HUMOR AND LAUGHTER, PLAYFULNESS AND CHEERFULNESS: UPSIDES AND DOWNSIDES TO A LIFE OF LIGHTNESS

EDITED BY: Willibald Ruch, Tracey Platt, René T. Proyer and Hsueh-Chih Chen
PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714
ISBN 978-2-88945-926-1
DOI 10.3389/978-2-88945-926-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

HUMOR AND LAUGHTER, PLAYFULNESS AND CHEERFULNESS: UPSIDES AND DOWNSIDES TO A LIFE OF LIGHTNESS

Topic Editors:
Willibald Ruch, University of Zurich, Switzerland
Tracey Platt, University of Sunderland, United Kingdom
René T. Proyer, Martin-Luther-University Halle-Wittenberg, Germany
Hsueh-Chih Chen, National Taiwan Normal University, Taiwan

Children's Games by Pieter Bruegel the Elder.
Image in the Public Domain.

The emergence of positive psychology has highlighted the importance of studying the good life and how to attain it. Positive life outcomes, such as well-being, thriving, flourishing, and happiness were discussed and investigated. Among them, different orientations to happiness were identified, such as a life of pleasure, life of meaning, and life of engagement. Other outcomes, such as subjective and objective fulfillment in life or societal recognition have been less studied. Among the characteristics that facilitate positive outcomes, the VIA-classification of strength and virtues distinguishes 24 strengths with humor/playfulness being one of them. Only a small segment of humor entered the definition of humor as character strengths, namely the parts that contain some "goodness". Humor as a character strength facilitates a lot of positive outcomes, such as positive emotions and positive relationships, and there is a "lightness" accompanying humor/playfulness.

The field is, however, broader. It transcends the definition of humor as used in positive psychology in at least two ways. First, there is an actual family of overlapping but still distinct concepts, both with different research traditions. We include, aside of humor (and types of humor), laughter, playfulness and cheerfulness. We think that more research is needed on how they do overlap and what makes them distinct. Second, while positive psychology is interested in the goodness of we do want to stress that there is the need to study the non-virtuous parts as well. That is, laughter may not only be expressing amusement but scorn directed at people, humor may be benevolent but there is also sarcasm, and playfulness may elicit positive emotions but also risk-prone and immature types of behavior.

Therefore, the aim of this Research Topic was to collect current perspectives on humor, playfulness, laughter, and cheerfulness in both adults and children, to study their full diversity but also interrelations and overlapping features, to introduce new instruments or ways for their assessment in future studies, and to study their causes and consequences in a variety of life domains. We encouraged studies on differences due to gender or nationality, the embodiment in different groups (e.g., class clowns, psychiatric patients), or whether or not they can be trained. We also welcomed contributions from adjacent disciplines (e.g., education, leisure studies, or therapy/counseling) and different regions of the earth.

The outcome is a set of 33 manuscripts from altogether 101 authors. Not all areas are covered and not all aims were met; while we made progress there is much left to do. In this sense, the merging of these topics may be the first milestone but like every milestone, it only marks the beginning of a long journey.

Citation: Ruch, W., Platt, T., Proyer, R. T., Chen, H.-C., eds. (2019). Humor and Laughter, Playfulness and Cheerfulness: Upsides and Downsides to a Life of Lightness. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-926-1

Table of Contents

06 ***Editorial: Humor and Laughter, Playfulness and Cheerfulness: Upsides and Downsides to a Life of Lightness***
Willibald Ruch, Tracey Platt, René T. Proyer and Hsueh-Chih Chen

HUMOR AND ITS KIN (CONCEPTS, MEASUREMENT, TRAINING, APPLICATION)

13 ***Broadening Humor: Comic Styles Differentially Tap Into Temperament, Character, and Ability***
Willibald Ruch, Sonja Heintz, Tracey Platt, Lisa Wagner and René T. Proyer

31 ***Pleasures of the Mind: What Makes Jokes and Insight Problems Enjoyable***
Carla Canestrari, Erika Branchini, Ivana Bianchi, Ugo Savardi and Roberto Burro

52 ***Exposure to Political Disparagement Humor and its Impact on Trust in Politicians: How Long Does it Last?***
Andrés Mendiburo-Seguel, Salvador Vargas and Andrés Rubio

64 ***The Social Life of Class Clowns: Class Clown Behavior is Associated With More Friends, but Also More Aggressive Behavior in the Classroom***
Lisa Wagner

75 ***Studying Irony Detection Beyond Ironic Criticism: Let's Include Ironic Praise***
Richard Bruntsch and Willibald Ruch

90 ***When Sugar-Coated Words Taste Dry: The Relationship Between Gender, Anxiety, and Response to Irony***
Anna Milanowicz, Adam Tarnowski and Barbara Bokus

107 ***Psychometric Comparisons of Benevolent and Corrective Humor Across 22 Countries: The Virtue Gap in Humor Goes International***
Sonja Heintz, Willibald Ruch, Tracey Platt, Dandan Pang, Hugo Carretero-Dios, Alberto Dionigi, Catalina Argüello Gutiérrez, Ingrid Brdar, Dorota Brzozowska, Hsueh-Chih Chen, Władysław Chłopicki, Matthew Collins, Róbert Ďurka, Najwa Y. El Yahfoufi, Angélica Quiroga-Garza, Robert B. Isler, Andrés Mendiburo-Seguel, TamilSelvan Ramis, Betül Saglam, Olga V. Shcherbakova, Kamlesh Singh, Ieva Stokenberga, Peter S. O. Wong and Jorge Torres-Marín

126 ***Is an Ideal Sense of Humor Gendered? A Cross-National Study***
Sümeyra Tosun, Nafiseh Faghihi and Jyotsna Vaid

134 ***Experimentally Manipulating Items Informs on the (Limited) Construct and Criterion Validity of the Humor Styles Questionnaire***
Willibald Ruch and Sonja Heintz

148 ***Assessing Theory of Mind by Humor: The Humor Comprehension and Appreciation Test (ToM-HCAT)***
Simge Aykan and Erhan Nalçacı

161 ***Humor Assessment and Interventions in Palliative Care: A Systematic Review***
Lisa M. Linge-Dahl, Sonja Heintz, Willibald Ruch and Lukas Radbruch

173 ***Investigating Humor in Social Interaction in People With Intellectual Disabilities: A Systematic Review of the Literature***
Darren David Chadwick and Tracey Platt

189 ***Feasibility of a Humor Training to Promote Humor and Decrease Stress in a Subclinical Sample: A Single-Arm Pilot Study***
Nektaria Tagalidou, Viola Loderer, Eva Distlberger and Anton-Rupert Laireiter

199 ***Who Benefits From Humor-Based Positive Psychology Interventions? The Moderating Effects of Personality Traits and Sense of Humor***
Sara Wellenzohn, René T. Proyer and Willibald Ruch

LAUGHTER & DISPOSITIONS TOWARDS LAUGHTER AND BEING LAUGHED AT

209 ***Telling Friend From Foe: Listeners are Unable to Identify In-Group and Out-Group Members From Heard Laughter***
Marie Ritter and Disa A. Sauter

217 ***Social Context Disambiguates the Interpretation of Laughter***
William Curran, Gary J. McKeown, Magdalena Rychlowska, Elisabeth André, Johannes Wagner and Florian Lingenfelser

229 ***Candidate Performance and Observable Audience Response: Laughter and Applause–Cheering During the First 2016 Clinton–Trump Presidential Debate***
Patrick A. Stewart, Austin D. Eubanks, Reagan G. Dye, Zijian H. Gong, Erik P. Bucy, Robert H. Wicks and Scott Eidelman

244 ***Extraversion is a Mediator of Gelotophobia: A Study of Autism Spectrum Disorder and the Big Five***
Meng-Ning Tsai, Ching-Lin Wu, Lei-Pin Tseng, Chih-Pei An and Hsueh-Chih Chen

255 ***Eye Contact and Fear of Being Laughed at in a Gaze Discrimination Task***
Jorge Torres-Marín, Hugo Carretero-Dios, Alberto Acosta and Juan Lupiáñez

269 ***Assessing Dispositions Toward Ridicule and Laughter in the Workplace: Adapting and Validating the PhoPhiKat-9 Questionnaire***
Jennifer Hofmann, Willibald Ruch, René T. Proyer, Tracey Platt and Fabian Gander

279 ***Evaluation of a Picture-Based Test for the Assessment of Gelotophobia***
Willibald Ruch, Tracey Platt, Richard Bruntsch and Róbert Ďurka

293 ***Relations of Dispositions Toward Ridicule and Histrionic Self-Presentation With Quantitative and Qualitative Humor Creation Abilities***
Karl-Heinz Renner and Leonie Manthey

303 ***Fear of Being Laughed at in Children and Adolescents: Exploring the Importance of Overweight, Underweight, and Teasing***
Carl-Walter Kohlmann, Heike Eschenbeck, Uwe Heim-Dreger, Michael Hock, Tracey Platt and Willibald Ruch

318 ***Fear of Being Laughed at in Borderline Personality Disorder***
Carolin Brück, Stephanie Derstroff and Dirk Wildgruber

PLAYFULNESS

323 ***The Education of Playful Boys: Class Clowns in the Classroom***
Lynn A. Barnett

341 ***The Relation Between Teachers' and Children's Playfulness: A Pilot Study***
Shulamit Pinchover

349 ***The Positive Relationships of Playfulness With Indicators of Health, Activity, and Physical Fitness***
René T. Proyer, Fabian Gander, Emma J. Bertenshaw and Kay Brauer

365 ***How Playfulness Motivates – Putative Looping Effects of Autonomy and Surprise Revealed by Micro-Phenomenological Investigations***
Katrin S. Heimann and Andreas Roepstorff

380 ***An Initial Cross-Cultural Comparison of Adult Playfulness in Mainland China and German-Speaking Countries***
Dandan Pang and René T. Proyer

CHEERFULNESS

395 ***Assessing the Temperamental Basis of the Sense of Humor: Adaptation of the English Language Version of the State-Trait Cheerfulness Inventory Long and Standard Form***
Jennifer Hofmann, Hugo Carretero-Dios and Amy Carrell

410 ***Trait Cheerfulness Does not Influence Switching Costs but Modulates Preparation and Repetition Effects in a Task-Switching Paradigm***
Raúl López-Benítez, Hugo Carretero-Dios, Alberto Acosta and Juan Lupiáñez

424 ***Are Clowns Good for Everyone? The Influence of Trait Cheerfulness on Emotional Reactions to a Hospital Clown Intervention***
Sarah Auerbach

frontiers in Psychology

EDITORIAL
published: 09 April 2019
doi: 10.3389/fpsyg.2019.00730

Editorial: Humor and Laughter, Playfulness and Cheerfulness: Upsides and Downsides to a Life of Lightness

Willibald Ruch[1]*, *Tracey Platt*[2], *René T. Proyer*[3] and *Hsueh-Chih Chen*[4]

[1] Department of Psychology, University of Zurich, Zurich, Switzerland, [2] Department of Psychology, University of Sunderland, Sunderland, United Kingdom, [3] Department of Psychology, Martin-Luther-University Halle-Wittenberg, Saxony-Anhalt, Germany, [4] Department of Psychology, National Taiwan Normal University, Taipei, Taiwan

Keywords: humor, playfulness, laughter, cheerfulness, gelotophobia, wit

Editorial on the Research Topic

Humor and Laughter, Playfulness and Cheerfulness: Upsides and Downsides to a Life of Lightness

INTRODUCTION

This research topic brings together the four research areas of humor, laughter, playfulness, and cheerfulness. There are partial overlaps among these phenomena. Humor may lead to laughter but not all laughter is related to humor. Playfulness is considered the basis of humor (a play with ideas), but not all play is humorous. Cheerfulness is considered the temperamental basis of good humor, a disposition for laughter and for keeping humor in face of adversity but it mostly overlaps with the socio-affective component of humor. Laughter was considered a play signal and to indicate the annulment of seriousness, but there is play without laughter and laughter outside of play. Cheerfulness might facilitate play and cheerful state might be raised due to play but again the conceptual overlap is only partial. They all contribute to levity in life and their apparent similarity suggests studying them together to map out the territory; i.e., to see where they overlap and what is specific. While these traits and behaviors have the potential to contribute to a good life, there is the danger of overlooking their non-virtuous facets; that is, laughter may not only be expressing amusement but scorn directed at people, humor may be benevolent but there is also sarcasm, and playfulness may elicit positive emotions but also risk prone behaviors. While this research topic solicited articles to these four domains without the aim to connect them, a few articles did and it is expected that growing together will be one outcome of this compilation of articles.

Currently, these fields are studied mostly in isolation. A literature search (using the psychology database of Web of Science Core Collection from 1900, 06.08.2018) yielded that humor is clearly leading in terms of number of publications (n = 3,006), followed by laughter (n = 1,412), playful(ness) (n = 629), and cheerful(ness) (n = 204). As a comparison, antonyms were studied as well, and yielded higher numbers, such as for crying (n = 1640), serious-mindedness (or seriousness) (n = 892), and sadness (n = 3,654). The latter indicates that sadness is 18 times more frequently researched than cheerfulness.

Next, the frequency of articles combining terms was investigated. Combinations of humor and one of the other key terms are rather infrequent with the exception of "humor and laughter" (n = 454), suggesting that about 10% of all articles on humor also refer to laughter. Humor and playfulness (n = 59) and humor and cheerfulness (n = 53) represent only 2% of all articles on

OPEN ACCESS

Edited and reviewed by:
Nadin Beckmann, Durham University, United Kingdom

***Correspondence:**
Willibald Ruch
w.ruch@psychologie.uzh.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *22 February 2019*
Accepted: *15 March 2019*
Published: *09 April 2019*

Citation:
Ruch W, Platt T, Proyer RT and Chen H-C (2019) Editorial: Humor and Laughter, Playfulness and Cheerfulness: Upsides and Downsides to a Life of Lightness. Front. Psychol. 10:730. doi: 10.3389/fpsyg.2019.00730

humor, and these numbers are still much higher than any combination among the other three. This clearly shows that work is needed integrating these areas to examine how the concepts overlap both regarding their defining substance but also in predicting third variables. It should be mentioned that in a pioneering publication preceding the renaissance of empirical humor research three of the keywords were considered together. Toronto-based English psychologist (Berlyne, 1969) gave an account of laughter, humor, and play in a chapter in a handbook of social psychology. The compilation of research in the four fields is aimed at deepening our understanding of these concepts and stimulating research combining them.

OVERVIEW OF STUDIES

There are 32 manuscripts in this research topic. Not surprisingly, most articles are on various aspects of humor, followed by laughter (including dispositions to ridicule and being laughed at), playfulness and cheerfulness. To highlight some prevalent issues beforehand: Individual studies relate to introducing new concepts, or new scales or working on existing ones (Aykan and Nalçaci; Bruntsch and Ruch; Heintz et al.; Hofmann et al.; Hofmann et al.; Ruch and Heintz; Ruch et al.; Ruch et al.). Furthermore, substantial attempts are made to develop and evaluate trainings and interventions (Auerbach; Linge-Dahl et al.; Tagalidou et al.; Wellenzohn et al.). There is also a significant number of cross-cultural comparisons (Heintz et al.; Pang and Proyer; Tosun et al.) and systematic literature reviews (Chadwick and Platt; Linge-Dahl et al.). What research questions were posed and what have we learned in the different fields?

Humor and Humor-Related Traits

Seven contributions relate to humor. Two are systematic reviews summarizing the use of humor in their related fields. Chadwick and Platt's paper draws upon the 32 existing articles on humor with regards to intellectual disability, which they found grouped into eight emergent themes. The paper showed humor to be of importance in social interactions, not only for people with intellectual disabilities but those who support them and highlighted both the positive and the negative role of humor for both groups. However, the authors suggest that future studies should aim for more empirical rigor when investigating this important, yet complex construct. As Heintz et al. highlighted, the terminology of a dichotomized thinking of positive and negative humor may be a too simplistic approach, especially when thinking about fostering positive relationships. For example, employing carers with a propensity for benevolent humor may help forge more than a work relationship, but a friendship.

In the study of humor assessment and interventions in palliative care, Linge-Dahl et al. reviewed 13 papers. The review found that although the papers were difficult to compare, it was clear that humor is an appropriate and useful resource in palliative care of terminally ill patients (in different settings, such as hospices or oncology wards). Given this review accounts for the last 20 years, the authors note that research is still exceptionally limited, although humor interventions showed promising results on many well-being outcomes.

Humor as a quality that can be trained and developed evidently has potential not only to increase well-being in the terminally ill but also to reducing stress, depressiveness, and anxiety in a population of sub-clinical individuals (Tagalidou et al.). This pilot intervention demonstrated encouraging evidence that a humor training can have a stable, long-lasting impact on increasing positive affective states and reducing levels of stress, depressiveness and anxiety. This study also reported a relatively low attrition rate, which would suggest that participants were enjoying themselves, whilst having an overall positive impact on their mental health.

Wellenzohn et al. studied who benefits from online humor-based positive psychology interventions. In Study 1, personality traits were tested and it was the extraverts that benefitted more from the three funny things intervention than introverts did. Remembering emotional events allows reliving the emotion and the extraverts' tendency to positive emotions (i.e., the amusement due to the funny events during the day) apparently contributed to increasing their level of happiness and to lowering their depressive symptoms. In Study 2, no moderating effects were found for sense of humor on the effectiveness of the five humor-based interventions tested. Interestingly, however, changes in sense of humor from pretest to the 1-month follow-up predicted later changes in happiness and depressive symptoms. Thus, increases in sense of humor during and after the intervention are associated with the interventions' effectiveness.

Instruments that measure aspects of humor were investigated in five studies. Heintz et al. investigate responses to the BenCor in 25 samples from 22 countries. The BenCor measures humor aiming at the good and may be seen as a character (as different from personality or temperament) approach to humor. Benevolent humor treats human weaknesses and wrongdoings benevolently, while corrective humor aims at correcting and bettering them. The 12 items exhibited sufficient psychometric qualities in most of the samples. Metric measurement invariance was supported across the 25 samples, and scalar invariance was supported across age and across gender. This study supported the suitability of the 12 marker items of benevolent and corrective humor in different countries, enabling cross-cultural research and eventually applications of humor aiming at the good. Importantly, benevolent and corrective humor were clearly established as two positively related, yet distinct dimensions of virtue-related humor.

Ruch and Heintz study the construct and criterion validity of the HSQ (Martin et al., 2003), which assesses humor styles. They argue that each item entails construct-relevant content (i.e., humor) but also (unwanted) variance produced by the item context. The 32 items were experimentally manipulated to strip off the context or to substitute the humor content by non-humorous alternatives (i.e., only assessing context). Study 1 shows that humor is not the primary source of the variance in three of the HSQ scales with the self-defeating humor style being primarily determined by the context. Study 2 shows that also the relationships of the HSQ with personality were reduced and those with subjective well-being vanished when the non-humorous

contexts in the HSQ items were controlled for. For self-defeating, removing the context rendered the results to a positive rather than a negative view of the humor in this humor style. The results suggest that the items of humor instruments warrant careful examination.

Ruch et al. enlarge the list of styles of humor by adding fun, benevolent humor, non-sense, wit, irony, satire, sarcasm, and cynicism and by providing first evidence for the reliability and validity of a set of 48 marker items for their assessment, the Comic Style Markers (CSM). Exploratory and confirmatory factor analyses showed that the eight styles could be distinguished in English- and German-speaking samples, and studying self- and other-reports supported both convergent and discriminant validity. Studies also showed that the scales tapped differentially into personality, intelligence, and character strengths; for example, wit correlated with verbal intelligence, fun with indicators of vitality and extraversion, and while benevolent humor was related to strengths of the heart, the styles related to mock/ridicule (i.e., sarcasm, cynicism, but also irony) correlated negatively with character strengths. The results suggest that more styles may be distinguished than was done hitherto, which is also confirmed by Heintz and Ruch (2019).

Two more studies examine irony in more detail and distinguish between two forms. Bruntsch and Ruch investigate irony in ironic criticisms (i.e., mock positive evaluation of negative circumstances) and ironic praise (i.e., mock negative evaluation of positive circumstances). They introduce the TOVIDA (Test of Verbal Irony Detection Aptitude) containing 26 scenario-based items for the detection of ironic criticism vs. ironic praise. Initial validation is provided by exploring personality and ability correlates of the two TOVIDA scales. Relatedly, Milanowicz et al. study mocking compliments and ironic praise from an interactional gender perspective. The ability to create irony is assessed and related to state and trait anxiety. Male responses were consistently more ironic but both genders used more irony in response to male ironic criticism than to female ironic praise. Anxiety predicted irony comprehension and willingness to use irony. The results enrich the discussion within the framework of linguistic intergroup bias and natural selection strategies.

Also Aykan and Nalçaci introduce a new instrument (ToM-HCAT) for assessment of ToM (i.e., theory of mind) by humor comprehension and appreciation suitable for healthy adult populations. This performance test consisting of cartoons measures perceived funniness, reaction time to perceived funniness decision, and meaning inference. While a first validation is presented (individuals high and low in the Autism Spectrum Quotient differ in the meaning-inference scores of the subscale with the ToM cartoons) it awaits further validation to support the claim it is useful to detect variations in ToM ability in the healthy adult population.

While Heintz et al. study country differences in measured humor traits, Tosun et al. explore lay conceptions of an ideal sense of humor in three countries, namely Iran, United States, and Turkey. As in prior US studies they find that the embodiment of an ideal sense of humor is predominantly a male figure. Country and gender had an impact on relative number of specific humor characteristics. For example, Americans mentioned hostility/sarcasm and caring more often than participants from the other countries. Further work is needed to replicate the observed group differences and to identify their sources.

Canestrari et al. use the Theory of the Pleasures of the Mind to study the enjoyment derived from both humor and insight problem solving as they share similar cognitive mechanisms. The results show that finding the solution to a problem is associated with a positive evaluation, and curiosity, virtuosity and violation of expectations are the most frequent explanations. Understanding a joke is accompanied by the joy of verification and a feeling of surprise. However, the choice for the most enjoyable cartoons related to other factors, such as recognizing a violation of expectations and experiencing a diminishment in the cleverness attributed to the characters in the cartoon.

Mendiburo-Seguel et al. investigate the effects of political humor on an individual's trust toward politics and politicians. They conducted two experiments, in which participants were exposed to political disparagement humor to non-humorous political information, or to non-political humor. Study 1 showed that an exposure to political disparagement humor and non-humorous political contents negatively affects trust in politicians immediately after the exposure. Study 2, in which semidaily messages were sent to the participants, did not yield significant effects.

The study by Wagner nicely demonstrates how close upside and downside of humor are together by showing that class clown behavior was positively related to different indicators of social status and peer-rated popular-leadership behavior, but also to aggressive-disruptive behaviors and negatively to prosocial behaviors. Thus, humor is involved in making a student popular but it may also be used in destructive ways. The study also demonstrates that it is important to distinguish among different dimensions of class clown behavior, as they yielded different results.

Laughter and Dispositions to Ridicule and Being Laughed at

Laughter is both a social signal and an expression of emotion with several behavioral and physiological components (e.g., respiratory, acoustic, facial, postural, hormonal). There are different motivations for laughter (with laughing *with* and laughing *at* being a minimal distinction made by many) and there are individual differences to be considered regarding both the laughing person and the one perceiving the laughter. Laughter is studied among the healthy but also within psychopathology. Clearly, the section of this research topic devoted to laughter and laughter-related dispositions received a variety of submissions.

Ritter and Sauter investigated whether listeners can identify in- and out-group members from laughter. They showed that listeners were unable to accurately identify group identity from laughter and the exposure to a group did not affect the classification performance. In conclusion, group membership cannot be inferred from the way people laugh.

Curran et al. test the notion that laughter is an ambiguous signal, which is only interpreted correctly in the context it occurs.

They provide supportive data from two experiments in which participants judged the genuineness of audio–video recordings of social interactions containing laughter (either original or replacement laughter). When replacement laughter was matched for intensity, genuineness judgments were similar to judgments of the original recordings. When replacement laughter was not matched for intensity, genuineness judgments were generally significantly lower.

Stewart et al. used the 2016 US presidential debates to study laughter together with other responses of audience, such as applause, cheering, laughter, and even booing. In three interconnected studies the impact of the norm-violating audience behavior on those watching or listening was studied. Applause–cheering significantly enhanced liking of the speaking candidate, whereas laughter did not, and party identity mediated the response to applause–cheering, but not for laughter. Thus, in such settings, cheering may be more socially contagious and laughter more stereotypic and likely to be mimicked.

The study by Auerbach confirms that it is important to distinguish between Duchenne Displays as an indicator of joy and non-Duchenne displays. Only the former go along with a variety of indicators of positive experience during a visit of hospital clowns in a rehabilitation center. Thus, also in such interventions it pays off to invest into the fine-grained assessment of facial expressions; i.e., to use the *Facial Action Coding System* to code the patients' affective responses. Only the Duchenne displays are affected by trait cheerfulness and they can serve as an indicator that hospital clown interventions are beneficial for patients.

The study of laughter also includes the dispositions to laughter—more precisely individual differences in qualities relating to laughing at and being laughed at. They are still the new kid on the block of variables related to humor and laughter with a research tradition of about 10 years. Gelotophobia (i.e., the fear of being laughed at) represents one form of humorlessness and gelotophobes see humor and laughter as weapons directed at them not as a basis for a pleasant experience to be shared with others. Together with gelotophilia (i.e., the joy of being laughed at) and katagelasticism (i.e., the joy of laughing at others) gelotophobia forms the dispositions to being laughed at and ridicule.

Two of the articles in the present collection of articles relate to their assessment. Ruch et al. utilize a picture completion task to derive a more unobtrusive semi-projective test of gelotophobia. This alternative instrument for the assessment of gelotophobia turns out to yield comparable results to the standard assessment. Hofmann et al. fulfill the need for an ultra short instrument for the assessment of these three dispositions and extends research into the workplace. They propose (and confirm in a nationally representative sample of employees) that if friendly teasing and laughter of co-workers, superiors, or customers are misperceived as malicious, one may feel less satisfied with work and life and experience more work stress. Conversely, gelotophilia went along with positive evaluations of one's life and work, and katagelasticism was negatively related to work satisfaction and positively related to work stress. Torres-Marín et al. provide evidence that gelotophobia is related to a potential bias in gaze discrimination in two experiments. Interestingly, the nature of the emotion did not play the expected role raising the question what elements are necessary for smiling faces to elicit the effect among gelotophobes.

Renner and Manthey investigate humor creation abilities in their study of self-presentation styles and dispositions to ridicule and being laughed at. They derive scores for quantitative (e.g., number of punch lines) and qualitative (e.g., wittiness of the punch lines and wittiness of the person as evaluated by three independent raters) aspects of humor creation abilities. Results show that both gelotophilia and histrionic self-presentation are supported by fluency and quality of humor creation abilities.

Three manuscripts examine gelotophobia in circumscribed groups. Kohlmann et al. investigated the associations between the experience of weight-related teasing and mockery with overweight, self-perceptions of weight, and gelotophobia in youth. Deviations from normal weight were related to experiencing teasing, which in turn was related to the fear of being laughed at. The four studies suggest that research on well-being of youth with weight problems would benefit from studying weight-related teasing and mockery in connection with gelotophobia. Tsai et al. study the relation between the dispositions toward ridicule and being laughed at, personality, and presence of autism spectrum disorder (ASD) in high school students. As in prior studies, the ASD group was found to have a higher level of gelotophobia and the present study reveals that they also have lower levels of gelotophilia and katagelasticism. However, extraversion fully accounted for the observed lower gelotophobia scores among the ASD sample, and partly for the differences found for gelotophilia. Brück et al. investigated the prevalence of gelotophobia among Borderline Personality Disorder patients. They showed an extraordinarily high level of the fear of being laughed at (i.e., 87%) compared to other clinical and non-clinical reference groups.

Playfulness

The section on playfulness consists of five contributions of which two have a qualitative approach, while the others are quantitative in nature. Two contributions focus on *play* (the behavior associated with trait playfulness) and playfulness in school and the others employ adult samples. With 1,235 Tweets reaching an upper bound of 3,945,511 followers (March 25th, 2019)[1], Barnett's article attracted much attention on social media. Her analyses show that teachers react differently—more negatively—toward playfulness expressed by boys than by girls (kindergarten-aged children followed up across 3 years). In contrast, playfulness in girls did not seem to be a concern for the teachers. The methodology employed and the study of gender differences provides a valuable update on earlier literature. Overall, the emerging question is how teachers, schools and societies in general may benefit from playfulness in the classroom.

Pinchover's pilot study examines the interplay of playfulness in teachers and their students. Taking the limitations of this initial study into account, this may indicate that teacher behavior impacts children's playfulness. Given that there is initial evidence for a contribution of playfulness to academic achievement and

[1]https://frontiers.altmetric.com/details/33125117/twitter

more robust data on a beneficial use for stress coping, some functions of playfulness may be helpful for students in their learning experience and development.

The idea that a playful state of mind contributes to innovativeness and creativity has received much interest in the literature (for overviews see Proyer et al., 2019) and, for example, it has been argued that "[…] a child who experiences truly "playful play" learns cognitive and behavioral processes that enhance his creative potential" (Bishop and Chace, 1971; p. 321). Heimann and Roepstorff introduce microphenomenological interviews as a method for research in playfulness. In this initial study, they found that autonomy and self-expression were of particular importance for achieving a playful state of mind.

Proyer et al. test associations of playfulness with self-reported health, activity, and physical fitness. Self- and peer-ratings (i.e., ratings by knowledgeable others; Study 1) and a series of behavioral tests (Study 2) to assess playfulness were collected. Overall, playfulness is linked to some facets of physical functioning. Future research will have to clarify the pathways and moderators of these associations (e.g., causality or indirect ways of impacting greater physical activity).

Finally, Pang and Proyer present first data on a comparison of playfulness scores in samples from two regions in the P.R. China and a sample from German-speaking countries—using measures from both, the East and the West. The article provides details on cultural differences and linguistic challenges in the translation of the term playfulness. Overall, the findings indicate that differences are smaller than expected, but that the differentiation between private and public situations impacts how people in the two regions enjoy expressing their playfulness. This study narrows a gap in the literature by providing initial data on cross-cultural differences (see also Barnett, 2017) and highlights that larger scale cross-cultural comparisons are encouraged.

These five studies support the notion that playfulness has an impact on various domains of life, but also that more research will be needed for a better understanding of its role across different age groups.

Cheerfulness

Cheerfulness has a tradition in psychological research for more than 100 years (e.g., Morgan et al., 1919). Trait cheerfulness, seriousness, and bad mood have been proposed to form the temperamental basis of humor. Bypassing the vague folk concept of the "sense of humor" they were expected to predict humor-related thoughts, feelings, and actions. Washburn in her early studies claimed that a person in the attitude of cheerfulness is incapable of a depressing thought, and meanwhile there is ample evidence that trait cheerful individuals maintain being in a cheerful state (i.e., keep humor) in the face of adversity. The contributions of the present collection of articles are diverse. First, a training of humor yielded outcomes for cheerfulness, seriousness, and bad mood) in the desired direction with medium to large effect sizes (Tagalidou et al.). Different to a recent study (Ruch et al., 2018) the state version was utilized. Congruent with the assumption that cheerfulness predicts smiling and laughter, Auerbach shows that trait cheerful patients showed more genuine smiling and laughter during a hospital clown intervention than low trait cheerful individual do. Hofmann et al. present an adaptation of the instrument measuring state and trait cheerfulness using samples from the USA and the UK to providing the basis for studies with English-speaking participants. Next to the long version with 106 items, they provide the standard short form with 60 items and deliver initial validation data. López-Benítez et al. investigate a cognitive mechanism associated with trait cheerfulness. Utilizing a task-switching paradigm they find that while trait cheerfulness does not influence switching costs it modulates preparation and repetition effects. Studies like this are needed to further illuminate the processes associated with the traits be it cheerfulness, playfulness, or humor. Bruntsch and Ruch find trait cheerfulness and low bad mood facilitating the detection of ironic praise.

CONCLUSIONS

The individual contributions show how humor, laughter, playfulness, and cheerfulness are related and yet heterogeneous. Each field would profit from starting to talk to each other, see overlaps in scope, finding common structure, common language, and work on theories connecting these fields. Combining the domains in the prediction of important criteria might be important too. The topics studies in this research topic (plus others) may be understood as nodes in a larger net and the interrelations need to be better explored.

It is positive to see that integrative models within the domains are now developed. This indeed needs to be the prime goal, namely to work on a solid structure within the four fields. It took research of personality and intelligence more than half a century to arrive at models that are shared by many. Also in these fields we once had "schools" that did believe into one model and defended it a lifetime. Later generations of researchers then found that the competing models were incomplete variants and do fit into a more general, often hierarchical model. We recommend concerted efforts to solve those basic questions, perhaps by compiling special issues on pertinent topics.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We would like to thank all the authors who agreed to participate in this Topic with their original contributions, and to all the reviewers who promoted the quality of research and manuscripts with their comments. Furthermore, special remarks go to Frontiers staff and Professors Marcel Zentner and Anat Bardi for the opportunity they gave to us.

REFERENCES

Barnett, L. A. (2017). The inculcation of adult playfulness: from west to east. *Int. J. Play* 6, 255–271. doi: 10.1080/21594937.2017.1383010

Berlyne, D. E. (1969). "Laughter, humor, and play," in *Handbook of Social Psychology, Vol. 3,* eds G. Lindzey and E. Aronson (Reading, MA: Addison-Wesley), 795–852.

Bishop, D. W., and Chace, C. A. (1971). Parental conceptual systems, home play environments, and potential creativity in children. *J. Exp. Child Psychol.* 12, 318–338. doi: 10.1016/0022-0965(71)90028-2

Heintz, S., and Ruch, W. (2019). From four to nine styles: an update on individual differences in humor. *Pers. Individ. Diff.* 141, 7–12. doi: 10.1016/j.paid.2018.12.008

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., and Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: development of the humor styles questionnaire. *J. Res. Pers.* 37, 48–75. doi: 10.1016/S0092-6566(02)00534-2

Morgan, E., Mull, H.K., and Washburn, M.F. (1919). An attempt to test the moods or temperament of cheerfulness and depression by directed recall of emotionally toned experiences. *Am. J. Psychol.* 30, 302–304.

Proyer, R. T., Tandler, N., and Brauer, K. (2019). "Playfulness and creativity: a selective review," in *Creativity and Humor,* eds S. R. Luria, J. Baer, and J. C. Kaufman (San Diego, CA: Academic Press), 43–60. doi: 10.1016/B978-0-12-813802-1.00002-8

Ruch, W., Hofmann, J., Rusch, S., and Stolz, H. (2018). Training the sense of humor with the 7 Humor Habits Program and satisfaction with life. *Humor Int. J. Humor Res*. 31, 287–309. doi: 10.1515/humor-2017-0099

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ruch, Platt, Proyer and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 18 January 2018
doi: 10.3389/fpsyg.2018.00006

Broadening Humor: Comic Styles Differentially Tap into Temperament, Character, and Ability

Willibald Ruch[1], Sonja Heintz[1], Tracey Platt[2], Lisa Wagner[1] and René T. Proyer[3]*

[1] Personality and Assessment, Department of Psychology, University of Zurich, Zurich, Switzerland, [2] Institute of Psychology, University of Wolverhampton, Wolverhampton, United Kingdom, [3] Department of Psychology, Martin-Luther University Halle-Wittenberg, Halle, Germany

The present study introduces eight *comic styles* (i.e., fun, humor, nonsense, wit, irony, satire, sarcasm, and cynicism) and examines the validity of a set of 48 marker items for their assessment, the Comic Style Markers (CSM). These styles were originally developed to describe literary work and are used here to describe individual differences. Study 1 examines whether the eight styles can be distinguished empirically, in self- and other-reports, and in two languages. In different samples of altogether more than 1500 adult participants, the CSM was developed and evaluated with respect to internal consistency, homogeneity, test–retest reliability, factorial validity, and construct and criterion validity. Internal consistency was sufficiently high, and the median test-retest reliability over a period of 1–2 weeks was 0.86 (N = 148). Exploratory and confirmatory factor analyses showed that the eight styles could be distinguished in both English- (N = 303) and German-speaking samples (N = 1018 and 368). Comparing self- and other-reports (N = 210) supported both convergent and discriminant validity. The intercorrelations among the eight scales ranged from close to zero (between humor and sarcasm/cynicism) to large and positive (between sarcasm and cynicism). Consequently, second-order factor analyses revealed either two bipolar factors (based on ipsative data) or three unipolar factors (based on normative data). Study 2 related the CSM to instruments measuring personality (N = 999), intelligence (N = 214), and character strengths (N = 252), showing that (a) wit was the only style correlated with (verbal) intelligence, (b) fun was related to indicators of vitality and extraversion, (c) humor was related to character strengths of the heart, and (d) comic styles related to mock/ridicule (i.e., sarcasm, cynicism, but also irony) correlated negatively with character strengths of the virtues temperance, transcendence, and humanity. By contrast, satire had a moral goodness that was lacking in sarcasm and cynicism. Most importantly, the two studies revealed that humor might be related to a variety of character strengths depending on the comic style utilized, and that more styles may be distinguished than has been done in the past. The CSM is recommended for future explorations and refinements of comic styles.

Keywords: humor, ridicule, fun, satire, wit, irony, personality, character

OPEN ACCESS

Edited by:
Monika Fleischhauer, Medizinische Hochschule Brandenburg Theodor Fontane, Germany

Reviewed by:
Kai Tobias Horstmann, Humboldt-Universität zu Berlin, Germany
Ursula Beermann, University of Innsbruck, Austria

****Correspondence:***
Willibald Ruch
w.ruch@psychologie.uzh.ch;
willibald.ruch@bluewin.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *08 September 2017*
Accepted: *03 January 2018*
Published: *18 January 2018*

Citation:
Ruch W, Heintz S, Platt T, Wagner L and Proyer RT (2018) Broadening Humor: Comic Styles Differentially Tap into Temperament, Character, and Ability. Front. Psychol. 9:6. doi: 10.3389/fpsyg.2018.00006

INTRODUCTION

Humor research has to accommodate that people habitually differ in terms of humor behaviors both quantitatively (i.e., some show more humor production, appreciation, etc., than others) and qualitatively (i.e., people indulge in different forms in humor). The quantitative part was taken care of by regarding humor as a continuum ranging from low to high, rather than solely distinguishing between having it and not having it. The qualitative part acknowledges that humor might have a different tone (*comic tonality*), *flavor* (just as people have different tastes), form, type, distinctive quality, or *style*. These qualitative differences may reflect mood (e.g., cheerful or bitter), degree of refinement (e.g., from low comedy often based on physical incongruity to high comedy of manners with more emphasis on language), structure (i.e., genre) or modality (for an overview, see Ruch, 2008).

So far, these differences have not been studied. For a systematic classification of styles, we not only need to consider the different qualities (i.e., the horizontal level) but also different levels of abstraction (i.e., the vertical level). Humor styles might be very specific, representing a narrow scope of behaviors (e.g., teasing, bantering), or they might be more general, covering a broader range of different behaviors (e.g., socially warm humor). The lower in the level of abstraction, the closer one is to a form of behavior that could actually be shown in a specific situation or that could be trained or modified. Conversely, this is less likely the higher up the hierarchy a style is located, as more general styles are abstractions representing dispositions to behaviors. Thus, the more a style is a composite of many behaviors, the less easily it can be shown, trained, or modified as a block (in its entirety). Lower-level approaches to styles are more informative but also more redundant, while more abstract approaches are more parsimonious, but at the expense of detailed descriptions. Hence, a comprehensive approach needs to consider different levels of aggregations.

The two approaches to humor styles introduced in the past two decades can be located at an intermediate level. The Humorous Behavior Q-sort Deck (HBQD; Craik et al., 1996) distinguishes 10 styles of everyday humorous conduct that were allocated to five bipolar dimensions; namely, socially warm versus cold, reflective versus boorish, competent versus inept, earthy versus repressed, and benign versus mean-spirited humorous styles. The 10 styles were derived from the intercorrelations of 100 items depicting everyday humor behaviors as represented in thoughts, behaviors, and attitudes. The Humor Styles Questionnaire (HSQ; Martin et al., 2003) distinguishes four trait-like humor styles, namely, the affiliative, self-enhancing, aggressive, and self-defeating humor styles. The former two are considered to be adaptive and the latter two maladaptive functions of humor. In both instruments, the "humor styles" do not represent elementary flavors, types, or distinctive qualities of humor (such as sarcasm or nonsense) but compounds of humor behaviors (that go together) or humor functions at a more general level of abstraction. None of them relates to a "known" or established category of humor, but they represent new constructs that cluster more diverse humor behaviors or functions.

The present approach aims at supplementing the existing styles by investigating lower-level styles, namely *humor, fun, nonsense, wit, irony, satire, sarcasm,* and *cynicism* as described by Schmidt-Hidding (1963). We argue that these styles reflect established categories of humor (in the broad sense) and that they are narrower than the ones in the HBQD and HSQ, which allows for a more fine-grained differentiation of humor-related styles. For example, in the HBQD, sarcasm is one element of the mean-spirited humor style (that also involves wit to keep people at distance), and nonsense is part of the benign humor style (which also includes appreciating intellectual word play and regularly exchanging topical jokes). Also, the aggressive humor style in the HSQ does not allow distinctions between different forms of mockery, such as satire, sarcasm, and cynicism (see Ruch and Heintz, 2016a). Employing narrower but distinct styles also allows speaking of "using" a humor style, as these represent smaller units that can be enacted, trained, and modified more easily. Such a list is now sought after.

Selecting and Describing the List of Comic Styles

The present manuscript represents the first step in a larger endeavor aimed at eventually arriving at a comprehensive list of lower-level styles, their classification, reliable and valid assessment, the study of their origin and consequences, life-span development, and the development and evaluation of interventions controlling (i.e., increasing, decreasing, or modifying) their use. We selected a manageable list from prior work in psychology, esthetics, philosophy, and other disciplines. In literary studies, humor styles (or: comic styles[1]) already have a longer tradition, and this crystalized knowledge can be transferred to the domain of assessment of individual differences. Different authors introduced various lists of styles. For example, Lauer (1974) distinguished nine styles; namely, humor, self-irony, comic in a narrow sense, fun, wit, irony, satire, sarcasm, and cynicism. Milner Davis (2003) identified a very comprehensive list covering farce/slapstick or low comedy,

[1]When talking about "comic styles," it becomes obvious that these authors adhere to a different terminological system that is used in some academic disciplines in some countries. As this is in conflict with contemporary English-speaking psychology, these two should be briefly contrasted as they assign different roles to the key term "humor." The historical nomenclature in literature stems from the field of esthetics where the funny (or: the comic) — defined as the faculty of being able to make someone laugh or to amuse — is distinguished from other esthetic qualities, such as beauty, harmony, or the tragic. In this tradition humor is simply one element of the funny — as are wit, fun, nonsense, sarcasm, ridicule, satire, or irony, and humor is in opposition to them (e.g., humor and sarcasm are excluding each other). Humor is not a neutral term here but exclusively positive. The alternative, almost incompatible, current use of "humor" in contemporary psychological research is an umbrella term for all phenomena of the funny, including the capacity to perceive, interpret, and enjoy but also create and perform non-serious incongruous communications (for an overview, see Ruch, 2008). Obviously, in this terminology humor has replaced the comic/funny as the supreme term and is treated as a neutral concept; that is, humor is not restricted to positive occasions for laughter. From this perspective it is not a contradiction to speak about "sarcastic humor"; that is, sarcasm that is funny. In the present article, we stick to the notion of "comic styles" to mark their origin, knowing that they can as well be called "humor styles" when one works within the other frame of reference, which is most prevalent in contemporary psychology. We consider the operational definition important rather than the labels.

comedy of manners/wit or high comedy, romantic/festive or sentimental comedy (e.g., sitcoms), ironic/parodic or burlesque comedy, nonsense humor/absurdist comedy, "sick"/disgust comedy (e.g., U.S. "shock-u" comics), satire/satirical comedy, "black"/gallows/existential comedy, and tragicomedy. By perusing the table of index words of different handbooks and encyclopedias of humor (e.g., Raskin, 2008; Attardo, 2014; Wirth, 2017) a few more candidates could be added, especially if different disciplines, countries, etc. are involved. Yet it is also apparent that a smaller set of styles is more frequently mentioned, and those that have clear behavioral implications (i.e., that can be applied to distinguish individuals) can be selected for psychological investigation.

The work of Schmidt-Hidding (1963) helps bridging the gap between the comic styles rooted in literary studies of humor and a personality perspective interested in describing individual differences regarding humor use (for a detailed account of his approach, see Ruch and Heintz, 2016b). Schmidt-Hidding described eight styles with seven features (exemplified with sarcasm); namely, (1) intention, goal (to hurt the partner), (2) object (the corrupt world), (3) attitude of the agent as subject (e.g., derisive, feels like an undiscovered genius, thus often maliciously critical), (4) behavior toward others (e.g., hostile), (5) the ideal audience (e.g., subordinate and dependent people, who don't dare to disagree), (6) method (e.g., ruthless exposure), and (7) linguistic peculiarities (e.g., ironic, with emphasis). Some of these features can be clearly located in the person, and others provide valuable additions to the definition of the constructs. For example, behavior is motivated by the intentions or goals which might be either conscious or not. The "object" might define individuals, as this is what they adhere to or find important. The attitude of the agent clearly is trait-like as is the behavior toward others. The ideal audience might be significant as people might search for the right audience to use their comic style. The last two features are less central to personality, but the use of these methods might be indicative of personality and individuals may or may not use the linguistic peculiarities well. Taken together, these seven features allow creating distinct prototypes of styles, and we supplemented these accounts by the study of other sources, and finally developed descriptions of the eight comic styles.

The prototypes of the styles, crystallized mainly from Schmidt-Hidding's descriptions, can be described as follows. Four styles may be considered dark ones (as opposed to lighter ones; see below), as they constitute a family of mockery/ridicule. In particular, *sarcasm* aims at hurting others. The sarcastic person is described, among others, as being hostile and derisive and as using ruthless exposure to highlight the corrupt world. The ideal audience consists of subordinate and dependent people. High scorers would see themselves as malignant and critical when decrying the corruption, depravity, vice, or evil. They are prone to scorn and *schadenfreude*. *Cynicism* is aimed at devaluing commonly recognized values. Cynics exhibit a negative and destructive attitude. They use disillusionment and mockery to highlight weaknesses in the world. Cynics do not lack moral values in general, yet they disdain certain common norms and moral concepts and find them ridiculous. *Satire* (a.k.a. corrective humor; Ruch and Heintz, 2016b) shares with sarcasm and cynicism the detection of weaknesses and is aggressive. However, this is paired with attempts at goodness. This involves not only deprecating the bad and foolish, but also the intention of improving the world and correcting fellow humans. A satirist takes the ethical world as a measure of the real one and attempts to improve conditions by disclosing the true circumstances. The satirist is critical, often negative, tense and superior, but prefers the world to be moral and uses ridicule to better the world. Although the aggressive tendency is the common element, the mockery is not done on the basis of sheer pleasure, but it is grounded in a moral-based criticism. People with a critical mindset typically approve satire. The goodness of satire appeals to change inappropriate behaviors or mindsets without seriously damaging the interpersonal relations. *Irony*, as expressed in interactions, aims at creating a mutual sense of superiority toward others by saying things differently than they mean it. It does not entail lying as one assumes that smart people will understand what was actually meant irrespective of what was said. Ironic people are courting and letting in the intelligent, thereby at the same time mocking the stupid. Irony is a means of confusing the non-insiders and finding out who is a knowledgeable informed insider. Others may see them as conceited, superior, and frequently negative-critical.

There are lighter styles that do not contain these skeptic elements. They are very diverse despite sharing a more positive basis of interpersonal cooperation, benevolence, positive emotions, and cognitive capabilities. Specifically, *fun* (joking, jesting) is aimed at spreading good mood and good comradeship. People using this comic style are considered to be social, jovial, and also agreeable. In everyday life situations, they use teasing (waggish, impish) with friends and people accustomed to bawdy matters. They might see themselves as funny jokers and like to make mischievous jests. They play harmless tricks on friends and like to jest and act clownish. Next, *humor* (a.k.a. benevolent humor; Ruch and Heintz, 2016b) aims at arousing sympathy and an understanding for the incongruities of life, the imperfections of the world, the shortcomings of fellow humans, and the own mishaps and blunders. People with humor are realistic observers of human weaknesses, but treat them benevolently, often including themselves in the judgment rather than directing it exclusively at others. There is an understanding for humanity in all weaknesses, which are observed and shared with a jovial, relaxed, and contemplative audience. Humor comes "from the heart" and reflects a tolerant, loving attitude toward others that includes accepting their shortcomings. A person with humor in this sense knows that, both on a large and small scale, the world is not perfect. Still, with a humorous outlook on the world even the adversities of life can be amusing and be smiled at. A person using this comic style manages to arouse understanding and sympathy for imperfections and the human condition through humor. *Nonsense*, as intellectual and playful, cheerful fun, aims at exposing the ridiculousness of the sheer sense, though basically without any purpose. People enjoying nonsense describe themselves as playful and cheerful. They let their mind play, for example, by being creative with language and by playing with sense and nonsense. For them, incongruities do

not need to be resolved, but the opposite holds true; that is, the more absurd, and the funnier. They create an upside-down world, use language in its imperfection, and find bizarre and fantastic stories amusing.

Finally, one style can be seen as part of the lighter styles despite also containing elements characteristic of the darker styles. *Wit* intends to illuminate like a flashlight, typically with a surprising punch line that uses unusual combinations created on the spot. A person using wit plays with words and thoughts, and they might be callous, malicious, and generally without sympathy for the "victims" in order to maximize the funny impact. Producing wit requires skills: It entails quickly reading situations and nailing non-obvious matters to the point in a funny way. They surprise others with funny remarks and accurate judgments of current issues, which occur to them spontaneously. They make relationships between disconnected ideas or thoughts and thus create a comical effect quickly and pointedly. Witty people might be tense, vain, and take themselves seriously, and look for an educated society that appreciates brief pointed utterances as an ideal audience.

Considerations on the Structure of the Styles

The use of many narrower styles will yield interrelated scales, and the intercorrelations could be used to derive fewer (and more) abstract styles, which poses the question what these different levels are good for. To uphold the use of the narrower styles, it is important to demonstrate (a) that they can be separated conceptually and empirically and (b) that each style predicts different phenomena and is not redundant. Some individuals might "use" certain styles more often than others, but each style is still functionally different. Being sarcastic does not necessarily mean that one is also more cynical, although these two styles will be highly correlated. Training to be witty might enhance wit, but not necessarily satire, although both might correlate as well. This suggests that it is best to keep the concepts at this level of abstraction, rather than, for example, cluster them together and use aggregated styles.

However, one can look at the interrelations among the styles (based on covariations of individual differences in a sample) and conduct a second-order factor analysis for two reasons. First, one can examine how these interrelations can be represented in a smaller space and describe the styles at an aggregated level. While there is no intention of reducing these styles to a fewer number of concepts, it might provide insights into the structure of the styles and indicate where they overlap. Should styles correlate too highly, one might consider dropping some or combining them at a conceptual level to form a new scale (but not a factor derived from it). Second, structure-building methods could be applied to empirically test the assumptions of different authors about the structure inherent in this list of styles. For example, Lauer (1974) ordered the styles (in the sequence listed above) to reflect different mixtures of two tendencies, namely self-assertion (as a consciousness-limiting tendency) and participation (as a consciousness-expanding tendency). Humor assumes a special role in this model, as it allows for an optimum of "euphoric" self-assertion and participation, while cynicism is lowest in this respect. This allows for predictions about the relative proximity of these two styles as well as the postulate that two factors might be sufficient to represent most of the variance. Interestingly, Schmidt-Hidding (1963) ordered the styles similarly. However, these are spread along a rhomboid that is marked by what he considered to be key terms (i.e., the most frequent terms) in the field of the comic, namely humor, wit, fun, and mock/ridicule. These and some satellite words (with lower frequency) as well as the comic styles are depicted in a topographical model (see **Figure 1A**). While the generation of the model is not fully explicated and it is also not clear whether these terms would be

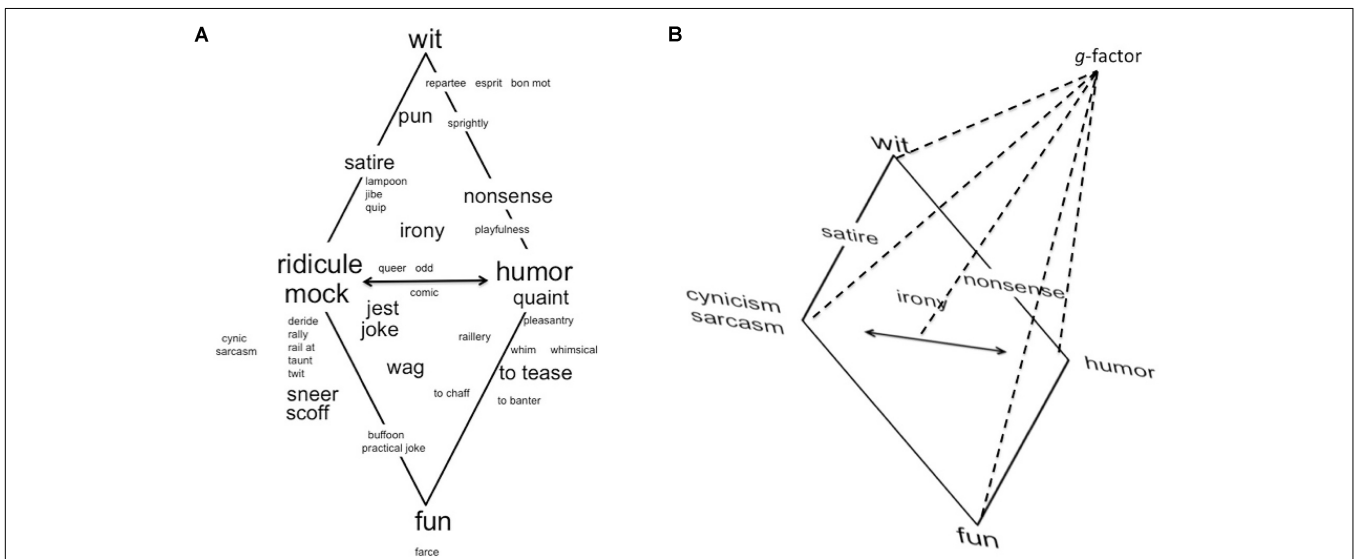


FIGURE 1 | The eight comic styles in a schematic representation together with the key terms and other humor-related words (**A**, left) and in an individual-differences model (**B**, right). Adapted from Schmidt-Hidding (1963, p. 48).

still the most frequent nowadays, this configuration can be taken allowing for hypotheses about the structure of the eight comic styles to be tested empirically (Study 1). Furthermore, Schmidt-Hidding (1963) saw "energizing forces" behind the key terms and the satellite words. Accordingly, humor can be contrasted from the other three key terms as being based on a "sympathetic heart" (guided by love), not a "superior spirit" (like wit), moral critique, or even haughtiness (guided by hatred) like mock/ridicule, or vitality/high spirits (like fun). These descriptions allow deriving the hypothesis that unique predictors for comic styles may come from the domains of ability (for wit) and character (for the virtuous forms) in addition to traditional personality traits (Study 2).

Two types of testing the structure of comic styles seem appropriate; namely, the relation among the depicted key terms/comic styles (see **Figure 1A**) and one that considers individual differences in the comic styles (see **Figure 1B**). The former test can be based on a multidimensional scaling of a matrix representing the degree of similarity or dissimilarity between all comic styles or all terms listed (based on either judging similarities/dissimilarities directly, or computing them from raw scores) or based on a second-order factor analysis of ipsative data (eliminating the third dimension; i.e., individual differences). Then one can examine whether humor and mock/ridicule (represented by sarcasm and cynicism) are indeed opposing each other, in as much as love and hate are opposites. There is also a north–south distinction, but it is not seen as a bipolar dimension in **Figure 1A**. There is intellectual wit in the north marking more clever, hidden, and verbal comic creations, with satellite words connecting toward mock/ridicule (e.g., satire, jibe, quip and lampoon) and humor (e.g., nonsense, playfulness). Fun is in the south position, with satellite terms connecting to mock (e.g., sneer, scoff) and humor (e.g., tease, banter). While this is not a bipolar dimension, the former is more akin to high comedy and the latter to low comedy. The order of the eight styles can be examined as well. From the arrangement of terms, cynicism and sarcasm are expected to be the most difficult to distinguish as they are very close to each other. Comparing self- and other-reports helps to show whether these comic styles can actually be separated from one another (discriminant validity) and whether one's self-evaluation and the perceptions of others converge (convergent validity).

The second type of testing the structure of comic styles involves a second-order factor analysis of individual differences in the use of comic styles, and somewhat different results may be expected due to the inclusion of the third dimension that represents a general factor (g-factor) of comic styles use (see **Figure 1B**). For example, while mock and humor are opposite as concepts in **Figure 1B** (i.e., implying a negative relationship), some individuals might engage in both and others in neither of them (i.e., suggesting even a positive relationship). This variance overlies the pattern of relations among the styles and alters the size (and potentially even sign) of the correlations. While there are people that clearly prefer mock over humor (and others that prefer humor over mock), this might happen at different levels of comic style use. Thus, by controlling the level, an initially perfect negative relation might turn into a slightly positive one. **Figure 1B** posits that the relations depicted in the rhombus (**Figure 1A**) only exist if the third dimension is kept constant (i.e., when individual differences do not occur or matter). Prior work with two different sets of preliminary markers for the eight comic styles (documented in Ruch, 2012) suggested that two or three second-order factors might be sufficient to represent the eight styles.

STUDY 1

Aims of Study 1

The overarching aim of Study 1 is to design and evaluate marker items for the eight comic styles (the Comic Style Markers, CSM) that can be used for both self- and other-reports, that represent the comic styles as identified in literary studies, and that allow measuring differences among individuals. In detail, this entails (a) confirming the item-level factor structure, (b) selecting suitable marker items (based on factor loadings and item statistics), (c) examining the reliability (internal consistency) and retest reliability of the CSM, (d) replicating the psychometric properties in a different language (English), (e) examining whether there is convergent and discriminant validity in self-other agreement, (f) determining socio-demographic correlates, and (g) examining the structure of the comic styles by looking at their intercorrelations, and vertical and hierarchical configurations (by means of hierarchical and ipsative second-order factor analysis).

Methods[2]

Participants

Overall, five samples were employed in Study 1 (see **Table 1**). Sample 1 was used to select the best items from the pilot version of the CSM for the final version. Sample 2 was employed to test whether the final item selection could be replicated in an independent sample. Sample 3 investigated the test-retest reliability of the CSM after 1–2 weeks. This sample partially overlaps with another study in which everyday humor behaviors and the HSQ were investigated (Heintz, 2017b). Sample 4 investigated the self-other agreement by having two close others rate the participants on an other-report form of the CSM. This sample partially overlaps with another study in which the construct validity of the HSQ was investigated (Heintz, 2017a). Sample 5 investigated the English version of the CSM.

Instruments

A pilot version of the CSM was generated, which was designed to mark the comic styles fun, humor, nonsense, wit, irony, satire, sarcasm, and cynicism (Schmidt-Hidding, 1963) as clearly as possible. The pilot version of the CSM comprised 73 marker items that depict the eight comic styles. Descriptions of the styles were compiled incorporating the elements discussed by Schmidt-Hidding (1963) and supplemented by other sources, such as descriptions of the comic styles in the literature, encyclopedias,

[2]Data and materials of Study 1 can be obtained from the corresponding author upon request.

TABLE 1 | Overview of the samples including basic descriptive statistics, measures, and analyses of Studies 1 and 2.

Samples	Gender *(M/F)*	Age *M* (*SD*)	Education	Nationality	Measures	Analyses
Study 1						
Sample 1			62.4% university-entrance diploma	69.3% German	CSM	Reliability, CITC
(*N* = 550)	26.7%/	25.04	29.3% university degree	14.5% Swiss	(pilot version)	Unidimensionality, EFA
	73.1%	(6.06)		11.3% Austrian		
Sample 2			41.7% university degree	44.0% German	CSM	Reliability, CITC
(*N* = 468)	33.3%/	37.15	38.0% university-entrance diploma	31.2% Austrian	(pilot version)	Unidimensionality, EFA
	65.6%	(15.11)		21.4% Swiss		
Sample 3			48.0% university students	65.5% Swiss	CSM	Test-retest reliability
(*N* = 148)	32.4%/	27.77	25.0% university-entrance diploma	25.7% German		CFA
	67.6%	(11.32)	23.0% university degree			
Sample 4			59.5% university students	83.3% Swiss	CSM	Self-other agreement
(*N* = 210)	26.7%/	26.35	17.6% university entrance diploma	10.5% German	CSM	CFA
	73.3%	(11.02)	16.2% university degree		(other- report form)	
Sample 5			–	38.6% American	CSM	Reliability, CITC
(*N* = 303)	33.0%/	29.77		30.7% British	(English adaptation)	Unidimensionality, CFA
	67.0%	(14.76)				
Study 2						
Sample 1			51.0% university-entrance diploma	57.3% German	CSM	Partial correlations and
(*N* = 999)	30.1%/	30.77	35.3% university degree	20.3% Austrian	MRS-25	multiple regressions
	69.9%	(12.81)		18.0% Swiss		
Sample 2	23.8%/	39.24	36.9% school or university students	51.8% Swiss	CSM	Partial correlations and
(*N* = 252)	76.2%	(16.52)		42.3% German	VIA-IS	multiple regressions
Sample 3			100% university students	–	CSM	Partial correlations and
(*N* = 214)	20.7%/	24.13			I-S-T 2000 R	multiple regressions
	79.3%	(3.88)			MSEI	

CSM, Comic Style Markers; MRS-25, Inventory of Minimal Redundant Scales; VIA-IS, VIA Inventory of Strengths; I-S-T 2000 R, Intelligence Structure Test 2000 Revised; MSEI, measure for self-estimated intelligence; CITC, corrected item-total correlation; EFA, exploratory factor analysis; CFA, confirmatory factor analysis.

dictionaries, and so on. Special care was taken that these elements could be related to individuals and eventually be transformed into corresponding items. This was achieved by studying definitions of the styles and transforming them into statements depicting everyday thoughts, feelings, and actions, while taking care of sticking to the definitions as purely as possible. A seven-point Likert format (1 = "strongly disagree" to 7 = " strongly agree") was utilized. There were between 6 and 13 marker items per comic style in the pilot version. Sample items are "I quickly read situations and can nail non-obvious matters to the point in a funny way" (wit) and "I accept the imperfection of human beings and my everyday life often gives me the opportunity to smile benevolently about it" (humor).

The revised version of the CSM includes 48 marker items, with six marker items for each comic style. The same seven-point Likert format is utilized. The items are listed in the Electronic Supplementary Material (Supplementary Table S1). For now, the main aim was to preserve the meaning of the styles and to be able to study the concepts. A final questionnaire to measure the comic styles comprehensively (e.g., by adding other relevant styles, facets of styles, and additional descriptions of the comic styles) will be developed at a later point in time.

The other-report version consists of the same 48 marker items as the CSM. The only difference is that the marker items were rephrased to capture other-reports. Specifically, pronouns and verb forms were adapted, and "I" was replaced by the participant's first name. It employs the same seven-point Likert scale. The English version of the CSM was adapted in a translation back-translation procedure. Inconsistencies were jointly resolved in a group discussion among the first, second, and third author of this paper.

Procedure

The five samples were collected online via www.surveymonkey.com (Samples 1, 2, and 5) or www.unipark.info (Samples 3 and 4). Other variables were collected that are not relevant for the present study. The study was conducted in compliance with the local ethical guidelines and participants provided online informed consent. In Sample 3, participants completed the final version of the CSM twice in a period of 1–2 weeks. In Sample 4, participants were provided with a link to an online survey including the other-reports of the CSM, which they forwarded to two close others.

Analyses

The rationally derived 73 items of the pilot version of the CSM (listed in the Supplementary Table S2) were subjected to three analyses to select the final items: Descriptive item analyses, corrected item-total correlations (CITC), and loadings on the first unrotated principal component (FUPC) to ensure the

unidimensionality of the scales. These CITC and FUPC analyses were refined in two rounds, as the item pool influences the outcomes of these analyses. The analyses were conducted in Sample 1 and then replicated in Sample 2.

To examine the factor structure of the revised version of the CSM, both exploratory and confirmatory factor analyses were conducted. Because these item-level analyses require large sample sizes, the exploratory and confirmatory factor analyses were conducted in pooled samples (Samples 3 + 4, and Samples 1 + 2, respectively). The exploratory factor analysis (EFA) was a principal axis factoring with oblimin rotation, as the factors were expected to be dependent (conducted with SPSS 20). The confirmatory factor analysis (CFA) was estimated with the *lavaan* package (Rosseel, 2012) in *R* (R Core Team, 2015). The MLR estimator was employed to yield robust standard errors, and the factors were allowed to correlate with each other. Fit indices were evaluated by the recommendations for acceptable fit of Schermelleh-Engel et al. (2003): comparative fit index (CFI) $\geq$ 0.95, root mean square error of approximation (RMSEA) $\leq$ 0.08 with a confidence interval close to the RMSEA, and standardized root mean square residual (SRMR) $\leq$ 0.10. For the CFI somewhat lower values were expected in the present analysis due to the large number of variables per factor, which can lead to low CFI values even if the model is correctly specified (see Kenny and McCoach, 2003).

To investigate test–retest reliability, the scores from the first assessment of the CSM were correlated with the scores from the second assessment (Sample 3). In Sample 4, self-other convergence (convergent validity) was tested by correlating the self-reports of the CSM with the aggregated other-reports (aggregated across two raters per participant).

Results

Identification of Markers: Reduction of Items

First, the descriptive statistics of the 73 marker items in the pilot version of the CSM were analyzed. The distribution of the items should approximate a normal distribution, so items that were skewed or kurtotic (values > |2|) were removed (2 items from the irony scale). Second, the CITC of the items were computed and compared to the correlations of the items with the other seven scales. Similarly, the FUPC of the items belonging to one scale was extracted in a principal component analysis, and the factor score was saved. Then the 73 marker items were correlated with each of the eight factor scores. These two steps should ensure that (a) each marker item relates to the scale/factor it belongs to, and (b) each marker item relates more strongly to the scale/factor it should belong to than to the other scales/factors. This procedure contributes to the reliability (internal consistency and unidimensionality) and factorial validity of the resulting scales. The marker items should have CITC of $\geq$0.30 (see Traub, 1994) and a loading on the FUPC of $\geq$0.40 (see Stevens, 2012). Also, the correlations of the marker item with the other scales should be at least 0.05 lower than the CITC, and the correlations of the marker items with the other factors should be at least 0.10 lower than the loading on the FUPC. Based on these criteria, 15 items were deleted (0–5 items from each scale), resulting in a second pilot pool of 56 items. In the second round, the remaining items were investigated with similar CITC and FUPC analyses. This resulted in an exclusion of 8 additional marker items (0–5 items from each scale), resulting in 48 marker items (six marker items per comic style). Importantly, the marker items that were excluded in Sample 1 were also those that showed the lowest CITC and loadings on the FUPC in Sample 2, replicating the selection of the revised 48 marker items (i.e., the CSM).

Reliability and Factor Structure of the Comic Styles

Table 2 shows the psychometric properties of the revised set of marker items of the CSM in the pooled construction and replication samples (Samples 1 and 2). As shown in **Table 2**, the psychometric properties supported the reliability of the eight scales. Internal consistencies ranged from 0.66 (humor) to 0.89 (cynicism), with most values being > 0.80. The CITCs ranged from 0.33–76, indicating that the marker items related to their scales, yet they were not redundant. Homogeneity (or unidimensionality) was supported in CFAs, indicated by high loadings on the latent factor (all > 0.40) and by mostly acceptable model fits, ranging from $\chi^2_{(9)} = 46–129$ ($ps < 0.001$), CFI = 0.92–0.96, RMSEA = 0.06–0.12 (90% confidence intervals [0.05–0.10, 0.08–0.13], and SRMR = 0.03–0.05. Supplementary

TABLE 2 | Descriptive statistics, reliability, factor structure, and test–retest correlations of the Comic Style Markers (CSM) in the German-speaking samples.

CSM	*M*[a]	*SD*[a]	α[a]	CITC[a]	Homogeneity[a]	EFA[b]	CFA[a]	r_{tt}[c]
Fun	4.37	1.16	0.86	0.57–0.70	0.62–0.77	0.51–0.75	0.62–0.81	0.88
Humor	5.01	0.83	0.66	0.33–0.49	0.41–0.66	0.20–0.67	0.45–0.58	0.74
Nonsense	4.93	1.09	0.85	0.58–0.71	0.62–0.79	0.37–0.85	0.65–0.78	0.89
Wit	4.80	1.06	0.87	0.61–0.71	0.50–0.81	0.56–0.80	0.67–0.77	0.89
Irony	4.46	1.17	0.82	0.47–0.72	0.65–0.77	0.37–0.71	0.52–0.82	0.86
Satire	4.22	1.02	0.75	0.41–0.58	0.47–0.70	0.31–0.64	0.47–0.67	0.78
Sarcasm	3.65	1.33	0.85	0.50–0.76	0.55–0.86	0.43–0.65	0.55–0.85	0.83
Cynicism	3.55	1.38	0.89	0.67–0.74	0.71–0.80	0.36–0.74	0.72–0.80	0.88

CITC, range of the corrected item-total correlations; Homogeneity, range of the loadings on the latent factor (separate for each comic style); EFA, exploratory factor analysis (principal axis factoring with oblimin rotation, loadings on the expected factors); CFA, confirmatory factor analysis (range of the loadings on the latent factors); r_{tt}, test–retest reliability across 1–2 weeks. [a]Results from pooled Samples 1 and 2 (N = 826–1018). [b]Results from pooled Samples 3 and 4 (N = 358). [c]Results from Sample 3 (N = 148).

Table S3 additionally shows the descriptive statistics of the CSM in all samples.

In the EFA, the eight factors explained 58.2% of the total variance (eigenvalues 1.18–10.93, rotated sums of squared loadings 3.63–5.62). While the scree test indicated the retention of either four or six factors, the parallel analysis suggested the retention of nine factors, and the revised minimum average partial test suggested the retention of seven factors. However, we decided to extract eight factors for the following reasons: (a) we theoretically expected eight factors, (b) the factor loadings were more clearly interpretable compared to the other solutions, (c) the communalities were mostly high (range = 0.13–0.70, $Mdn = 0.51$), (d) the items always loaded highly on their intended factors (ranging from 0.20 to 0.75; see **Table 2**), and (e) these loadings were always higher than the loadings on any of the other factors (maximum |0.50|). Only one humor item ("I am a realistic observer of human weaknesses, and my good-natured humor treats them benevolently") loaded negatively on sarcasm (−0.27) and positively on satire (0.27), which was slightly higher than the loading on humor (0.25).

The CFA model indicated a mostly acceptable fit: $\chi^2_{(1052)} = 3310$ ($p < 0.001$), CFI = 0.86, RMSEA = 0.05 (90% confidence interval [0.049,0.053], and SRMR = 0.07). Loadings were high for each factor, ranging from 0.45 to 0.85 (see **Table 2**). Finally, the test-retest reliability was high for all scales (0.74–0.89, $Mdn = 0.87$). As the time interval was rather short (1–2 weeks), this indicates at least short-term stability of the eight scales.

The intercorrelations of the eight scales ranged from essentially 0 to 0.67 (sarcasm and cynicism), with a median correlation of 0.37 (pooled Samples 1 + 2, $N = 1{,}018$). The zero correlations suggest that there will be no general factor in the field of the comic. However, it should be mentioned that the zero correlations all either involved sarcasm or cynicism and hence the other comic styles showed a positive manifold (i.e., only positive intercorrelations). The factor correlations were similar to the scale intercorrelations. In the factor analyses, the factor correlations were highest between sarcasm and cynicism (0.44 in the EFA and 0.81 in the CFA) with a median correlation of 0.21 (EFA) and 0.45 (CFA). As the CFA correlations were true-score correlations, this supports the notion that sarcasm and cynicism were similar, yet not interchangeable.

English Version of the Comic Style Markers

Table 3 shows the descriptive statistics, reliability, and factor structure of the CSM in the English-speaking sample. The reliabilities of the comic styles were also sufficient, ranging from 0.79 (irony) to 0.88 (wit and satire). The expected factor structure was supported in a CFA, which showed a mostly acceptable model fit: $\chi^2_{(1'052)} = 2'005$ ($p < 0.001$), CFI = 0.86, RMSEA = 0.06 (90% confidence interval [0.05,0.06], and SRMR = 0.07). Loadings were high for each factor, ranging from 0.49 to 0.85. Homogeneity of the factors was also supported, indicated by high loadings (>0.40) and by mostly acceptable model fits, ranging from $\chi^2(9) = 16$–49 (ps < 0.07), CFI = 0.92–0.99, RMSEA = 0.05–0.12 (90% confidence intervals [0.01–0.09,0.08–0.15], and SRMR = 0.03–0.05). Also Tucker's phi was computed, which indicates the factor congruence between the eight EFA factors from the German- and English-speaking samples. The nonsense scale could be considered equal across both languages, while fair similarity was obtained for fun, humor, and wit (and to some degree for irony, satire, and cynicism). A lack of similarity was only obtained for sarcasm. Tucker's phi at the item level indicated sufficient similarity for four of the six items (>0.84). Two items ("I am a sharp-tongued detractor" and "My laughter is occasionally derisive and expresses schadenfreude") showed lower loadings on the sarcasm factor and higher loadings on wit and satire, resulting in low convergence (0.64 and 0.15, respectively). Thus, similarity between the English- and the German-speaking samples was sufficient for all comic styles except for two sarcasm items.

The intercorrelations among the comic styles were slightly higher than in the German-speaking samples, ranging from small positive correlations to 0.74 (sarcasm and cynicism), with a median correlation of 0.49. In the CFA, the factor correlations were highest between sarcasm and cynicism (0.84) with a median correlation of 0.56.

Demographic Differences in the Comic Styles

Next, it is of interest whether the comic styles differed across several demographic variables. **Table 4** shows the correlations and analyses of covariance of the CSM with the demographic variables. Gender and age showed several

TABLE 3 | Descriptive statistics, reliability, and factor structure of the Comic Style Markers in the English-speaking sample.

Comic styles	*M*	*SD*	α	CITC	Homogeneity	CFA	Tucker's Phi
Fun	4.91	1.24	0.85	0.54–0.70	0.61–0.79	0.60–0.79	0.91
Humor	5.14	1.01	0.82	0.46–0.71	0.47–0.85	0.49–0.85	0.90
Nonsense	5.11	1.09	0.86	0.60–0.68	0.65–0.75	0.68–0.73	0.95
Wit	5.06	1.15	0.88	0.61–0.78	0.59–0.74	0.66–0.80	0.93
Irony	4.53	1.09	0.79	0.49–0.63	0.65–0.84	0.56–0.71	0.80
Satire	4.07	1.27	0.88	0.67–0.73	0.72–0.79	0.72–0.80	0.84
Sarcasm	3.91	1.34	0.87	0.56–0.76	0.58–0.85	0.61–0.82	0.70
Cynicism	3.94	1.24	0.84	0.55–0.69	0.59–0.79	0.65–0.75	0.83

N = 303. CITC, range of the corrected item-total correlations; Homogeneity, range of the loadings on the latent factor separate for each comic style; CFA, confirmatory factor analysis (range of the loadings on the latent factors).

TABLE 4 | Demographic differences in the Comic Style Markers.

	Correlations		*F*-values of ANCOVA (age and gender as covariates)			
Comic styles	**Gender**	**Age**	**Education[1]**	**Nation[2]**	**Family[3]**	**Housing[4]**
Fun	−0.08**	−0.07*	0.49	1.71	1.50	4.02**
Humor	−0.03	0.19***	2.01	1.38	1.30	3.88**
Nonsense	−0.09**	0.10**	1.20	4.61*	2.67*	1.81
Wit	−0.13***	0.08**	3.10*	1.88	1.16	1.98
Irony	−0.14***	−0.23***	2.54*	0.54	1.71	0.64
Satire	−0.20***	−0.05	0.59	0.32	1.26	1.88
Sarcasm	−0.17***	−0.13***	1.43	0.33	0.44	1.30
Cynicism	−0.27***	−0.12***	1.27	0.67	2.59	2.26*

*N = 1'013. ANCOVA, analysis of covariance. Gender was coded as 1 = male, 2 = female. Spearman's rank correlations were used for education and for age. *p < 0.05; **p < 0.01; ***p < 0.001. [1]Five education categories: <10 years (n = 15), apprenticeship (n = 93), university entrance diploma (n = 520), university degree (n = 354), and doctoral degree (n = 25). [2]Three nation categories: Germany (n = 579), Switzerland (n = 168), and Austria (n = 207). [3]Four family categories: single (n = 486), in a relationship (n = 310), married or registered relationship (n = 156), and divorced or widowed (n = 46). [4]Six housing categories: living alone (n = 274), living with partner (n = 225), living with partner and children (n = 100), living with children (n = 18), living in a shared apartment (n = 251), and living with parents/relatives (n = 135).*

meaningful correlations with the CSM. While most of them were small and significant due to the large sample size, a few noteworthy and theoretically expected relationships emerged. First, men tended to score higher in all comic styles than women (except for humor), with the strongest effects found for cynicism, satire, and sarcasm. This is in line with the more mocking and critical nature of these comic styles. Regarding age, humor and, to a lesser extent, nonsense tended to be shown more often by older than by younger people. Conversely, younger people engaged more often in irony, sarcasm, and cynicism than older people. When considering the love vs. hate dimension underlying the comic styles, the more love-related comic styles tended to increase with age, while the more hate-related ones tended to decrease with age.

The education level made a difference regarding wit ($\eta_p^2 = 0.012$) and irony ($\eta_p^2 = 0.010$). Follow-up pairwise comparisons (Bonferroni-corrected) showed that people who held a doctoral degree scored significantly higher on wit ($M = 5.48$, $SD = 0.70$) than those with an apprenticeship ($M = 4.64$, $SD = 1.16$; $p = 0.010$, $d = 0.78$), while no pairwise comparison was significant for irony. The three nations showed significant differences only in nonsense ($\eta_p^2 = 0.010$); that is, Austrians ($M = 5.15$, $SD = 0.99$) scored higher than Germans ($M = 4.82$, $SD = 1.13$; $p = 0.008$, $d = 0.30$). The family situation showed a significant difference in nonsense ($\eta_p^2 = 0.008$); that is, those in a relationship ($M = 4.94$, $SD = 1.07$) scored higher than those who were divorced or widowed ($M = 4.71$, $SD = 1.09$; $p = 0.037$, $d = 0.21$). The housing situation made a difference regarding fun ($\eta_p^2 = 0.020$), humor ($\eta_p^2 = 0.019$), and cynicism ($\eta_p^2 = 0.011$). Those who lived in a shared apartment scored significantly higher on fun ($M = 4.61$, $SD = 1.16$) and on humor ($M = 5.05$, $SD = 0.77$) than those living alone ($M = 4.22$, $SD = 1.21$, and $M = 4.89$, $SD = 0.95$; $p = 0.008$, $d = 0.33$, and $p = 0.004$, $d = 0.18$ respectively), while no pairwise comparison were significant for cynicism. Overall, a few meaningful, but small demographic differences emerged. The only large effect was found for wit, which was influenced by the level of education of the participants.

Construct Validity-I: Self-Other Convergence

Table 5 shows the convergent and discriminant correlations of the self- and other-reports of the CSM. The convergent correlations were large for each comic style (ranging from 0.44–0.56, $Mdn = 0.50$), supporting the convergent validity of the CSM. Importantly, the convergent correlations were always larger than the discriminant correlations (both regarding the median and maximum discriminant correlations). This also supports the discriminant validity of the CSM.

Construct Validity-II: Structure of the Comic Styles

Intercorrelations of the scales

Table 6 shows the intercorrelations among the eight comic styles. As in Samples 1 and 2, the correlations between humor and satire and cynicism were close to zero, and correlations were largest among sarcasm and cynicism. No negative correlations were found among any of the comic styles.

Second-order factor analyses

First, an analysis of the ipsative scores was conducted (principal components analysis). For each individual the mean across the eight styles was computed and subtracted from the eight scores. This way every individual had the same mean (but the standard

TABLE 5 | Convergent and discriminant correlations of self-reports and other-reports (averaged across two close others) of the Comic Style Markers.

		Discriminant correlations		
Comic styles	**Convergent**	**Median**	**Minimum**	**Maximum**
Fun	0.55***	0.17	0.04	0.33
Humor	0.44***	0.13	−0.12	0.27
Nonsense	0.44***	0.16	0.11	0.33
Wit	0.56***	0.19	0.08	0.36
Irony	0.49***	0.25	−0.04	0.37
Satire	0.40***	0.20	0.11	0.32
Sarcasm	0.50***	0.17	−0.12	0.40
Cynicism	0.52***	0.15	−0.12	0.40

*N = 210; ***p < 0.001.*

TABLE 6 | Intercorrelations among the eight Comic Style Markers.

Comic styles	Fun	Humor	Nonsense	Wit	Irony	Satire	Sarcasm
Humor	0.33***						
Nonsense	0.44***	0.38***					
Wit	0.40***	0.46***	0.28***				
Irony	0.31***	0.16***	0.24***	0.34***			
Satire	0.28***	0.39***	0.25***	0.36***	0.46***		
Sarcasm	0.19***	0.00	0.19***	0.18***	0.53***	0.46***	
Cynicism	0.15**	0.03	0.28***	0.18***	0.45***	0.53***	0.69***

*Samples 3 + 4 (N = 358); **p < 0.01; ***p < 0.001.*

deviations could vary). The scree test indicated two factors, which are displayed in **Figure 2**. The configuration was similar to **Figure 1A**, with a few peculiarities. Regarding the darker styles, sarcasm and cynicism were closely together and irony and satire are close to where wit was. On the lighter side, the arrangement of fun, humor and wit were as expected, yet nonsense was closer to fun, rather than between humor and wit. Finally, the main axis separated the lighter and darker styles, which is not as salient in **Figure 1A**, but still suggests that this was the major bipolar dimension in the comic styles.

Second, a PCA was performed on the normative data of the eight comic styles for Samples 3 and 4. The scree test suggested the extraction of two or three factors (first four eigenvalues: 3.49, 1.59, 0.77, 0.60, 0.50). Therefore all solutions between a FUPC and four oblique factors were studied. Thus, a hierarchical factor analysis (Goldberg, 2006) was conducted (see **Table 7**).

The eight styles all loaded on the first unrotated factor (explaining 43.6% of the variance) and then split up in the positively correlated light and dark styles at step two. While both factors were largely unipolar, there was a tendency for humor to load negatively on the dark styles factor and sarcasm to load negatively on the light styles factor. The three-factor solution had the lighter styles split up into two, and a four-factor solution was clearly an overextraction, as specific factors emerged. Thus, the three-factor solution (explaining 73.2% of the variance) was selected for interpretation. Factor 1 (tentatively labeled "mockery, ridicule") was highly loaded by sarcasm and cynicism as well as by satire (i.e., morally based ridicule) and irony (i.e., a technique that may be used for ridicule). The common element was that people mock and ridicule in a funny way. The second factor ("good humor") was primarily loaded by wit, humor and satire, but also to a lower degree by fun and irony. The commonality was that they are the more competent and even virtuous comic styles. The loading of satire was due to the moral goodness that is merged with mockery when ridicule is done to better a situation. The third factor ("enjoyment of humor") was primarily loaded by two scales, namely nonsense and fun, and slightly also by humor. This factor was definitely underdefined and needs more markers for a precise interpretation in the future.

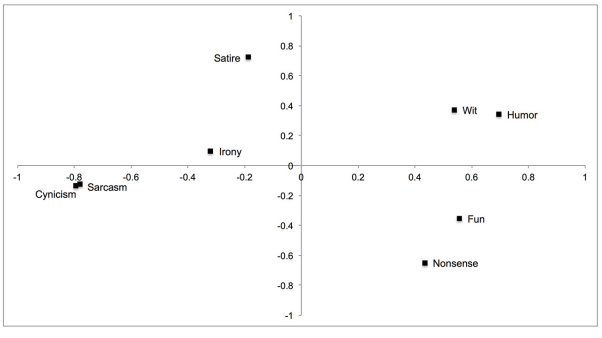


FIGURE 2 | Plot of factor loadings of the ipsative comic styles (based on principal components analyses).

Discussion

The main aim of Study 1 was accomplished, namely to design and validate a set of marker items that represent the eight comic styles based on the descriptions derived from literary studies (the CSM). The descriptions were useful for formulating marker items and the scales got refined in a first empirical analysis. Six marker items proved to be adequate to measure the styles with sufficient reliability (i.e., internal consistency, unidimensionality/homogeneity, and test–retest reliability). Regarding validity, the factorial validity of the items was established by CFAs; that is, the marker items measured the styles they were intended to measure. Furthermore, the self-other correspondence was sufficiently high, and it was even possible to distinguish between sarcasm and cynicism. The self-other correspondence with a median of 0.50 was much higher than for the earlier one-item measure (Ruch, 2012) and in the range typical for personality instruments. While most analyses were done with a German-speaking sample, the first testing of an English version proved successful too. Most importantly, discriminant validity was supported; that is, all styles (including cynicism and sarcasm) could be distinguished from each other. Thus, for now the CSM can be recommended for use in future studies. Once more styles are identified and once items from the experiential world of laypeople supplement these prototypes, a final instrument, the *Comic Styles Profiler*, will be introduced.

Thus, the eight styles were conceptually and empirically different. Nevertheless, some styles were more similar to each other than others, and when eliminating individual differences (i.e., a g-factor) through ipsatizing the scores, the proposed bipolarity of mockery styles (sarcasm, cynicism) and good-natured humor was verified. Furthermore, cynicism and sarcasm were close to each other with satire (as a moral

TABLE 7 | Factor pattern (oblimin rotation) of principal components analyses based on the intercorrelations among the eight comic styles.

		Two-factor solution		Three-factor solution		
Comic styles	**FUPC**	**Factor 1**	**Factor 2**	**Factor 1**	**Factor 2**	**Factor 3**
Fun	**0.58**	0.05	**0.70**	0.23	**0.44**	**0.79**
Humor	**0.49**	−0.19	**0.85**	0.05	**0.84**	**0.42**
Nonsense	**0.57**	0.11	**0.62**	0.26	0.28	**0.88**
Wit	**0.61**	0.07	**0.71**	0.27	**0.80**	0.36
Irony	**0.71**	**0.67**	0.18	**0.72**	0.35	0.30
Satire	**0.75**	**0.59**	0.33	**0.69**	**0.61**	0.20
Sarcasm	**0.68**	**0.92**	−0.13	**0.88**	0.08	0.20
Cynicism	**0.68**	**0.90**	−0.09	**0.87**	0.10	0.23
Correlation with F1			0.31		0.23	0.23
Correlation with F2						0.37

N = 358. FUPC, first unrotated principal component. Loadings > |0.40| in bold.

critique) and irony also having a higher proximity to wit. On the light side, the (vertical) order of wit, humor and fun was also found, with the exception that nonsense was not located between humor and wit, but close to fun. It remains to be studied whether the location of nonsense was due to emphasizing the fun element in the enjoyment of nonsense in the marker items or whether this was simply the more appropriate location not anticipated by the more intuitive model (depicted **Figure 1A**), which was not based on measurement.

However, when individual differences were allowed for, there was no strict opposition of light and dark styles (as some individuals might be high or low in both), but they rather defined the first unrotated factor together. They then tended to fall into a lighter and darker cluster, and the former ones fell into a shallower (non-serious cheerfulness) and a more profound (resourceful) subgroup. All these factors intercorrelated uniformly positive, suggesting that only one factor (i.e., the g-factor as depicted in **Figure 1B**) was needed to account for the intercorrelations. Such a hierarchical model (i.e., entailing eight lower order styles, three style factors, and a general factor) is possible; however, it will need to be built on more variables helping to identify the factors more clearly (see also Supplementary Figure S1 for a schematic representation of this model). Mockery (or "laughing at") is a factor that emerged in the present study as well as in previous studies with preliminary measures (Ruch, 2012). Mockery combines all dark styles, with cynicism and sarcasm being at its core, and satire and irony having high but not pure loadings. Schmidt-Hidding (1963) suggested that the use of these styles implies having malicious, mean-spirited goals and attitudes, intentions of hurting other people and demonstrating superiority. Therefore, using this set of comic styles will hurt or upset others. Still, there are nuances in this factor, and future research needs to study these styles further and also examine their relation to katagelasticism (i.e., the joy of laughing at others; Ruch et al., 2014) and to the aggressive (Martin et al., 2003) and mean-spirited (Craik et al., 1996) humor styles (see Ruch and Heintz, 2016a, for a preliminary investigation of the overlap of these different conceptualizations of humor styles).

This was different for the light styles, which have different goals, but typically go along with positive affect. They came in two clusters, a more basic enjoyment of humor factor and a more profound good humor factor. The former can be seen as enjoyment of the non-seriousness in communication and social interaction; it is more socio-affective and refers indulging in playing pranks, clowning around, good-natured kidding, brightening others up, indulging in gibberish talk, and playing with meaning, sense and nonsense. This factor will be more similar to the socially warm and boorish humor styles (Craik et al., 1996) and the hilarity component of cheerfulness, namely the facets of low threshold for smiling and laughter, a broad range of active elicitors of cheerfulness and smiling/laughter, and a generally cheerful interaction style (Ruch et al., 1996). Also, this factor is expected to predict enjoyment of various forms of humor stimuli, including non-sophisticated forms and low comedy. The good humor factor is marked by the more profound styles of humor, wit, and also satire, which, taken together, entail more cognitive efforts (i.e., mindfully observing incongruities in daily lives), resilience when facing adversity (ability to see the funny side in adversities or short-comings), and a general aiming at the good. It is more related to the cheerful composedness of trait cheerfulness (Ruch et al., 1996), the self-enhancing humor style (Martin et al., 2003), and the reflective and benign humor styles (Craik et al., 1996), without being identical with any of these. There are more resources needed for this style, like mindfully detecting the incongruities in life, the capacity to describe them, and the relaxedness to deal with them in a lighthearted way. While both forms of light styles represent cheerfulness, the former might be also related to low seriousness and the latter might be related to a robustness of mood (i.e., low bad mood) and character strengths.

STUDY 2: CONSTRUCT VALIDITY

Comic styles are trait-like, either as typical behavior (i.e., temperament/personality), maximal performance (i.e., ability), or, more recently, morally valued traits (i.e., character). Humor instruments have been studied mostly in relation to personality

(as represented, for example, by the five-factor model, FFM; see McCrae and Costa, 2013) for a long while now (see Ruch, 2008, for an overview of studies). Liking to laugh, entertaining others, telling jokes, and experiencing positive emotions are part of components of extraversion, and hence we expect the light styles to be correlated positively with extraversion. Neuroticism represents a disposition to negative emotions and worry, and thus we expect it to relate negatively to humor in face of adversity (i.e., being able to laugh at oneself, to cope with stress) and humor performance (e.g., wit). Agreeableness vs. antagonism determines the tone toward others; that is, cooperative and friendly vs. critical or hostile. In line with this, we expect agreeableness to be negatively related to the dark styles (especially sarcasm and cynicism) and positively related to the benevolent treatment of shortcomings (i.e., humor). Openness to experience (or culture, intellect) provides the capacity for generating humor, and we expect it to correlate with wit, but also the other styles involving a production of humor. Conscientiousness refers to components like order, dutifulness, and self-deliberation, but also low spontaneity. Hence it is difficult to imagine conscientiousness being positively related to any particular style, and we do not make specific predictions for conscientiousness. In sum, to the extent that positive and low negative emotions, imagination and friendliness vs. antagonism are involved in a comic style, we expect it to show a correlation with extraversion, emotional stability, culture/openness to experience, and agreeableness, respectively.

Ability is maximal performance, and in humor (in the broad sense), there are a few components that represent ability in processing, creating, and delivering humor. For example, producing funny punch lines on the spot will require verbal ability. Perceiving incongruities and combining them in a witty statement requires mental capacities as well. Hence, it is not surprising that humor and ability were rarely studied together except for wit or humor production. Nevertheless, there are studies showing that people of higher intelligence displayed a higher appreciation of nonsense (Terry and Ertel, 1974; Wierzbicki and Young, 1978; Hehl and Ruch, 1985). Taken together, we expect that mostly wit has some relation with intelligence, and of the different components of intelligence it is verbal (but not numerical or figural) intelligence that displays the highest coefficients.

More recently, character has been introduced to the study of personality through the postulate of character being composed of virtues, character strengths, and situational themes (Peterson and Seligman, 2004). Virtues are seen as the core characteristics valued by moral philosophers and religious thinkers, and character strengths are the psychological ingredients—processes or mechanisms—that define the virtues, or distinguishable routes to displaying one or another of the virtues. Factor analyses of the strengths often reveal five factors, namely emotional, interpersonal, intellectual and theological strengths, as well as strengths of restraint (Peterson and Seligman, 2004).

We expect several links between the comic styles and this model of character. First, what is laughable and what is not laughable has historically been shaped by virtues (Ruch, 1998) and hence positive (as well as negative) correlations are expected between some comic styles and virtues. In detail, strengths molded by humanity may relate positively to humor (and negatively to the mockery styles), the cognitive strengths defining wisdom and knowledge might relate to wit, and temperance (e.g., prudence, self regulation) and transcendence (e.g., gratitude, spirituality) strengths suggest lower engagement in mockery.

Second, as humor is one of the 24 strengths, it is interesting to see what contents went into the definition and how the scale relates to the eight comic styles. For Aristotle (Chase, 1890), the virtuous form of humor (i.e., the ready-witted form) is to joke and amuse without hurting. For Aristotle, "ready wit" is moderation in the desire to amuse others, and the excess desire is buffoonery (amuse others too often, striving for laughter at all costs, laughing excessively, relentless mockery), and the deficient desire is boorishness (e.g., not getting involved in joking at all, feeling negatively about it). Similarly, Peterson and Seligman (2004, p. 530) define the humorous person as someone "skilled at laughing and teasing, at bringing smiles to the faces of others, at seeing the light side, and at making (not necessarily telling) jokes." They note that in the domain of humor, some forms are mean (e.g., mockery, ridicule, sarcasm) or on the border (e.g., parody, practical jokes), and they only include forms that "serve some moral good—by making the human condition more bearable by drawing attention to its contradictions, by sustaining good cheer in the face of despair, by building social bonds, and by lubricating social interaction" (Peterson and Seligman, 2004, p. 530). This suggests stronger overlaps between humor as character strength and fun, humor and wit, and zero conceptual overlap with sarcasm and cynicism. Additionally, satire forms something morally good (i.e., correcting wrongdoings with the aim to better society or people) but also involves some criticism, which might result in lower correlations. Thus, we expect satire and the mockery styles (sarcasm, cynicism) to differentially relate to the strengths and virtues; when controlling for the mockery element (by partialling out sarcasm and cynicism), we anticipate satire to more strongly positively relate to strengths and virtues. Low correlations are expected for irony, which may involve criticizing through a compliment or state something positive through negative words. Taken together, we anticipate the correlations between the comic styles and humor as character strength to differ between highly positive to virtually zero.

Study 2 aims at extending the construct validity investigations of the CSM. Construct validity is examined by studying the relation between comic styles and more general traits of personality, character, and ability. Special attention will be given to examine humor as character strength.

Methods[3]

Participants

An overview of the samples of Study 2 is given in **Table 1**. Sample 1 consisted of subsets of Samples 1 and 2 from Study 1 that also completed a personality measure. Sample 2 completed the CSM and a measure of character strengths. This sample overlaps with another study in which the comic styles and

[3]Data and materials of Study 2 can be obtained from the corresponding author upon request.

subjective well-being were investigated (Ruch et al., 2018). Sample 3 completed the CSM and measures of intelligence (self-reported and psychometrically tested).

Instruments

The Inventory of Minimal Redundant Scales (*MRS-25*; Schallberger and Venetz, 1999) lists 25 pairs of bipolar adjectives for the assessment of the Big Five personality dimensions *extraversion*, *agreeableness*, *conscientiousness*, *emotional stability*, and *culture*. The answers format is a bipolar six-point scale. In the present sample, internal consistencies were satisfactory, ranging from $\alpha = 0.76$ (agreeableness and culture) to 0.87 (conscientiousness).

The *VIA Inventory of Strengths* (VIA-IS; Peterson et al., 2005; German adaptation by Ruch et al., 2010) is a 240-item questionnaire for the assessment of the 24 character strengths (10 items per strength) covered by the VIA classification (Peterson and Seligman, 2004). It employs a 5-point Likert-style scale ranging from 1 ("very much unlike me") to 5 ("very much like me"). A sample item is "I never quit a task before it is done" (persistence). Internal consistencies in the present sample ranged from $\alpha = 0.68$ (honesty) to 0.91 (religiousness) with a median of 0.78. To obtain aggregated scores a principal component analysis of the VIA-IS scales with subsequent varimax rotation of five factors was conducted. The five-factor solution closely resembled the solution reported in Ruch et al. (2010), with Tucker's Phi coefficients being 0.95 (emotional), 0.90 (interpersonal), 0.88 (intellectual), 0.98 (theological), and 0.95 (restraint).

The *Intelligence Structure Test Revised* (I-S-T 2000 R; Amthauer et al., 2001) consists of nine subtests. It allows the assessment of fluid as well as crystallized intelligence. For the present study, subtests for verbal (analogies), numerical (arithmetical tasks), and spatial (cube tasks) intelligence were used. The tests for verbal, numeric, and spatial intelligence were taken together as a total score of intelligence. All tests are speed tests; i.e., the administration was timed. The I-S-T 2000 R is widely used and well established in the German-speaking countries. Norm scores were computed according to German age and gender norms ($M = 100$, $SD = 10$).

Measure for self-estimated intelligence (MSEI; Proyer and Ruch, 2009). Participants had to rate their ability on a line from "low" to "high ability" for the domains of verbal, numeric, and spatial intelligence. Each position on the scale ranging from lowest to highest self-estimated ability may be marked (on a scale from 0 "lowest ability" to 100 "highest ability"). A total score was computed from all self-estimations as a general self-estimated ability score. The single dimensions were explained by a short sentence [e.g., "verbal: Dealing with language and words (e.g., eloquence)"].

Procedure

For Samples 1 and 2, all questionnaires were presented online. Participants were recruited via different channels; for example, social media, mailing lists, or newspaper articles containing the link to the respective study. All participants provided consent and participated voluntarily. Sample 3 consisted of students at the University of Zurich who were attending a lecture on psychological assessment and the data was collected in paper–pencil format for this study. All studies were performed in accordance with the local ethical guidelines and online or written informed consent was supplied.

Analyses

To assess the overlap between the criteria and the comic styles, correlations were computed. Since the measures showed small but consistent correlations with age and gender, these variables were controlled for in partial correlations (Supplementary Table S4 also shows the zero-order correlations as well as the descriptive statistics of all measures). Furthermore, standard multiple regressions were computed to assess how much variance could be explained in total in each of the comic styles and in each of the criteria.

Results

Personality

The correlations between the FFM traits (MRS-25) and the eight comic styles were computed and are presented in **Table 8** (controlling for age and gender). Multiple correlations with the FFM traits as criteria showed that the variance in extraversion, agreeableness, culture and emotional stability was well explained by comic styles, while conscientiousness had a significant, but low contribution. Likewise, multiple correlations computed for the comic styles as criteria showed that wit and humor were most potently predicted, followed by fun, sarcasm, and cynicism, and eventually nonsense, satire, and irony.

Table 8 shows that each of the comic styles had a unique pattern of correlations with personality. In more detail, the prime correlation of sarcasm and cynicism was low agreeableness and

TABLE 8 | Partial correlations and multiple regressions of the comic styles with the Inventory of Minimal Redundant Scales (controlled for age and gender).

Personality traits	Fun	Humor	Non.	Wit	Irony	Satire	Sarc.	Cyn.	*R*/Adj. R^2
Extraversion	0.42***	0.31***	0.14***	0.37***	0.06*	0.16***	−0.04	−0.09**	0.49/0.23
Agreeableness	0.04	0.25***	0.05	0.00	−0.17***	−0.11***	−0.41***	−0.32***	0.49/0.24
Conscientiousness	−0.13***	−0.09**	−0.15***	−0.08*	−0.04	−0.11***	−0.10**	−0.13***	0.21/0.04
Emotional stability	0.19***	0.35***	0.13***	0.30***	0.03	0.09*	−0.10*	−0.02	0.40/0.16
Culture	0.24***	0.34***	0.28***	0.40***	0.08*	0.21***	−0.03	0.04	0.47/0.22
R	0.44	0.51	0.33	0.53	0.21	0.29	0.42	0.36	
Adj. R^2	0.19	0.25	0.11	0.27	0.04	0.09	0.17	0.13	

*N = 999. Non., nonsense; Sarc., sarcasm; Cyn., cynicism; Adj., adjusted. *p < 0.05; **p < 0.01; ***p < 0.001.*

to a smaller extent low conscientiousness. Additionally, cynics tended to be introverted and sarcastic individuals tended to be emotionally instable. Satire and humor both correlated with all FFM traits, but there were also differences. While satire (like the two styles of mockery and irony) correlated negatively with agreeableness, humor yielded a positive correlation. Furthermore, the positive correlations with extraversion, culture, and emotional stability were numerically higher for humor than for satire. Wit, fun and nonsense shared the predictors of humor, except for agreeableness; however, regarding fun, extraversion was the best predictor, and regarding both wit and nonsense, culture yielded the highest coefficients.

Interestingly, in this sample humor was uncorrelated with both cynicism and sarcasm, and hence it is unlikely that a third variable correlated with them with a different sign (as agreeableness did with humor and the mockery styles). To establish bipolarity between the love vs. hate comic styles (see **Figure 1A**), an index was computed by subtracting the average of sarcasm and cynicism from humor. This index correlated positively with agreeableness ($r = 0.48$), thus suggesting that agreeableness vs. antagonism as a personality dimension was aligned with humor-related benevolence vs. mockery. Removing individual differences enhanced the correlations that were found for the individual styles. Other high multiple correlations also showed bipolarity in the predictors, however, again with a predominance of one side, namely extraversion (cynics were introverted) and emotional stability (sarcasm was on the neuroticism side).

Character Strengths

The correlations between the five strengths factors (VIA-IS) and the eight comic styles (controlled for age and gender) were computed and are presented in **Table 9**. The correlations between the 24 characters strengths and the eight comic styles (controlled for age and gender, as well) are depicted in Supplementary Table S5.

Table 9 shows that each of the strengths factors was involved in the prediction of comic styles and each comic style was predicted by character strengths, and there were positive as well as negative coefficients. In detail, emotional strengths (loaded by zest, hope, bravery, but also humor) predicted fun, wit, and humor well, were still significant for nonsense, irony, satire, and sarcasm, and were uncorrelated with cynicism. Interpersonal strengths (loaded by fairness, teamwork, kindness, leadership, forgiveness) predicted fun positively and sarcasm, cynicism, and irony negatively. Strengths of restraint (loaded by prudence, humility, self-regulation, persistence) tended to go along with low scores in most comic styles, in particular with fun. Intellectual strengths (loaded by love of learning, creativity, open-mindedness but also appreciation of beauty and excellence) strongly predicted wit, and (albeit less strongly) other comic styles with a cognitive emphasis, namely, nonsense, humor and irony. Only fun was uncorrelated with intellectual strengths. Theological strengths (loaded by religiousness, gratitude, and appreciation of beauty and excellence) positively predicted fun and correlated negatively with sarcasm and cynicism.

Most interestingly, Supplementary Table S5 shows that the VIA-IS humor scale correlated significantly positively with every comic style, but to a different extent. Fun ($r = 0.63$), wit ($r = 0.61$), and humor ($r = 0.58$), had high coefficients, followed by nonsense ($r = 0.38$), satire ($r = 0.39$), and irony ($r = 0.33$, all ps < 0.001). Sarcasm ($r = 0.15$) and cynicism ($r = 0.13$, $p < 0.05$) had small but significant zero-order correlations. However, a multiple regression analysis predicting the VIA-IS humor scale yielded significant and positive beta weights only for fun ($\beta = 0.41$, $p < 0.001$), wit ($\beta = 0.30$, $p < 001$), humor ($\beta = 0.25$, $p < 001$) and satire ($\beta = 0.14$, $p = 0.027$), and negative ones for sarcasm and cynicism ($\beta = -0.17$, $p = 0.011$ each).

Next the assumption was tested that in satire two elements blend, namely mockery of someone combined with a good intention. Satire, or corrective humor, is not decrying something foolish or immoral for malicious pleasure, but for changing things to the better, leaving the good relationship intact. To highlight the good character element in satire, partial correlations were computed between the five strengths factors and satire, controlling for age and gender, but also for sarcasm and cynicism. Satire, bereft of the critical tone, was exclusively positively related to character, namely emotional strengths ($r = 0.24$), interpersonal strengths ($r = 0.23$), intellectual strengths ($r = 0.26$), and theological strengths ($r = 0.21$, all ps < 0.001). The factor describing strengths of restraint ($r = -0.06$) had no significant correlation, as it was a constant in all styles. Altogether 18 of the 24 strengths yielded significant positive correlations, underscoring the involvement of good character in corrective humor (see Supplementary Table S5).

TABLE 9 | Partial correlations and multiple regressions between the character strengths factors (derived from the VIA-Inventory of Strengths) and the Comic Style Markers (controlled for age and gender).

Strength factors	Fun	Humor	Non.	Wit	Irony	Satire	Sarc.	Cyn.	*R*/adj. R^2
Emotional	0.41***	0.40***	0.25***	0.47***	0.24***	0.23***	0.17**	0.04	0.56/0.29
Interpersonal	0.13*	0.12	−0.02	0.02	−0.18**	−0.07	−0.26***	−0.27***	0.43/0.16
Restraint	−0.26***	−0.11	−0.18**	−0.14*	−0.04	−0.13*	−0.15*	−0.13*	0.34/0.09
Intellectual	0.10	0.22***	0.27***	0.46***	0.23***	0.27***	0.13*	0.23***	0.55/0.30
Theological	0.16*	0.08	0.08	−0.01	−0.11	−0.02	−0.28***	−0.21**	0.46/0.18
R	0.53	0.49	0.44	0.70	0.48	0.44	0.56	0.53	
Adj. R^2	0.27	0.23	0.18	0.47	0.21	0.18	0.30	0.27	

*N = 252. Non., nonsense; Sarc., sarcasm; Cyn., cynicism; adj., adjusted. *p < 0.05; **p < 0.01; ***p < 0.001.*

Ability

The correlations between intelligence (based on a test and self-reports) and the eight comic styles (controlled for age and gender) were computed and are presented in **Table 10**.

As expected, the correlation between measured verbal intelligence and wit was indeed positive and significant, and it was the only significant correlation. The coefficient was rather low, but considering that ability and personality are rarely related, the size of the coefficient was as expected.

Interestingly, and not surprisingly, there were more (and higher) correlations for self-rated intelligence. First, the correlation of self-rated verbal intelligence and wit was much higher, underscoring that using the same method (self-reports) yielded higher relationships than divergent methods (self-reports, test). However, there were also positive correlations with humor, albeit smaller than for wit. The total intelligence score correlated most strongly with wit, followed by humor. Additionally, there were also small correlations with nonsense, irony, and satire. Thus, in five of the styles, people who scored higher also assumed that they were higher in several of the intelligence scales (and the total score). The differences between measured and rated intelligence are plausible but also striking, underscoring that there is method variance involved. However, it is also clear that the use of wit is also based on a higher (measured) verbal intelligence.

Discussion

Comic styles tap differently into personality, ability and character, and each of the comic styles had a unique set of predictors, underscoring the necessity to be separated. A person's involvement in the ludicrous is an expression of one's personality traits (including the valued traits) and selectively also verbal intelligence. Wit requires an astute mind that allows to quickly read situations and nailing non-obvious matters to the point in a funny way. Obviously, ability in the verbal (rather than numerical or figural) domain provides the link between measured intelligence and wit. Moreover, self-rated verbal ability, but also creativity (VIA-IS scale), cognitive strengths (as a strengths factor), and culture/intellect (MRS-25) assume the position to be especially predictive of wit. Future studies will need to show whether these predictors overlap, and whether performance measures of wit (e.g., being able to write witty punch lines to caption-removed cartoons) is predictive of wit in the present instrument. Thus, this component of humor (in the broad sense) can indeed be seen as also drawing on individual differences in ability, a domain neglected in humor research.

Character, the moral subdomain of personality, was demonstrated to be relevant as well, and the use of fine-grained measures of both character and humor allowed for a more comprehensive investigation. For once, the use of both sarcasm and cynicism was regulated by theological (e.g., gratitude, religiousness) and interpersonal (e.g., fairness, forgiveness) strengths (see also Beermann and Ruch, 2009; Müller and Ruch, 2011). These components of the good character counteracted a frequent expression of mockery, while cognitive strengths and emotional strengths (sarcasm only) favored it. Character was also involved in satire. While sarcasm and cynicism may be fueled by the joy of mockery without the involvement of a moral sense, in satire there are good intentions of correcting misdoings. Focusing on the motivation for corrections (i.e., removing mockery), the moral sense was revealed in nearly all factors of strengths (except restraint). Thus, the comparatively lower zero-order correlations for satire were the product of negative (i.e., pointing out flaws in others) and positive (i.e., moral justification for the criticism) tendencies. Wit and humor were most highly correlated with humor as character strength, and these comic styles were well predicted by the individual character strengths and by the strength factors. Humor had only positive character correlates both at the level of the individual strengths and the strength factors and was hence the best indicator of good character. Wit and fun had overwhelmingly positive correlates with character, but also a negative correlation with the strengths of restraint factor, which was based on individual strengths, namely, prudence (fun) and humility (wit). Agreeableness, just as emotional strengths (primarily loaded by strengths of humanity and courage), was also indicative of humor and negatively related to mockery. As there was no simultaneous assessment of personality and character, it cannot be decided whether and where character provides incremental predictions of the virtuous comic styles over personality.

TABLE 10 | Partial correlations of the Comic Style Markers with self-rated and measured intelligence (controlled for age and gender).

Intelligence	Fun	Humor	Nonsense	Wit	Irony	Satire	Sarcasm	Cynicism
Measured								
Verbal	0.06	0.08	0.05	0.19**	0.09	−0.04	0.08	−0.03
Numerical	0.01	0.11	−0.07	−0.03	0.08	0.09	−0.02	0.02
Spatial	−0.07	−0.03	0.01	−0.08	0.08	−0.01	−0.07	0.06
Total	−0.03	0.06	−0.02	0.04	0.12	0.06	0.00	0.03
Self-rated								
Verbal	0.10	0.28***	0.08	0.48***	0.01	0.12	0.04	−0.03
Numerical	−0.01	0.08	0.04	0.01	0.13	0.04	0.04	0.09
Spatial	−0.02	0.05	0.07	0.05	0.17*	0.06	−0.09	0.08
Total	0.04	0.21**	0.13*	0.28***	0.17*	0.16*	0.02	0.14*

*N = 214 (self-rated intelligence, assessed with the Measure for self-estimated intelligence) and N = 199 (measured intelligence, assessed with the Intelligence Structure Test 2000 Revised). *p < 0.05; **p < 0.01; ***p < 0.001.*

There were distinct correlations of personality with comic styles, which fit well to predictions. Some strengths underlay all styles, some either the light or the dark, and other strengths additionally underlay some specific styles. Conscientiousness (like strengths of restraint) tended to yield low negative correlations with all styles, suggesting that this is a minor constant in engaging in humor at all. Extraversion, emotional stability, and culture correlated higher with the light styles than with the others. While low agreeableness was related to the dark styles (most strongly for sarcasm and cynicism) and negative affect related to sarcasm and cynicism, both were also linked inversely to humor, suggesting that these traits were sensitive to the motivational difference between laughing at and laughing with. Other traits also had humor on both sides of the dimension, namely extraversion (cynics were introverted), emotional stability (low in sarcasm) and antagonism (humor was agreeable). There was no style that involved low culture/openness.

Thus, in sum, the results provide indirect and at least partial support for the assumption that the comic styles reflect different domains of human functioning, with fun, humor, wit, and mock/ridicule reflecting forces of vitality/high spirits, a sympathetic heart, a superior spirit, and moral sense or haughtiness/maliciousness, respectively (depicted in **Figure 1A**). Specifically, cognitive strengths and verbal intelligence indicate a "superior mind", agreeableness, emotional strengths and humanity reflect a "sympathetic heart," zest and extraversion represent "vitality," and low agreeableness represents "haughtiness." There was no direct predictor for moral sense, and hence satire remains without a direct potent predictor. While these results confirm the lay psychologist view on humor and personality, future studies will emphasize the contemporary models of personality, character, and ability.

OVERALL DISCUSSION

Overall, the present studies represent a starting point for research defining and measuring more narrow styles of humor (in the broad sense), as they have been discussed for a long time in the literature, but have not yet been utilized in psychology. These styles represent a broad variety of humor and tap into personality and character as well as ability. These styles will be more easily "used," trained, and modified than the existing ones in the literature. The overlap in the scales allows aggregating the styles to more general styles, which, in turn, potentially might form a general factor of humor. This analysis, again, might be the basis for forming a hierarchical model with three levels that needs to be completed and evaluated in future studies. Individual differences in humor might be described by the general level (concerning the overall humor potential of a person), by a profile in aggregated styles (informing about engagement in specific domains of humor), and by a profile in specific styles (that describes differences more fine-grained and is closest to behavior). These levels will be useful for different types of studies. For example, humor trainings best address the lower level; relations to health or work-related variables might be most parsimoniously studied at a midlevel (where the discovery of more general patterns will be sufficient). Again, if economy is important, the overall humor potential might be sufficient. However, the list of styles is not yet exhaustive and hence more research is needed in order to build a comprehensive model. This allows showing whether the assumption of a general factor is tenable or not. In particular, for a more complete description of the domain of humor, components of the ineptness in humor use or forms of humorlessness are needed.

As the present studies replicated the results found with prior markers (Ruch, 2012), the validity of the three-factor structure was substantiated. However, although the factors explained 70% of the variance, this is still lower than the reliability of the scales, and thus the scales had unique variance that gets lost when analyzing the factors only. Hence, the major level of analyses should still be the level of styles. Further research will show what the unique contributions of the individual styles are and where aggregation is meaningful. In an EEG study of 52 participants, potential brain mechanisms underlying different types of humor were investigated (Papousek et al., 2017). It provided evidence for the unique status of humor among the light styles, and the overlapping effects of sarcasm, cynicism and irony among the mocking comic styles. Specifically, phasic changes in the functional coupling of prefrontal and posterior cortex (EEG coherence) during other people's auditory displays of happy (i.e., laughter) and sad mood (i.e., crying) were recorded and related to comic styles. The results support the view that typical comic styles develop in accordance with the rewarding values of their implicit outcomes (e.g., interaction partners are joyful or upset), which in turn reflect the individuals' interpersonal goals. While there are four light comic styles, the results underscored that they were heterogeneous and that there was indeed only humor that had the "laughing with" quality. As in the structural analyses, the dark styles were more homogeneous, yet satire (i.e., the only dark style where the hurting aim might be diluted by the positive intentions in corrective humor) acts differently. Other studies also provided preliminary validation; for example, Ruch et al. (2018) provided evidence that the styles had different relations to well-being (e.g., wit, humor and fun correlated positively with life satisfaction, while cynicism correlated negatively). Further validation studies of the CSM will add knowledge to the uniqueness and common core of the different styles.

Limitations

One obvious limitation is that in a first step the measurement of the styles was restricted, and there are other comic styles that could be considered for inclusion, such as black (gallows, sick) humor or absurd humor. These should be identified and examined in future studies to see if they entail elements that are not yet covered by the CSM. Furthermore, this approach will not lead to the description of the various forms of ineptness in humor, as these are typically not described in the literature. Hence, this is not a complete model of humor, and it needs to be supplemented by forms of humorlessness (e.g., Ruch and Hofmann, 2012;

Ruch et al., 2014). Second, as the concepts of interest are complex by nature, they also require several elements to be present in the items to capture them adequately (e.g., "I am a realistic observer of human weaknesses, and my good-natured humor treats them benevolently"). Still, future research could develop different assessment strategies to tease apart these elements in separate items and test if they also mark the eight comic styles. Third, many of the findings were based on self-reports, and future studies should employ several assessment methods. Fourth, the samples employed were mostly well-educated, and thus replications with samples with a more varied educational background are needed.

CONCLUSION

In two studies, we presented and tested the Comic Style Markers (CSM), a set of 48 marker items that represent individual differences in eight comic styles: Fun, humor, nonsense, wit, irony, satire, sarcasm, and cynicism. Both studies supported the construct validity of the CSM. Specifically, the eight comic styles were shown to be theoretically and empirically distinguishable and to relate to different outcomes (personality, character strengths, and intelligence). The CSM thus provides a starting point for more fine-grained investigations of humor-related styles, ultimately aiming at identifying a comprehensive list of narrow and specific comic styles that can be enacted, trained, and modified.

AUTHOR CONTRIBUTIONS

WR initiated the project and designed the concepts, all authors collected the data, SH, LW, and WR analyzed the data. All authors contributed to the writing of the manuscript, read it critically and gave consent to its publication.

ACKNOWLEDGMENTS

The authors thank Jessica Milner Davis and Alexander Stahlmann for commenting on an earlier version of this manuscript. The authors would also like to thank Claudia Hürzeler, Alex Junghans, Hildegard Marxer, and Jan Steiner for their help in collecting the data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00006/full#supplementary-material

REFERENCES

Amthauer, R., Brocke, B., Liepmann, D., and Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R [Intelligence Structure Test 2000 R]*. Göttingen: Hogrefe.

Attardo, S. (ed.). (2014). *Encyclopedia of Humor Studies*. Thousand Oaks, CA: Sage. doi: 10.4135/9781483346175

Beermann, U., and Ruch, W. (2009). How virtuous is humor? What we can learn from current instruments. *J. Posit. Psychol.* 4, 528–539. doi: 10.1080/17439760903262859

Chase, D. P. (1890). *The Ethics of Aristotle: The Nichomachean Ethics*. London: W. Scott.

Craik, K. H., Lampert, M. D., and Nelson, A. J. (1996). Sense of humor and styles of everyday humorous conduct. *Humor* 9, 273–302. doi: 10.1515/humr.1996.9.3-4.273

Goldberg, L. R. (2006). Doing it all bass-ackwards: the development of hierarchical factor structures from the top down. *J. Res. Pers.* 40, 347–358. doi: 10.1016/j.jrp.2006.01.001

Hehl, F. J., and Ruch, W. (1985). The location of sense of humor within comprehensive personality spaces: an exploratory study. *Pers. Individ. Dif.* 6, 703–715. doi: 10.1016/0191-8869(85)90081-9

Heintz, S. (2017a). Do others judge my humor style as I do? Self-other agreement and construct validity of the Humor Styles Questionnaire. *Eur. J. Psychol. Assess.* doi: 10.1027/1015-5759/a000440

Heintz, S. (2017b). Putting a spotlight on daily humor behaviors: dimensionality and relationships with personality, subjective well-being, and humor styles. *Pers. Individ. Dif.* 104, 407–412. doi: 10.1016/j.paid.2016.08.042

Kenny, D. A., and McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Struct. Equ. Modeling* 10, 333–351. doi: 10.1207/S15328007SEM1003_1

Lauer, W. (1974). *Humor als Ethos: Eine Moralpsychologische Untersuchung [Humor as Ethos: A Moral Psychology Investigation]*. Bern: Huber.

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., and Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: development of the Humor Styles Questionnaire. *J. Res. Pers.* 37, 48–75. doi: 10.1016/S0092-6566(02)00534-2

McCrae, R. R., and Costa, P. T. Jr. (2013). "Introduction to the empirical and theoretical status of the five-factor model of personality traits," in *Personality Disorders and the Five-Factor Model of Personality*, eds T. A. Widiger and P. T. Costa Jr. (Washington, DC: American Psychological Association), 15–27. doi: 10.1037/13939-002

Milner Davis, J. (2003). *Farce*, 2nd Edn. Piscataway, NJ: Transaction Publishing.

Müller, L., and Ruch, W. (2011). Humor and strengths of character. *J. Posit. Psychol.* 6, 368–376. doi: 10.1080/17439760.2011.592508

Papousek, I., Ruch, W., Rominger, C., Kindermann, E., Scheidl, K., Schulter, G., et al. (2017). The use of bright and dark types of humour is rooted in the brain. *Sci. Rep.* 7:42967. doi: 10.1038/srep42967

Peterson, C., Park, N., and Seligman, M. E. P. (2005). "Assessment of character strengths," in *Psychologists' Desk Reference*, eds G. P. Koocher, J. C. Norcross, and S. S. Hill (New York, NY: Oxford University Press), 93–98.

Peterson, C., and Seligman, M. E. P. (2004). *Character Strengths and Virtues: A Handbook and Classification*. New York, NY: Oxford University Press.

Proyer, R. T., and Ruch, W. (2009). Intelligence and gelotophobia: the relations of self-estimated and psychometrically measured intelligence to the fear of being laughed at. *Humor* 22, 165–181. doi: 10.1515/HUMR.2009.008

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raskin, V. (ed.). (2008). *The Primer of Humor Research*, Vol. 8. Berlin: Walter de Gruyter. doi: 10.1515/9783110198492

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Ruch, W. (1998). *The Sense of Humor: Explorations of a Personality Characteristic*. Berlin: Walter de Gruyter.

Ruch, W. (2008). "The psychology of humor," in *The Primer of Humor Research*, ed. V. Raskin (Berlin: Mouton de Gruyter), 17–100. doi: 10.1515/9783110198492.17

Ruch, W. (2012). "Towards a new structural model of the sense of humor: preliminary findings," in *Proceedings of the AAAI Fall Symposium FS-12-02: Artificial Intelligence of Humor*, Menlo Park: AAAI Press, 68–75.

Ruch, W., and Heintz, S. (2016a). The German version of the Humor Styles Questionnaire: psychometric properties and overlap with other styles of humor. *Eur. J. Psychol.* 12, 434–455. doi: 10.5964/ejop.v12i3.1116

Ruch, W., and Heintz, S. (2016b). The virtue gap in humor: exploring benevolent and corrective humor. *Transl. Issues Psychol. Sci.* 2, 35–45. doi: 10.1037/tps0000063

Ruch, W., and Hofmann, J. (2012). "A temperament approach to humor," in *Humor and Health Promotion*, ed. P. Gremigni (Hauppauge, NY: Nova Science Publishers), 79–112.

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. T. (2014). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., Kühler, G., and Van Thriel, C., (1996). Assessing the "humorous temperament": construction of the facet and standard trait forms of the State-Trait-Cheerfulness-Inventory–STCI. *Humor Int. J. Humor Res.* 9, 303–340. doi: 10.1515/humr.1996.9.3-4.303

Ruch, W., Proyer, R. T., Harzer, C., Park, N., Peterson, C., and Seligman, M. E. P. (2010). Values in action inventory of strengths (VIA-IS): adaptation and validation of the German version and the development of a peer-rating form. *J. Individ. Diff.* 31, 138–149. doi: 10.1027/1614-0001/a000022

Ruch, W., Wagner, L., and Heintz, S. (2018). Humor, the PEN model of personality, and subjective well-being: support for differential relationships with eight comic styles. *Riv. Ital. Studi sull'Umorismo* 1, 31–44. Available at: https://www.risu.biz/wp-content/uploads/2018/01/Ruch_et_al.-RISU-11-2018-31-44-1.pdf

Schallberger, U., and Venetz, M. (1999). *Kurzversionen des MRS-Inventars von Ostendorf (1990) zur Erfassung der Fünf "Grossen" Persönlichkeitsfaktoren [Brief Versions of Ostendorf's MRS Inventory for the Assessment of the Big Five Personality Factors],* Vol. 30. Zurich: Universität Zürich, 1–51.

Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res.* 8, 23–74.

Schmidt-Hidding, W. (1963). *Europäische Schlüsselwörter: Humor und Witz,* Band I *[European Keywords: Humor and Wit*, Vol. 1]. Munich: Huber.

Stevens, J. P. (2012). *Applied Multivariate Statistics for the Social Sciences*, 5th Edn. New York, NY: Routledge.

Terry, R. L., and Ertel, S. L. (1974). Explorations in individual differences in preferences for humor. *Psychol. Rep.* 34, 1031–1037. doi: 10.2466/pr0.1974.34.3c.1031

Traub, R. E. (1994). *Reliability for the Social Sciences: Theory and Applications*, Vol. 3. Thousand Oaks, CA: Sage.

Wierzbicki, M., and Young, D. (1978). The relation of intelligence and task difficulty to appreciation of humor. *J. Gen. Psychol.* 99, 25–32. doi: 10.1080/00221309.1978.9920891

Wirth, U. (2017). *Komik: Ein Interdisziplinäres Handbuch [The Comic: An Interdisciplinary Handbook]*. Stuttgart: J. B. Metzler. doi: 10.1007/978-3-476-05391-6

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ruch, Heintz, Platt, Wagner and Proyer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 January 2018
doi: 10.3389/fpsyg.2017.02297

Pleasures of the Mind: What Makes Jokes and Insight Problems Enjoyable

Carla Canestrari[1], Erika Branchini[2], Ivana Bianchi[3], Ugo Savardi[2] and Roberto Burro[2]*

[1] Department of Education, Cultural Heritage and Tourism, University of Macerata, Macerata, Italy, [2] Department of Human Sciences, University of Verona, Verona, Italy, [3] Section Philosophy and Human Sciences, Department of Humanities, University of Macerata, Macerata, Italy

In this paper, a parallel analysis of the enjoyment derived from humor and insight problem solving is presented with reference to a "general" Theory of the Pleasures of the Mind (TPM) (Kubovy, 1999) rather than to "local" theories regarding what makes humor and insight problem solving enjoyable. The similarity of these two cognitive activities has already been discussed in previous literature in terms of the cognitive mechanisms which underpin getting a joke or having an insight experience in a problem solving task. The paper explores whether we can learn something new about the similarities and differences between humor and problem solving by means of an investigation of what makes them pleasurable. In the first part of the paper, the framework for this joint analysis is set. Two descriptive studies are then presented in which the participants were asked to report on their experiences relating to solving visuo-spatial insight problems (Study 1) or understanding cartoons (Study 2) in terms of whether they were enjoyable or otherwise. In both studies, the responses were analyzed with reference to a set of categories inspired by the TPM. The results of Study 1 demonstrate that finding the solution to a problem is associated with a positive evaluation, and the most frequent explanations for this were reported as being Curiosity, Virtuosity and Violation of expectations. The results of Study 2 suggest that understanding a joke (Joy of verification) and being surprised by it (Feeling of surprise) were two essential conditions: when they were not present, the cartoons were perceived as not enjoyable. However, this was not enough to explain the motivations for the choice of the most enjoyable cartoons. Recognizing a Violation of expectations and experiencing a Diminishment in the cleverness or awareness initially attributed to the characters in the cartoon were the aspects which were most frequently indicated by the participants to explain why they enjoyed the joke. These findings are evaluated in the final discussion, together with their limitations and potential future developments.

Keywords: pleasures of the mind, humor, cartoons, insight problem solving, the "Aha!" experience, enjoyability

OPEN ACCESS

Edited by:
Willibald Ruch,
University of Zurich, Switzerland

Reviewed by:
Tad Brunye,
Natick Solider Research, Development, and Engineering Center (NSRDEC), United States
Ursula Beermann,
University of Innsbruck, Austria

****Correspondence:***
Carla Canestrari
carla.canestrari@unimc.it

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *14 August 2017*
Accepted: *18 December 2017*
Published: *24 January 2018*

Citation:
Canestrari C, Branchini E, Bianchi I, Savardi U and Burro R (2018) Pleasures of the Mind: What Makes Jokes and Insight Problems Enjoyable. Front. Psychol. 8:2297. doi: 10.3389/fpsyg.2017.02297

INTRODUCTION

Everyone would immediately agree that humor belongs to the category of pleasurable human activities. The majority of experimental work on humor has focused on appreciation (which is clearly related to pleasure), and various theories regarding the pleasure we get from humor have been put forward. However, there are still new aspects of this topic to investigate, and this paper

explores one of these by means of a comparison between the sensations of pleasure triggered in two different but related cognitive activities: humor and insight problem solving.

The processes which are activated in insight problem solving have many structural features which also relate to humor, for example, puzzlement, instantaneous understanding, surprise, a collision of contrasting cognitive schemas and a subsequent representational change to overcome this contrast (e.g., Gick and Lockhart, 1995; Kozbelt and Nishioka, 2010; Korovkin and Nikiforova, 2015). Parallels between what happens when people "get" a joke and when they successfully solve an insight problem have been already made from a number of different perspectives (e.g., Schiller, 1938; Koestler, 1964; Suls, 1972, 1983; Fagen, 1981; Pepiciello, 1989; O'Quin and Derks, 1997; Derks et al., 1998). In both cases there is a kind of conundrum which needs to be resolved. A conundrum in the case of humor, such as in a joke for instance, often involves an incongruity in the punch line. When the joke is understood, this incongruity is resolved and a feeling of satisfaction, and therefore pleasure may arise. In insight problem solving too, there is typically a conundrum which may be either visual or verbal (Dominowski and Dallob, 1995; Öllinger and Knoblich, 2009).

What occurs in both cases is that the problem solver suddenly realizes that a representational change needs to be made in order for the incongruity to be resolved. This change requires a shift outside the initial representation of the problem (Ohlsson, 1992; Knoblich et al., 1999, 2001; Öllinger et al., 2006, 2008). Instantaneous understanding (Kozbelt and Nishioka, 2010, p. 377) and a fairly automatic revision or reorganization of the initial representation (Gick and Lockhart, 1995, p. 224) are therefore two of the basic features of the restructuring process that are common to both understanding humor and solving insight problems.

As a result of this similarity, some studies have even addressed the issue of whether humor might function as a facilitator in insight problem solving (Gick and Lockhart, 1995; Martin, 2007; Kozbelt and Nishioka, 2010; Korovkin and Nikiforova, 2015). The rationale for this, as identified by some researchers, relates to attentional processes, that is humor relieves stress thereby diluting the degree of attention being devoted to the problem (Rowe et al., 2007). This in turn stimulates the problem solver's "peripheral focus," destabilizing perceptual and thought patterns and producing a positive effect in terms of overcoming fixities and helping people to change their perspective in order to restructure the problem (Korovkin and Nikiforova, 2015). It has also been argued that humor strongly promotes associative thinking, in particular stimulating remoteness of association and the creation of non-obvious connections (Koestler, 1964; Goodchilds, 1972; Besemer and Treffinger, 1981; Sitton and Pierce, 2004). These are all related to creativity (Mednick, 1962; Koestler, 1964; Ellwood et al., 2009; Gilhooly et al., 2012, 2013) and have a facilitatory effect in insight problem solving where the solution cannot be reached by simply reproducing familiar procedures. Creative or divergent processes are required (Dominowski and Dallob, 1995; Öllinger and Knoblich, 2009).

Whereas various studies have analyzed the points of convergence relating to the cognitive processes involved in both humor and problem solving, very little research has been done into whether humor and problem solving also share points of convergence relating to the pleasurable emotions they elicit (Schiller, 1938; Csikszentmihalyi, 1990; Kahneman et al., 1999; Kubovy, 1999), despite evidence that both activities imply associative thinking and are frequently accompanied by positive emotions and moods (Schiller, 1938; Bar, 2009; Korovkin and Nikiforova, 2010; Brunyé et al., 2013; Trapp et al., 2015). The present paper aims to explore this topic further by analyzing both humor and problem solving using the same conceptual tool. The basis for this tool is a general Theory of the Pleasures of the Mind (TPM) that was published by Kubovy (1999) in a book edited by the Nobel prize winner Kahneman in collaboration with Diener and Schwarz. The subject of the book regards a complex and challenging topic, *Well-being: the foundations of Hedonic Psychology* (1999). In the various chapters forming this book, the contributors address the puzzle of what humans like and dislike, within the mindset of experimental science. In the set of empirical evidence used by Kubovy to support his theory, the relationship between humor and problem solving is hinted at but not focused on in detail. Providing experimental evidence concerning the grounds of this relationship, however, might provide a significant contribution toward a further development of the TPM. This paper aims to delve into this connection, on the one hand by strengthening any evidence resulting from a comparison of the literature on these two cognitive activities and on the other hand by proposing an empirical paradigm in order to explore this relationship experimentally.

In section Placing the Pleasure Elicited by Humor and Insight Problem Solving within a General Research Framework for Exploring Pleasures of the Mind we will briefly present the TPM and outline the reasons why it has been chosen as a point of reference. We will then discuss how in our view this "general" perspective is connected to more "local" approaches, that is, approaches that have been developed specifically to study the enjoyment people derive from humor (section Connections between the TPM Approach and More "Local" Theories on Humor) or from insight problem solving (section Connections between the TPM Approach and More "Local" Theories Relating to the Emotions Elicited by Insight Problem Solving). In the second part of the paper, we present two descriptive studies (sections Study 1: Factors Determining Enjoyment and Lack of Enjoyment in Insight Problem Solving and Study 2: Factors Determining Enjoyment or Lack of Enjoyment in Humor) that were carried out with a two-fold aim: first, to explore the applicability of the common categories of the TPM in terms of operationalizing the enjoyment (or lack of enjoyment) relating to tasks involving visuo-spatial insight problem solving (Study 1) and to humorous cartoons (Study 2), and second, to ascertain whether the results of these two studies reveal any potential benefits of using the same operational categories to investigate these topics.

PLACING THE PLEASURE ELICITED BY HUMOR AND INSIGHT PROBLEM SOLVING WITHIN A GENERAL RESEARCH FRAMEWORK FOR EXPLORING PLEASURES OF THE MIND

Whereas it is fairly evident that people experience humor as a pleasant experience, it is less obvious how this construct can be operationalized. This type of pleasurable feeling has been referred to in terms of amusement, appreciation, mirth, exhilaration, cheerfulness, hilarity, merriment and even sudden glory (e.g., Zweyer et al., 2004; Martin, 2007). All these facets of what is in effect a generally complex construct can, taken individually, be empirically investigated (e.g., Ruch et al., 1996, 1997; for an overview see Ruch, 1998).

Kubovy (1999) discussed humor as an example of a pleasurable experience within a different theoretical framework, i.e., one which aims to define the universals of pleasurable intellectual experiences such as, for example, listening to music, reading poetry, solving puzzles, bird watching, and gardening. This general theory is not usually mentioned in the literature on humor, but it seems to us to represent a comprehensive approach which encompasses the perspectives on pleasure derived from humor which have been, more or less explicitly, developed elsewhere in mainstream approaches to the subject (e.g., Keith-Spiegel, 1972; Martin, 2007; Larkin-Galiñanes, 2017).

According to the TPM, there are three main notions which go toward defining the concept of "pleasure of the mind": (1) the stimuli and activities that induce pleasures of the mind give rise to certain patterned sequences of emotions; (2) a feeling of satisfaction occurs when a definite set of expectations (the so-called *prior state*) is violated (the *onset* moment), thereby triggering a search for an interpretation (i.e., *change*) which in turn leads to the resolution of a situation or problem, and (3) there are a number of emotions that are present to varying degrees in most pleasures of the mind (*curiosity, feeling of surprise*, *joy of verification*, *virtuosity*, and *diminishment*). This harks back to Scheffler's (1991) definition of cognitive emotions as emotions that rest on a supposition relating to the contents of a person's propositional attitudes (beliefs, predictions, expectations) and bear on its epistemological status (e.g., confirmation).

More specifically, pleasures of the mind are defined as a collection of emotions distributed over time (Kubovy, 1999, see also Kahneman, 1988, 1999). The basic structure of a pleasurable episode (or stimulus) comprises an initial set of kernels that elicits a *prior state* (i.e., a set of expectations and interpretations related to the episode), a following set of kernels (i.e., *onset*) that produces a violation of the prior state triggering a search for a new interpretation (i.e., *change*) of the initial set of kernels. The emotions associated with this sequence are *suspense* (at the onset stage), which can be accompanied by *fear* or *hope* and *automatic nervous system arousal* due to the violation of expectations. At this point, *curiosity*, that originates from the unknown, emerges and triggers a search for a new interpretation. When a decision on how to reconstruct the initial interpretation has been made at the change stage, various emotions arise: feelings of *surprise*, due to the switch from the initial set of interpretations to the final one; *joy* in verifying the aptness of the new interpretation; *satisfaction* with performing a new skill (i.e., feeling virtuous due to success in finding a new interpretation) and sometimes *superiority* on discovering that the new interpretation produces a diminishment of the value of the initial interpretation. The sensation of suspense which produces tension due to the inadequacy of the initial interpretation gives way to a final feeling of relief.

Kubovy (1999, p. 146) suggests that this analysis can also apply to humor and hints at the fact that it might apply to problem solving too. We used this as a starting point to our investigation.

Connections between the TPM Approach and More "Local" Theories on Humor

We carried out a detailed analysis of the emotions that, according to various studies and theories, are said to be sequentially elicited by humor, going beyond the references mentioned by Kubovy (1999) in his original paper. We found that the TPM is in fact consistent with the core concepts of the three main approaches to humor and it somehow unites them. These are: the cognitive approach (i.e., the incongruity-resolution theory); the psycho-physiological approach (i.e., the release theory); and the sociological approach (i.e., superiority and disparagement theories). If we consider how the TPM applies to the pleasure associated with hearing a good joke, we can understand how this works. The final part of a joke, that is, the punch line, often produces a sudden and unexpected incongruity (Suls, 1972) since it is not coherent with the preceding phase (usually called the set up) and with the expectations, predictions, interpretations which have been established as part of the set up (i.e., the *prior state* in the TPM). This incongruity (referred to as the *onset* in the TPM) elicits a specific feeling referred to as, variously, confusion of thought (Maier, 1932, p. 70), puzzlement (Schiller, 1938; Berlyne, 1972, p. 56) and embarrassment (Schiller, 1938). The violation of the *prior state* provoked by the punch line triggers a *change* in the interpretation of the initial kernel on which the prior state is based (according to the TPM), and this is consistent with what both cognitive approaches to humor (e.g., Koestler, 1964; Suls, 1972; Attardo and Raskin, 1991; Giora, 1991; Vaid et al., 2002; Forabosco, 2008) and comprehensive theories of humor would claim (e.g., Apter, 1982; Wyer and Collins, 1992; Attardo, 2017). With reference to the former, in particular, this change in interpretation is the result of the resolution of the incongruity. It has also been demonstrated that this pattern elicits pleasurable emotions in those who are telling the joke (Hull et al., 2016).

Leaving aside the structure of the kernels, let us now focus on the emotions that, according to the TPM, are produced by and typically characterize pleasurable experiences in order to determine whether studies on the enjoyment that people derive from jokes also identified the same specific sensations.

(a) *Curiosity*—the pleasure which comes from satisfying curiosity, that is, learning something new, involves a shift from an epistemic stance of the unknown or the uncertain to the known. This is something which has been identified

as often characterizing people's experience of humor (Watts, 1989; Canestrari et al., 2014).

(b) *Feeling of surprise*—various authors have emphasized that feeling surprised is a necessary condition for humor to be a pleasing experience although it is not the only necessary condition (e.g., Maier, 1932; Suls, 1972; for an overview of the early conceptions of surprise theories relating to humor, see Keith-Spiegel, 1972); with specific reference to jokes, the feeling of surprise has been operationalized in terms of an *optimal innovation*, that is, a pleasing balance between novelty and salience (Giora, 2002).

(c) *Joy of verification*—according to the TPM, this emotion typically occurs when people find the solution to a problem and it can be argued that this also applies to the processes related to understanding a joke when an incongruity is resolved. However, in cognitive literature on the resolution of incongruities in humor, it has often been pointed out that this resolution is incomplete since a residual incongruity persists even after the listener or reader "gets" the joke (e.g., Koestler, 1964; Rothbart and Pien, 1977; Apter, 1982; Ziv, 1984; Forabosco, 1992; Wyer and Collins, 1992). This may constitute a structural difference between problem solving and understanding humor: in the former case (i.e., when there is a "serious" incongruity), the resolution renders the initially problematic elements of the situation completely coherent with the solution, without any "residuals"; this does not apply to humor in which case a radical re-interpretation of the *prior state* is implied without the implications of the initial interpretation being eliminated. This has been described by, for example, Apter (1982) in terms of the simultaneous perception of two contradictory and synergetic viewpoints. Beattie also observed that "*an uncommon mixture of relation and contrariety, exhibited, or supposed to be united, in the same assemblage*" provokes a pleasant emotion whose external sign is laughter (Beattie, 1776, p. 454) and Koestler referred to the bisociation theory as a key element of humor, that is, "*the perceiving of a situation or idea […] in two self-consistent but habitually incompatible frames of reference*" (1964, p. 35, italics by the author).

(d) *Virtuosity*—virtuosity seems to fit in well with the idea of *cognitive mastery* (e.g., McGhee, 1974; Forabosco, 1992, 2008); in the context of humor, the term cognitive mastery refers to the cognitive competence required to make an incongruity congruous; the acquisition and evolution of this ability depend on cognitive development (e.g., McGhee, 1974; Forabosco, 2008), for example, the humorous incongruities which can be understood in early childhood are very simple (such as a funny face), but more complex forms of incongruity (for example irony) are only understood much later (Dews et al., 1995; Pexman et al., 2005; Angeleri and Airenti, 2014; Bianchi et al., 2017); the term virtuosity, which in the TPM refers to the pleasure derived from doing something that we could not do before (Kubovy, 1999, p. 147), is clearly applicable to this sense of mastery signaling a cognitive development (e.g., McGhee, 1979; Pien and Rothbart, 1980; Bergen, 1998), something which is in fact frequently impaired in a number of mental disabilities (e.g., Forabosco, 1998, 2008; Ivanova et al., 2014).

(e) *Violation of expectations*—in the TPM the sense of relief comes from a relaxation of the initial tension caused by a violation of the expectations established in the prior state; the incongruity which arises between the initial interpretation of the joke and the punch line results in a feeling of tension or suspense which is only relieved when the joke is understood; this is consistent with the psycho-physiological approach to humor which goes back to Spencer and Freud and was also later developed in relation to humor in art works (e.g., Berlyne, 1972; Wyer and Collins, 1992; Bonaiuto, 2006); Berlyne (1972) explicitly connects the feelings of confusion or tension elicited by the perception of an incongruity to the hedonic value of humorous stimuli since they result in an increase in arousal which is released when the incongruity is resolved and the humor is understood;

(f) *Diminishment*—in the TPM, and according to the theories developed by Wyer and Collins (1992) and Apter (1982), it is possible that a reinterpretation of the kernel may diminish some aspects of the initial interpretation thereby eliciting a feeling of superiority. This is in line with superiority theories which claim that humor often involves laughing at someone else's weakness, defect, or misfortune (for a review see Keith-Spiegel, 1972; Martin, 2007; Larkin-Galiñanes, 2017).

Connections between the TPM Approach and More "Local" Theories Relating to the Emotions Elicited by Insight Problem Solving

The Eureka moment or "Aha!" experience, that is the moment in which the solution pops up in problem solvers' minds, suddenly and unexpectedly (Durso et al., 1994; Wegner, 2002), can be regarded as the defining feature of insight. Studies aiming to describe the insight experience focused on the "Aha!" experience (Kaplan and Simon, 1990; Gick and Lockhart, 1995; Bowden and Jung-Beeman, 1998; Boden, 2004; Bowden et al., 2005; Kounios et al., 2006; Danek et al., 2013, 2014a,b; Fedor et al., 2015; Hedne et al., 2016; Salvi et al., 2016; Shen et al., 2016; Webb et al., 2016). It has been demonstrated that the "Aha!" experience is not a unitary construct but a multidimensional one in which there is an interplay of cognitive and emotional components. Some of these components map with the emotions that, according to the TPM approach, characterize pleasurable events in general (and also specifically humor).

(a) *Curiosity,* according to the TPM is characterized by an initial state of tension related to not knowing something and by a final state of relief when the new information is acquired. Danek et al. (2014b) stated that "the release of tension" is in fact an aspect characterizing the "Aha!" experience. In insight problems, tension arises from the very beginning, since there is no obvious solution to the problem, and unsuccessful problem solving attempts built the tension up further. If finally a solution is found, the tension rapidly declines. Drive, that is another aspect of the "Aha!" experience which consists of the motivation to work and to continue working on the

problem (Ohlsson, 1984; Danek et al., 2014a,b), also belongs to this category.

(b) *Feeling of surprise*, in problem solving, is associated with the disclosure of the solution. It has been proved that it can vary in strength, and it can be accompanied by either positive (delight) or negative (chagrin) emotions (Gick and Lockhart, 1995; Danek et al., 2014a,b; Hill and Kemp, 2016).

(c) *Joy of verification* corresponds to what, in the literature on problem solving, has been called the "intuitive sense of success," that is, the certainty that an insightful solution is correct (Gick and Lockhart, 1995; Bowden and Jung-Beeman, 2007; Danek and Wiley, 2017). This aspect has also often been described in relation to scientific discoveries (Gick and Lockhart, 1995).

(d) *Virtuosity* captures the emotions that in the literature on problem solving have been referred to as "performance-related aspects" (Danek et al., 2014b), which are manifested in the problem solvers' comments about their ability to find the solution to a problem.

(e) *Violation of expectations* corresponds to what has been described as suddenly realizing that certain features which were not obviously relevant, and in fact were not initially focused on and encoded ("selective elaboration" or "selective encoding," Ohlsson, 1984; Gick and Lockhart, 1995; Danek et al., 2014b), were in reality extremely relevant to the solution.

(f) *Dimishment* has been not mentioned as such in previous literature on problem solving, but it may be applied to the emotion of discovering that the solution to the problem was right under the problem solver's eyes but hidden (Dominowski and Dallob, 1995; Öllinger and Knoblich, 2009). This emotion has been conceptualized and discussed more in terms of negative than of positive emotions, i.e., a sort of chagrin due to the problem solvers' prior stupidity (Gick and Lockhart, 1995; Danek et al., 2014b; Hill and Kemp, 2016).

STUDY 1: FACTORS DETERMINING ENJOYMENT AND LACK OF ENJOYMENT IN INSIGHT PROBLEM SOLVING

In the previous section (section Connections between the TPM Approach and More "Local" Theories Relating to the Emotions Elicited by Insight Problem Solving), it was shown that the cognitive emotions referred to in the TPM are not extraneous to the emotions revealed in other studies on insight problem solving. We might also ask whether they constitute a systematic list to usefully support empirical investigations into self-reports from problem solvers.

In this study, we focused on visuo-spatial insight problems. Three different conditions were investigated. These differed in terms of the degree of direct engagement of the problem solver in the search for a solution: in a relatively "standard" condition, the participants were given 7 min to solve each problem (e.g., Schooler et al., 1993; Fleck and Weisberg, 2013; Ball et al., 2015); in another condition, the time at their disposal was reduced to 3 min, and, in the third condition, the participants were not asked to try to solve the problems, but were instead immediately given a sheet of paper showing the solutions. In all of the conditions which were tested, after the solutions were revealed, the participants were asked to indicate which two problems they liked the most, which two they liked the least, and to explain their choices. Their explanations were analyzed in terms of a set of categories which had been derived from the TPM and re-formulated as "operational categories" (see **Table 1**). This is a descriptive study. There were no specific expectations regarding how frequently the various different categories would occur and there were no precise predictions about whether successfully solving the problems (or not solving them) would have a linear effect on the motivations the participants gave for why they found the problems enjoyable or not. We were rather aiming to explore whether analyzing responses in terms of these categories would lead to a meaningful pattern which might in turn indicate a further predictive research phase.

Materials and Methods

Participants

Two hundred and sixteen Italian undergraduate students (101 males, 115 females, $M = 21.9$ years, $SD = 6.97$ years) participated in the study (72 in the 7 min condition, 72 in the 3 min condition, 72 in the no engagement condition). The experiment was carried out in a room at the University of Macerata, Italy. All of the participants gave their written informed consent. The study conforms to the ethical principles of the declaration of Helsinki (World Medical Association, 2013) and was approved by the ethical committees of the University Departments of the researchers involved in study.

Materials

Six visuo- spatial insight problems were used in all conditions (see **Figure 1**). The order of the six problems was randomized between participants.

Procedure

One booklet was given to each individual participant with the six problems printed on separate A4 sheets of paper (with the order randomized between individuals). The instructions were read out by the experimenter and projected on a screen. In the two engagement conditions (i.e., 7 min engagement and 3 min engagement), the participants were given 7 and 3 min, respectively, to read and solve each problem. They were instructed to raise their hands when they thought they had found the correct solution. If the solution was correct, they could stop, if not, they were encouraged by the experimenter to keep searching until the end of the time at their disposal. After participants had tried to solve all six problems, they were given a sheet of paper showing a table with the title of each problem, its solution and a brief explanation of the solution (solution sheet). In the third condition, no engagement, participants were simply given the initial booklet and then immediately afterwards the solution sheet.

In all three conditions, the participants were then requested to specify on a preference sheet the two problems that they

TABLE 1 | The operational categories used to analyze the explanations provided by participants in Study 1.

Curiosity	***The TPM:*** Being curious means that you get pleasure from learning something that you did not previously know. ***Definition in Problem Solving:*** Curiosity is experienced by problem solvers when, in the initial stage (problem setting), they feel a state of tension related to not knowing the solution. It is the experience of "missing something" (the solution) that prompts them to look for what they do not know yet (i.e., to move from the unknown to the known). When they know the solution, they know something new and this leads to a final state of relief. ***Examples (most enjoyable problems)***: "I was very intrigued by the problem… and then also by the solution" [pigs in a pen]; "This was the problem that from the beginning most roused my curiosity" [triangle]. ***Examples (least enjoyable problems)***: "I did not find the solution and I felt as if I was left up in the air, out of the picture" [pigs in a pen]; "The problem did not make me curious, even after I learned what the solution was" [five-square].
Virtuosity	***The TPM:*** Virtuosity refers to the pleasure you have when you feel you are doing something well. ***Definition in Problem Solving*:** Virtuosity relates to experiencing the mastery of being able to cope with the task, to reason about possible solution paths. When problem solvers find the correct solution, they feel proud of their reasoning skills. (note: both virtuosity and curiosity concern stepping from an initial state of not knowing the solution—and not knowing whether one will be able to discover it – to knowing it. However, the focus in curiosity is on "learning a new content," whereas in virtuosity the focus is on discovering, or confirming, one's reasoning skills). ***Examples (most enjoyable problems)***: "The kind of reasoning involved was both intuitive and mathematical and presupposed a bit of knowledge of the subject" [circumference]; "It was thought provoking in terms of the reasoning which it necessitated" [eight-coins]. ***Examples (least enjoyable problems)***: "I remained focused for too long on the lines forming the head [in the deer problem] and my brain got stuck; then it was impossible to start reasoning in other directions"; "In order to solve this problem [the circumference problem] you needed to apply geometrical rules that you were supposed to remember."
Violation of expectations	***The TPM:*** You search for an interpretation of the source of the violation of your expectation. You get pleasure from the violation of expectations followed by a return to a stable state. ***Definition in Problem Solving*:** A violation of expectations is experienced by problem solvers when they realize that a change in their initial mental representation of a problem is needed (since it is misleading) and that information that has been viewed as insignificant is in reality relevant to the solution (i.e., a shift in the focus of attention). ***Examples (most enjoyable problems)****:* "The solution leads us away from the usual way of thinking because we are used to thinking of a square as being oriented with two vertical and two horizontal lines while the solution requires them to be oblique" [pigs in a pen]; "I focused from the beginning on moving the lines representing the head and only those. I never thought of moving the legs!" [deer]. ***Examples (least enjoyable problems)****:* "The position that the deer has in the solution (sitting that way!) is not a normal position" [deer]; "I thought we had to move the coins only on the plane, while the solution was to put some on top of others" [eight coins].
Feeling of surprise	***The TPM:*** This emotion is familiar to scientists but is widespread in entertainment as well. It is a feature much sought after in the mystery genre. ***Definition in Problem Solving*:** The experience of surprise (positive surprise, negative surprise or no surprise) is associated with the disclosure of the solution. (note: Feeling of surprise is associated with unexpectedness, and is therefore often likely a consequence of the problem solvers' expectations being violated—see previous category. When, in order to explain their choices, participants explicitly referred to their expectations as being violated, responses were classified in the previous category. When they simply referred to the amazement (positive) or perplexity or no feeling of surprise (both negative) that they experienced when the solution was revealed, responses were classified in the present category). ***Examples (enjoyable problems)***: "The solution surprised me" [deer]; "The solution astonished me" [pigs in a pen]. ***Examples (least enjoyable problems)***: "The solution did not surprise me" [five square]; "The correct solution surprised me negatively: I found it meaningless"[eight coins].
Joy of verification	***The TPM:*** This emotion is familiar to scientists but is widespread in entertainment as well. The joy of verification is a characteristic of many puzzles. ***Definition in Problem Solving*:** Joy of verification is experienced by problem solvers in terms of proximity to the correct solution. (note: this category differs from Virtuosity in that participants do not explicitly refer to the pleasure of the reasoning acts they were engaged in but to their experience of verifying that their solution was in fact the correct one—or close to it). ***Examples (most enjoyable problems)***: When I received the response sheet, I verified that my solution was the right one" [circumference]; "I came very very close to the correct solution" [deer].

(Continued)

TABLE 1 | Continued

	Examples (least enjoyable problems): "I was close but still wrong" [deer]; "When I was given the solution, I realized I was very far away from the correct solution" [triangle].
Diminishment	***The TPM:*** If the reinterpretation paints a less desirable picture of the protagonist or the event (…), then you will find the event to be humorous. ***Definition in Problem Solving*:** Diminishment in problem solving is associated with a person realizing that the solution was simple while they had been trying much more complex reasoning paths. This can lead to enjoyment when the person makes fun of his/her own too convoluted reasoning (i.e., diminishes him/herself) or can lead to negative feelings when the person diminishes the problem to the status of a trivial one. ***Examples (most enjoyable problems)***: "When I read the solution, I found it so interestingly simple" [pigs in a pen]; "The solution was so simple and obvious, but at the same time very clever" [deer]. ***Examples (least enjoyable problems)***: "It was a very easy problem in the end" [circumference problem]; "The problem was too banal and the solution too elementary" [triangle].
Happiness	***Definition in Problem Solving***: A pure expression of amusement and/or enjoyment without any specific explanation for its cause. ***Examples (most enjoyable problems)***: "I liked it from the very beginning" [triangle]; "I found the solution nice" [deer]. ***Examples (least enjoyable problems)***: "Even after I learned the solution, I did not like it" [five-square]; "I do not like this problem" [eight coins].
Content type	***Definition in Problem Solving*:** An expression of amusement and pleasure related to the specific *kind* of process that needs to be activated in order to search for a solution, independently of feeling able to do it or not. ***Examples (most enjoyable problems)***: "I always enjoy working on problems with non-geometrical figures" [pigs in a pen] ; "I adore puzzles that require me to pay great attention to the words used in the text" [eight coins]. ***Examples (least enjoyable problems)***: "I don't like Geometry" [circumference]; "I have never liked solving Geometry problems" [circumference].
Superficial aspects	***Definition in Problem Solving*:** An expression of amusement and pleasure related to the superficial elements of the problem or to the images depicted in the problem. ***Examples (most enjoyable problems)***: "The problem had a simple structure and did not depend too much on the images" [triangle]; "The image was nice" [pigs in a pen]. ***Examples (least enjoyable problems)***: "It was too stylized " [deer problem]; "The elements in the image depended too much on the overall configuration" [five square].

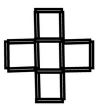
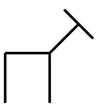
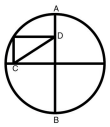
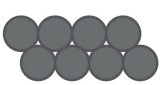
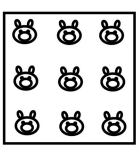
Figure	Formulation	Figure	Formulation
	Triangle (De Bono, 1969; Schooler et al., 1993): Make the triangle shape point downwards by moving only three circles.		Five-square (Katona, 1940): Reduce the number of squares from five to four by moving only three sticks.
	Deer (Origin unknown): Make the deer face in a different direction by moving just one of the lines.		Circumference (Köhler, 1969): Find the length of the oblique side CD, given the length of the diameter AB
	Eight-coin (Öllinger et al., 2013; Ormerod et al., 2002): Move two coins in such a way that each coin only touches three other coins		Pigs in a pen (Schooler et al., 1993; Ball et al., 2015): Construct two more square pens so as to ensure that each pig ends up in a pen of its own.

FIGURE 1 | The problems used in Study 1.

considered to be the most enjoyable and the two that they considered to be the least enjoyable. In both cases, they were also asked to explain their choices in an open-answer format. There were no time limits to this last phase, but all of the participants completed the task within 15 min. The language used in the task was Italian.

Categorization of Responses

Responses were analyzed based on the six different cognitive emotions described in the TPM (see **Table 1**) with three other categories (i.e., Happiness, Content type, Superficial aspects) which were added after an initial inspection of responses in order to exhaustively cover all the types of reasons referred to by the participants in the study.

Responses were classified by two independent judges with reference to each of the nine categories. Binomial coding was used, that is, the values 1 or 0 were assigned to each of the nine categories based on whether they were included in the responses or not. Each *response* (as a whole) was assigned to *at least* one category. However, it was also possible to assign it to more than one category depending on how many "chunks" (pieces of information) it could be divided into. For example, the response stating: "I found the end totally unexpected and I also liked the caricature of the faces of the subjects" was divided into two chunks since the first part refers to a violation of expectation and the second part to the superficial aspects of the cartoon (i.e., a different category). Each *chunk* was assigned to only one category. The categories that we used, technically, are partitions in that none of the categories is empty, and all the categories are disjoint sets. Both judges classified all of the responses. The inter-rater agreement was very good (*Cohen's* $\kappa = 0.901$, $SE = 0.043$). In the very few cases where the initial classifications done by the two judges did not match, a discussion took place with a third judge, and a final agreement was always reached.

Statistical Analyses

Responses were analyzed using Mixed-effect Models (Bates et al., 2015) which make it possible to deal with the variability of some factors as random effects and with the variability of other factors as fixed effects. In all the analyses, Subjects and Problems constituted random effects. In particular, we used Generalized Linear Mixed effects Models (GLMM) with the logit link function and binomial family in the case of proportions and the Poisson family in case of counts[1]. All analyses were carried out using the statistical software program R 3.3.1, with the "lme4" (Bates et al., 2015), "lsmeans" (Lenth, 2016), and "effects" (Fox, 2003) packages. We performed Mixed Model ANOVA Tables (Type 3 tests) via Wald chi-square tests implemented in the "car" package (Fox and Weisberg, 2011). Bonferroni corrections were applied to *post-hoc* comparisons. Frequency Bubble Plot were made with the "ggplot2" package (Wickham, 2016).

Results

The bubble plots shown in **Figure 2** provide a first indication of the overall frequency of the various types of explanation which the participants gave for their choice of most enjoyable or least enjoyable problems. As the plot on the left indicates, the explanations that were mentioned most frequently concerned Virtuosity, Violation of expectations and Curiosity. All of these three categories were also frequently used to explain why some problems were considered to be less enjoyable (bubble plot on the right in **Figure 2**), with the addition of considerations concerning feelings of happiness deriving from the activity of insight problem solving processes (Content type).

A GLMM (binomial, logit-link function, with Category, Condition and Enjoyability as Fixed effects) was conducted to test how responses were distributed in the three conditions, in relation to the two levels of Enjoyability (most and least enjoyable). This was done after the variability relating to the two random factors had been isolated (Subjects and Problems as random effects). The results are shown in **Figure 3** and summarized in **Table 2**.

The interaction between Category and Enjoyability turned out to be significant [$\chi^2_{(8,\, N = 216)} = 32.742$, $p \leq 0.001$], which indicates that the frequency of the various Categories significantly differed for the most enjoyable vs. the least enjoyable problems. As *post-hoc* tests revealed:

(a) significant differences emerged for only three Categories. Curiosity and Joy of verification were used more often in relation to the two most enjoyable as compared to the two least enjoyable problems. This means that being curious and searching for a solution, as well as discovering that the solution was correct (or nearly correct) triggered pleasurable sensations.

(b) feelings of Happiness relating to the specific kind of cognitive task (Content type) were conversely proportionally more frequent in the case of explanations for the choice of the least enjoyable problems as compared to the most enjoyable problems. In this case, the category was obviously used negatively (e.g., "I've never liked geometrical problems. . . .," see some of the examples provided in **Table 1**). This suggests that unpleasant feelings were linked more to an *a priori* negative evaluation of the type of task than to any specific difficulties encountered.

[1]Generalized Linear Mixed Models (GLMM) allowed us to deal with variability related to the items and to the subjects as a random effect. The items used in the experiment were in fact simply exemplars of visuo-spatial geometrical problems and humorous captioned cartoons and—in our experimental design—they were interchangeable with other items of the same type. Fixed effects are constant across individuals and random effects vary (Kreft and De Leeuw, 1998). Fixed effects are interesting in themselves; effects are random if the focus of interest in on the underlying population (Searle et al., 2008; Snijders and Bosker, 2011).

Link-function refers to the link between factors/covariates and responses. It explains how the expected value of the response relates to the linear predictor of explanatory variables. Linear regression assumes that the response variable is normally distributed (Dobson and Barnett, 2008). GLMM can have response variables with distributions other than the Normal distribution—they may even be categorical rather than continuous. Thus they may not range from $-\infty$ to $+\infty$; the relationship between the response and explanatory variables does not need to be in a simple linear form. This is why we need the link function: it links the mean of the dependent variable to the linear term in such a way that the range of the non-linearly transformed mean ranges from $-\infty$ to $+\infty$. Thus we can actually form a linear equation and use an iteratively re-weighted least squares method for a maximum likelihood estimation of the model parameters. Our dependent variable was coded as binomial-data: in this case link function is logit-function.

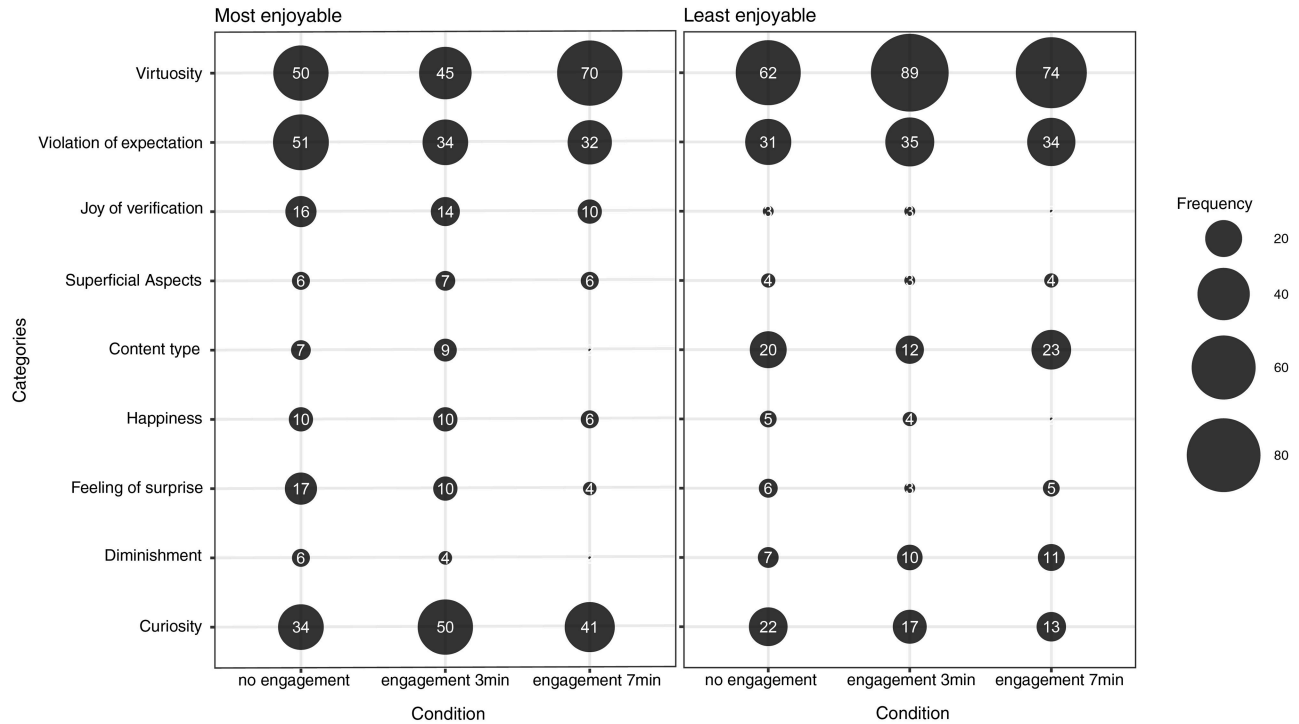


FIGURE 2 | Frequency Bubble Plots showing the overall frequency of the various Categories reported by participants in relation to their choices of the most enjoyable (graph on the left) and least enjoyable (graph on the right) problems.

A significant interaction involving Category, Enjoyability and Condition also emerged [$\chi^2_{(16,\, N\,=\,216)} = 38.442$, $p < 0.001$], while the interaction between Category and Condition did not turn out to be significant [$\chi^2_{(16,\, N=216)} = 17.445$, $p = 0.357$]. This latter result indicates that the three conditions did not lead, *per se*, to a different frequency with regard to the various Categories. Conversely, as the former finding indicates, differences only emerged between Category and Condition in interaction with the Enjoyability factor. Indeed, as *post-hoc* tests confirmed:

(a) in both of the engagement conditions, the participants referred to Curiosity more often in relation to the problems they liked the most as compared to those they liked the least;
(b) Virtuosity was mentioned more often as an explanation for choosing the two most enjoyable problems in the 7 min engagement condition than in the no engagement condition; (This result is consistent with the fact that feelings of virtuosity can only be experienced at the end of a successful search process. Learning the right strategy obviously implies engaging in the search for a reasonable amount of time, that is, sufficient time for the participant to feel virtuous. Those who did not engage in the search phase were obviously not in a position to feel virtuous);
(c) Curiosity was given as the explanation for the choice of the two most enjoyable problems more often by those participants who had been engaged in the search phase for only 3 min as compared to those participants who had not been engaged at all;
(d) a difference between being engaged in the search phase for only 3 min and not being engaged at all also emerged with respect to Virtuosity in the participants' explanations for their choice of the least enjoyable problems.

These findings suggest that participants in the 3 min condition were able to start the search phase and thus experience the typical emotions characterizing the early stages of problem solving (which are related to Curiosity). However, they could not experience the emotion characterizing the final phase, that is virtuosity, since they did not have sufficient time to find the correct solution. In fact, in the 3 min condition, Virtuosity was more frequently mentioned by participants in a negative sense, that is, they did not find the problem enjoyable since they did not have time to find the correct solution and therefore did not feel virtuous.

In a further analysis, we investigated whether the explanations given for both the most and least enjoyable problems changed depending on whether participants succeeded or not in finding the correct solution. To do this, we zoomed in on the two conditions in which the participants had been engaged in a search phase (7 min condition, and 3 min condition) and studied whether the frequency of the various Categories varied depended on whether or not they had been able to solve the problems. We conducted two new GLMMs (binomial family, logit link function), one to study the effects of Category (on 9 levels), Condition (3 min engagement, and 7 min engagement), and Success (problem solved correctly, problem not solved correctly) in relation to the two most enjoyable problems and another

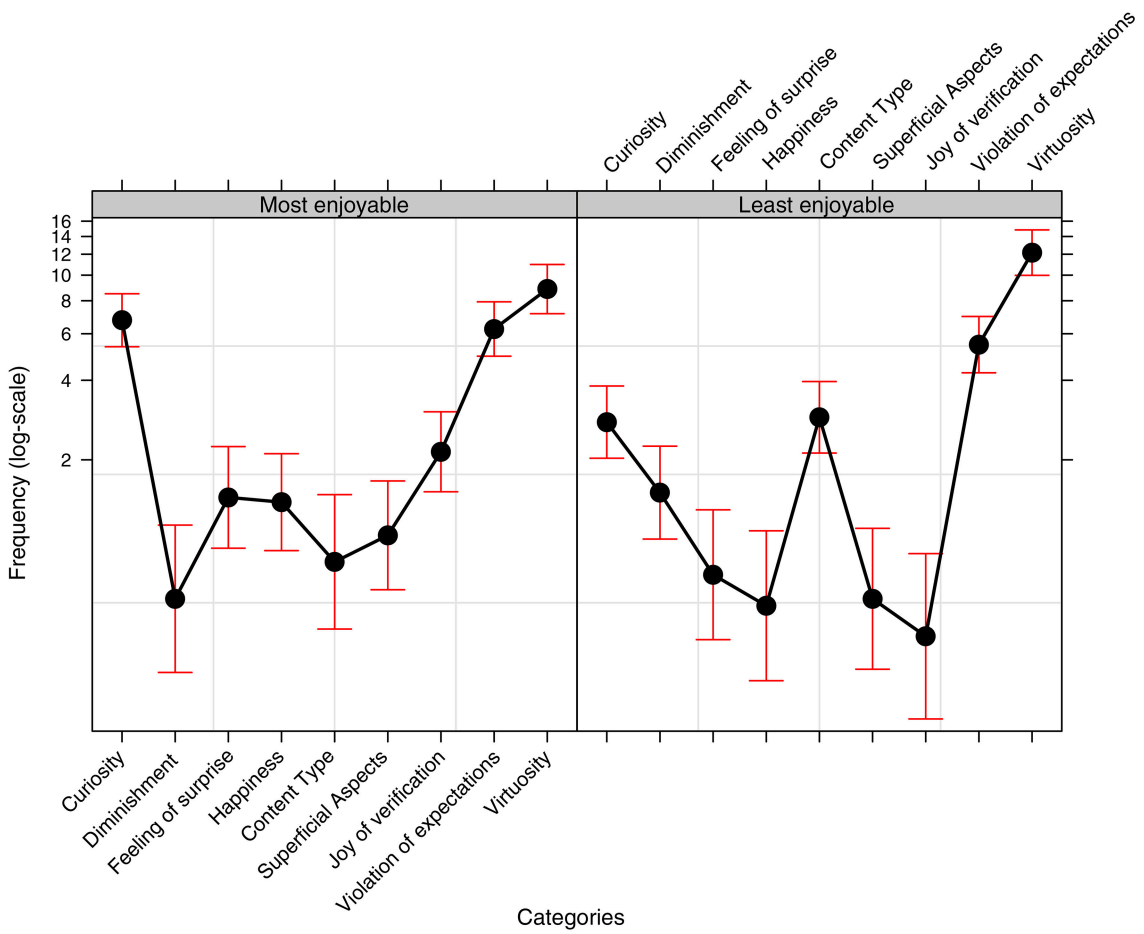


FIGURE 3 | Effect plot of the frequency data (with reference to the binomial model described in the main text) showing the proportional use (use over nonuse) of the various Motivation Categories relating to the participants' choices of the most and least enjoyable problems. Bars represent a 95% confidence interval.

TABLE 2 | Summary of the significant *post-hoc* tests resulting from the GLMM carried out on the explanations provided by participants to support their choices of the two most enjoyable and the two least enjoyable insight problems.

***Post-hoc* pairwise contrasts**	***z*-test**	***p*-value**	**Standard error**	**Effect size (odds ratio)**
Curiosity in the most enjoyable problems > Curiosity in the least enjoyable problems	5.689	<0.001	0.431	2.457
Joy of verification in the most enjoyable problems > Joy of verification in the least enjoyable problems	4.345	0.002	0.073	0.199
Content type in the least enjoyable problems > Content type in the most enjoyable problems	4.134	0.005	1.086	3.548
Curiosity in the most enjoyable problems (3 min condition) > Curiosity in the least enjoyable problems (3 min condition)	4.325	0.021	0.688	2.976
Curiosity in the most enjoyable problems (7 min condition) > Curiosity in the least enjoyable problems (7 min condition)	4.300	0.024	0.740	3.185
Virtuosity in the most enjoyable problems (7 min condition) > Virtuosity in the most enjoyable problems (no engagement condition)	4.430	0.013	0.580	2.573
Curiosity in the most enjoyable problems (3 min condition) > Virtuosity in the most enjoyable problems (no engagement condition)	4.307	0.023	0.638	2.751
Virtuosity in the least enjoyable problems (3 min condition) > Virtuosity in the least enjoyable problems (no engagement condition)	4.938	<0.001	0.532	2.629

in relation to the least enjoyable problems. In both cases, a significant interaction between Category and Success emerged. The results are summarized in **Table 3**.

In the case of the two most enjoyable problems [$\chi^2_{(8,N\ =\ 144)} = 18.780$, $p = 0.016$], the difference concerned the Violation of expectations category that was more frequently mentioned in relation to unsolved problems. The fact that this category was frequently mentioned by participants in relation to problems that they enjoyed but had not been able to solve, indicates that realizing that a switch in perspective was needed (even though this only became evident when the participants' response sheets were examined) elicited pleasurable emotions. In other words, people find pleasure in discovering that a change in the initial expectations is needed to find the solution, that fixating on the initial representation of the problem causes a block and that they can overcome this block by violating the initial expectations. "Unexpected" in this case means "enjoyable."

The interaction between Category and Success was also significant in the second GLMM [$\chi^2_{(8,N=144)} = 21.264$, $p = 0.008$] which focused on the problems which were chosen as the least enjoyable (see the section on the right in **Table 3**). Three categories were most frequently used in association with unsolved problems: Violation of expectations; Virtuosity and Curiosity. A tendency also emerged in the case of Content type. These results indicate that participants who had not being able to solve a problem and evaluated it as unpleasant/ not enjoyable reported that their negative feeling related to not having experienced being skilled enough to succeed in finding the correct solution (i.e., lack of Virtuosity), or not having felt stimulated by the problem (i.e., no Curiosity), or their frustration at not having being able to change their initial perspective (i.e., Violation of expectations).

STUDY 2: FACTORS DETERMINING ENJOYMENT OR LACK OF ENJOYMENT IN HUMOR

The results from **Study 1** showed which categories (in terms of the TPM) occurred the most frequently in the participants' explanations for their choice of the most and least enjoyable visuo-spatial insight problems of the six that they worked on. In this second study, again using the TPM as a point of reference, we aimed to explore the categories that were the most frequently included in the explanations given by the participants for their choice of the most and least enjoyable of the six captioned cartoons they were shown. In caption cartoons (also called mixed mode cartoons), both the pictorial and the textual aspects are pivotal to the interpretation of their humorous interpretation (Attardo and Chabanne, 1992; Tsakona, 2009). The reason for choosing this type of cartoon for the second study as compared to, for instance, verbal jokes, was that the six visual-insight problems used in Study 1 were also mixed mode since they consisted of both drawings and verbal texts.

Humorous stimuli are supposed to be understood quickly, otherwise the humorous effect diminishes or fails (Derks et al., 1998; Cunningham and Derks, 2005). For this reason, it was not possible to test different time conditions in Study 2 as in Study 1. The process of understanding humor is immediately activated by the presentation of a stimulus. We modulated the immediacy of the participant's access to the punch line by using one-panel and multi-panel versions of the same cartoons but the times involved were still very short. In visuo-spatial insight problems, the initial representation is provided together with a text describing the task, while the representation displaying the solution is shown at a later point (unless the problem solver immediately sees the solution but this is extremely rare). In one-panel cartoons, all the information is condensed into one image. In multi-panel cartoons, the information (i.e., the onset and resolution) is distributed across the panels and the resolution is only displayed in the last one. In this sense, spreading out the participant's access to the initial and to the final parts of the joke is more similar to what normally happens in problem solving tasks, although within a much longer timeframe.

Materials and Methods

Participants

One hundred and eighty four Italian undergraduate students (96 males, 88 females, $M = 21.8$ years, $SD = 6.44$ years) participated in the study (86 in the multi-panel condition, 98 in the single-panel condition). The experiment was carried out in a classroom

TABLE 3 | Summary of the significant *post-hoc* tests resulting from the two GLMMs conducted (one on the two most enjoyable problems, another on the two least enjoyable problems) to study the effect on the explanation category of having solved or not solved the problem.

***Post-hoc* pairwise contrasts**	***z*-test**	***p*-value**	**Standard error**	**Effect size (odds ratio)**
Violation of expectations in the most enjoyable problems (problem not solved correctly > problem solved correctly)	4.544	0.008	1.454	4.436
Violation of expectations in the least enjoyable problems (problem not solved correctly > problem solved correctly)	5.180	<0.001	8.613	23.863
Virtuosity in the least enjoyable problems (problem not solved correctly > problem solved correctly)	6.468	<0.001	9. 468	26.566
Curiosity in the least enjoyable problems (problem not solved correctly > problem solved correctly)	3.990	0.010	7.613	13.993
Content type in the least enjoyable problems (problem not solved correctly > problem solved correctly)	3.498	0.071	4.142	7.293

at the University of Verona (Italy) at the end of a class which was totally unrelated to the topic of the study. All of the participants gave their written informed consent. The study conforms to the ethical principles of the declaration of Helsinki (World Medical Association, 2013) and was approved by the ethical committees of the University Departments of the researchers involved in study.

Materials

Six caption cartoons were used. The cartoons had been taken from a website on the internet. All of them were one-panel cartoons but we modified them in order to obtain an additional multi-panel version (see **Figure 4**).

Procedure

One booklet containing the 6 cartoons was given to each individual participant with the order of the cartoons randomized between participants. The cartoons were all one-panel cartoons in one condition and all multi-panel cartoons in another condition.

The instructions were read out by the experimenter and projected on a screen. Participants were asked to look at and read the six cartoons. A sheet of paper containing a brief explanation for each cartoon was then provided (paralleling the solution sheet in Study 1). It was felt that this was needed to guarantee that everyone understood the jokes. The participants were then requested to specify which two cartoons they considered to be the most enjoyable and which two they considered to be the least enjoyable. They were also asked to explain their choices (open-answer). The format of the response sheet was identical to the one used in Study 1 with a space for them to indicate their choices and five lines for each choice in which they were requested to explain what made the cartoon particularly enjoyable or otherwise. There were no time limits, but all of the participants completed the task within 10 min. The language used for the task was Italian.

Categorization of Responses

Responses were analyzed with reference to the set of categories used in Study 1 (see **Table 1**) adapted for use with the cartoons (see **Table 4**). For the sake of simplicity, the cartoons are referred to as jokes since traditionally cartoons are frequently visual jokes (Attardo and Chabanne, 1992; Corcoran et al., 1997; Hempelmann and Samson, 2008). The application of this set of categories to humor was done on the basis of the TPM and of an initial inspection of the responses in order to guarantee that the operative tools used represented the complexity of the qualitative explanations of the participants. All of the responses were classified by two independent judges in terms of each of the nine categories. Binomial coding was used, that is, the values 1 or 0 were assigned to each of the nine categories based on whether they were contained in the responses or not. The categories were therefore not mutually exclusive. The inter-rater agreement was very good (*Cohen's* $\kappa = 0.879$, $SE = 0.051$). In the very few cases where the initial classifications done by the two judges did not match, a discussion took place with a third judge, and a final agreement was always reached.

Statistical Analyses

Responses were analyzed using Mixed-effect Models (using the same packages as those described in Study 1). In all the analyses, Subjects and Cartoons constituted random effects. We used Generalized Linear Mixed effects Models (GLMM) with the logit link function and binomial family in the case of proportions and the Poisson family in case of counts. Bonferroni corrections were applied to *post-hoc* comparisons.

Results

The bubble plots in **Figure 5** show the overall frequency of the various types of reasons which the participants gave for their choices of the most and least enjoyable cartoons. The plot on the left suggests that Violation of expectations is often referred to, and that Structural aspects concerning the subject or the graphics of the cartoon (Superficial aspects) are also frequently mentioned. Conversely, there was a greater range of reasons given for lack of enjoyability but lack of a Feeling of surprise, lack of autonomous understanding of the joke (Joy of verification) and again specific aspects relating to the subject or graphic aspect of the cartoon (Superficial aspects) were overall the Categories that the participants most frequently referred to.

A GLMM was conducted to test how responses were distributed (binomial, logit-link function, with Category, Condition and Enjoyability as Fixed effects, Subjects and Cartoons as random effects).

As the significant main effect of Categories indicates [$\chi^2_{(8,N=184)} = 82.803$, $p < 0.001$], some of the Categories were more frequently used by participants to explain their choices than others. In particular (as *post-hoc* tests confirmed), Feeling of surprise, Violation of expectations and aspects relating to the content or graphics of the cartoons (Superficial aspects) were the three categories most frequently used. However, the interaction between Category and Enjoyability also turned out to be significant [$\chi^2_{(8,N=184)} = 68.111$, $p < 0.001$], which means that the frequency of the various Categories significantly differed for the most enjoyable vs. the least enjoyable cartoons. In fact, as shown in **Figure 6** (and confirmed by the *post-hoc* tests reported in **Table 5**), Violation of expectations and Diminishment were used more often in relation to the two most enjoyable as compared to the two least enjoyable cartoons. Conversely, lack of Feeling of surprise and the absence of Joy of verification were more frequently referred to when explaining the choice of two least enjoyable cartoons. References to structural aspects were used with the same frequency for both the most enjoyable and the least enjoyable cartoons (*Odds-ratio* $= 0.618$, $SE = 0.041$, *z-ratio* $= 2.831$, $p < 0.708$).

Therefore, the findings which emerged from this study suggest that understanding a cartoon (Joy of verification) and being surprised by it (Feeling of surprise) are two conditions which are essential for pleasure: when they were not present, the cartoon was not perceived as being enjoyable. At the same time, being surprised by the punch line and understanding it do not seem to be enough to guarantee a greater degree of enjoyment: recognizing a violated expectation and experiencing a diminishment in the cleverness or awareness initially attributed

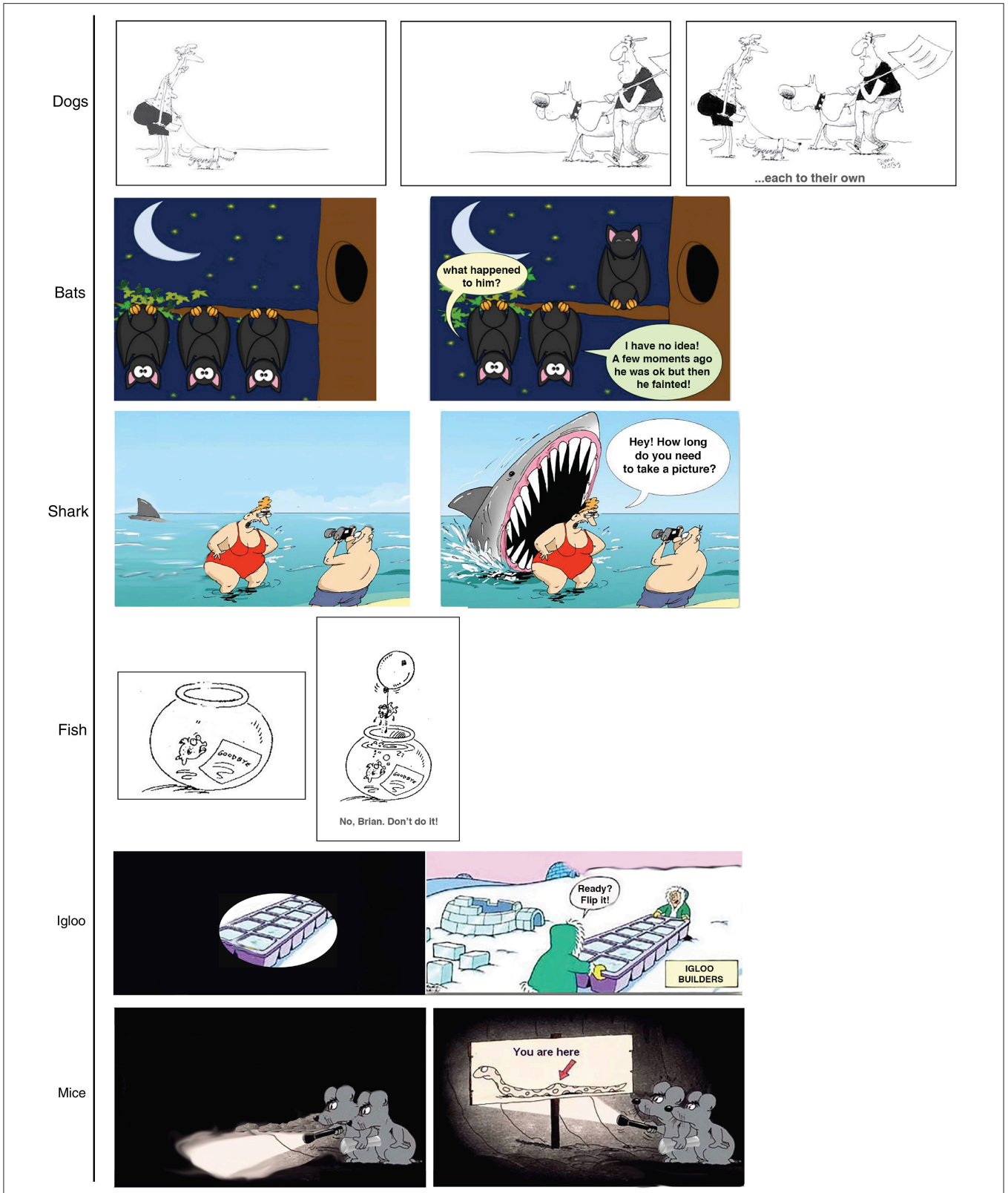


FIGURE 4 | The cartoons used in Study 2 as presented in the multi-panel condition. In the one-panel condition, only the final panel (i.e., the one on the right) was presented. The original versions of the cartoons (one-panel, in Italian) were retrieved from www.paginainizio.com on the 15th September, 2017 (by courtesy of Paginalnizio.com).

TABLE 4 | Operational categories used to analyze the explanations provided by participants in Study 2.

Curiosity	***Definition in Humor:*** Curiosity is associated with realizing that there is something incongruous in the text that needs to be understood, and this incongruity elicits a state of tension which activates the person to look for meanings. Once the resolution of a joke is achieved, a positive feeling of knowing something new (i.e., getting the joke) arises and this leads to a final state of relief. ***Examples (most enjoyable cartoons)***: "It stimulated my curiosity and made me want to understand the meaning" [bats]; "It made me curious and it interested me; this pushed me to look more carefully at the snake…it was clear that there was something humorous hidden somewhere…" [mice]. ***Examples (least enjoyable cartoons):*** "From the beginning, it left me indifferent and did not stimulate any reaction in me or any curiosity [dogs]*";* "It did not catch my attention, and therefore it did not make me want to understand its meaning" [igloo].
Virtuosity	***Definition in Humor:*** Virtuosity is the feeling of being able to understand a joke. It occurs more frequently when the joke is perceived as witty or is based on intellectual or specific domain knowledge than when it is perceived as trivial and can be understood by everyone. ***Examples (most enjoyable cartoons)***: "It was ironic to the right point: you needed to think about it a bit" [igloo]; "It offers a brilliant comparison; it is not immediately clear" [igloo]. ***Examples (least enjoyable cartoons):*** "It is much too convoluted and complicated, and did not make me laugh immediately" [bats]*;* "I was not able to understand immediately what the point of the joke was; I focused on an irrelevant aspect…" [dogs].
Violation of expectations	***Definition in Humor:*** Violation of expectations is related to discovering that the joke plays on contravening/contrasting expectations toward which the reader has been biased by the text/image at the beginning of the interpretative process, and to discovering that the resolution of the joke requires a re-structuring of the initial interpretation. ***Examples (most enjoyable cartoons):*** "It is paradoxical. Exactly the opposite of what is true for humans!" [bats]; "It is uncommon and unusual, and it excited my interest precisely because it is absurd" [mice]. ***Examples (least enjoyable cartoons):*** "It did not amuse me because the cartoon is too far from reality, it is unreal, absurd." [fish]; "I did not enjoy the igloo cartoon because the solution is too far from what actually happens" [igloo].
Feeling of surprise	***Definition in Humor:*** Feelings of surprise arise when a new interpretation, which is achieved by resolving an incongruity, is perceived as unusual. ***Examples (most enjoyable cartoons):*** "The punch line was unexpected" [fish]*;*"The punch line surprised me" [mice]. ***Examples (least enjoyable cartoons)****:* "It is banal and predictable" [shark]; "The end was predictable" [fish].
Joy of verification	***Definition in Humor:*** Joy of verification is experienced depending on the proximity of a person's understanding to the correct interpretation of the joke. (note: this category differs from Virtuosity in that the participants did not explicitly refer to their ability to understand quickly or to the subtlety of the jokes but rather mentioned the outcome of their understanding matching the "official"interpretation of the joke). ***Examples (most enjoyable cartoons):*** "It is an easy and immediate punch line, it does not require reasoning and you understand it quickly" [shark]; "I got it immediately and this made it very humorous" [igloo]. ***Examples (least enjoyable cartoons):*** "I could not get the meaning from the drawing and the text" [dogs]*;* "If I had not been given the explanation of the joke, I would have never have thought it was meant to make people laugh" [dogs].
Diminishment	***Definition in Humor:*** Diminishment is experienced when the reinterpretation of the text (i.e., the resolution of the incongruity) implies that the characters or the event on which the joke is focused are less attractive (e.g., honest, innocent, loyal, clever) than they seemed from the first impression. ***Examples (most enjoyable cartoons):*** "It is amusing that the little mice have not realized before that they have been eaten"*;* "The husband, with the excuse of taking a picture of his wife, tries to give the shark time to eat his wife, who apparently has no idea what is happening [shark]. ***Examples (least enjoyable cartoon):*** "It plays on the perceived stupidity of mice" [mice]*;* "I was sorry for the mice and felt bad when I realized that they had understood that they were in a snake's belly." [mice]
Happiness	***Definition in Humor:*** a pure expression of amusement (i.e., appreciation of a joke), without any specific explanation for its cause. ***Examples (most enjoyable cartoons):*** "The idea of a fish flying tied to a balloon makes me laugh." [fish]*;* "It was ironic and playful" [shark]. ***Examples (least enjoyable cartoons):*** "There is nothing amusing about this" [dogs]; "I did not enjoy it for no specific reason but I simply find it not very ironic" [mice].

(Continued)

TABLE 4 | Continued

Content type	***Definition in Humor:*** an expression of appreciation and amusement connected to a specific humorous genre or humorous topic. ***Examples (most enjoyable cartoons):*** "The stereotypical topic of an annoying wife who exasperates her husband is always humorous" [shark]; "It made me laugh because it plays on the customary parody of wife and husband. The relationship between the two is often compared to the formula "love-hate relationship." [shark] ***Examples (least enjoyable cartoons):*** "It represents typical masculine humor that is based on the idea that you need to get rid of the no longer desired wife, without caring about her general wellbeing. Male chauvinism." [shark] ; "It is not amusing because it relates to the issue of suicide, so I think it is black humor" [fish].
Superficial aspects	***Definition in Humor:*** An expression of amusement and appreciation related to the superficial and formal aspects of a joke. ***Examples (most enjoyable cartoons):*** "I enjoyed the characters and the facial expressions used to convey the humorous meaning" [fish]; "I found the caricature of the characters funny" [shark]. ***Examples (least enjoyable cartoons):*** "I did not appreciate it mostly because of the style of the drawing" [shark]; "The characters in the cartoon are animals which I do not like" [mice].

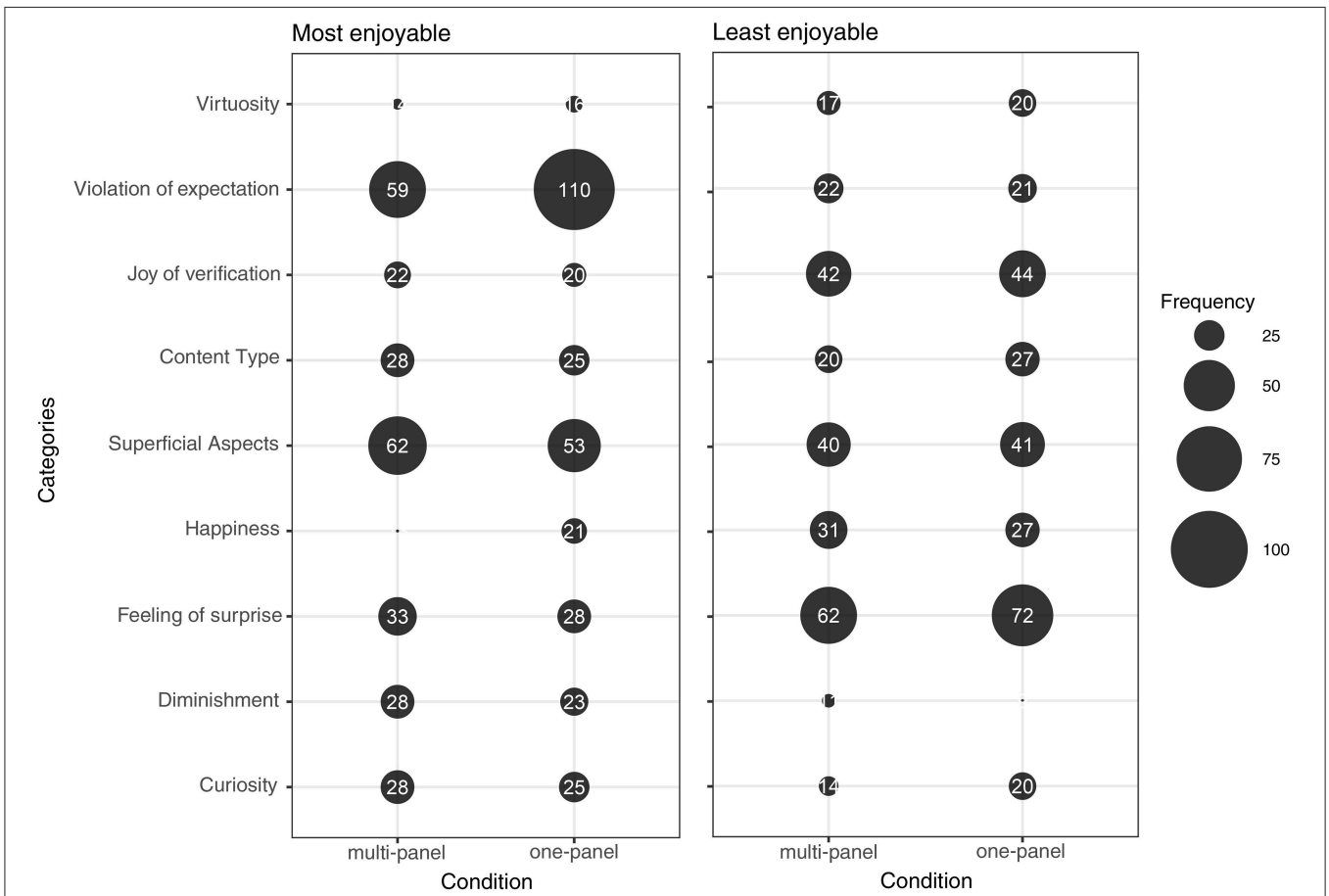


FIGURE 5 | Frequency Bubble Plots showing the overall frequency of the various Categories reported by participants in relation to their choices of the most enjoyable (graph on the left) and least enjoyable (graph on the right) cartoons.

to the characters of the joke were the two aspects which were specifically more frequently associated with the most enjoyable cartoons.

No significant differences in the distribution of responses in the one-panel as compared to the multi-panel condition emerged, there was only a trend [$\chi^2_{(8,N=184)} = 14.701$, $p < 0.065$]. A *post-hoc* inspection revealed that this related to a relatively lower frequency of the Category entitled Violation of expectations as a reason for a cartoon being chosen as the most enjoyable in the multi-panel condition (*Odds-ratio* = 0.408, *SE* = 0.088, *z-ratio* =

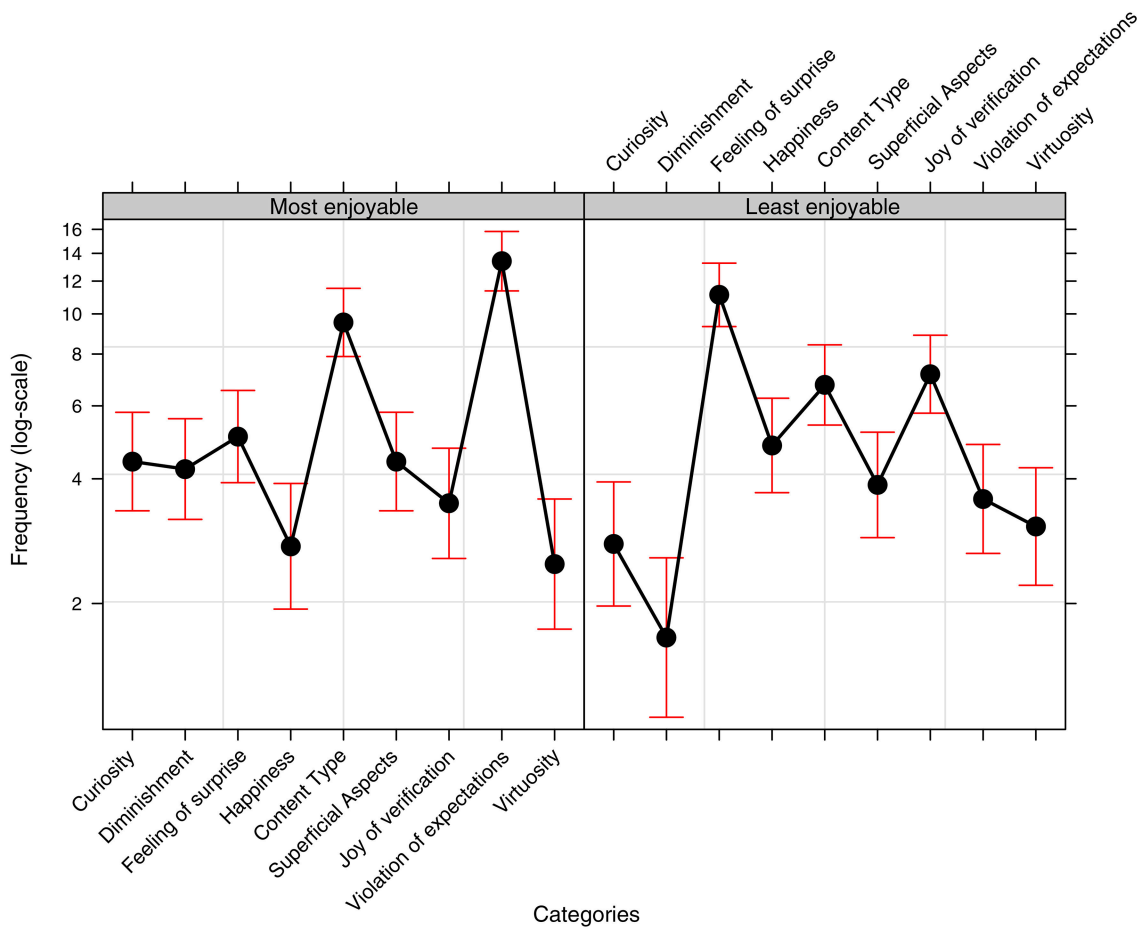


FIGURE 6 | Effect plot of the frequency data (with reference to the binomial model described in the main text) showing the proportional use (use over nonuse) of the various Categories relating to the participants' choices of the most and least enjoyable cartoons. Bars represent a 95% confidence interval.

TABLE 5 | Summary of the significant *post-hoc* tests resulting from the GLMM carried out on the explanations provided by participants to support their choices of the two most enjoyable and the two least enjoyable cartoons.

***Post-hoc* pairwise contrasts**	***z*-test**	***p*-value**	**Standard error**	**Effect size (odds ratio)**
Violation of expectation in the most enjoyable cartoons > Violation of expectation in the least enjoyable cartoons	9.315	<0.001	0.662	6.173
Diminishment in the most enjoyable cartoons > Diminishment in the least enjoyable cartoons	3.732	0.029	0.752	2.809
Feeling of surprise in the least enjoyable cartoons > Feeling of surprise in the most enjoyable cartoons	5.926	<0.001	0.511	2.872
Joy of verification in the least enjoyable cartoons > Joy of verification in the most enjoyable cartoons	4.188	<0.001	0.486	2.365

-4.146, $p = 0.021$). We will go back to this finding in the final discussion.

DISCUSSION

In this study, we explored whether new elements relating to the enjoyment experienced in problem solving and understanding humor might be discovered by comparing these two cognitive activities within a general theory of the Pleasures of the Mind. The theory we assumed as a framework (Kubovy, 1999) is based on the idea that all pleasures of the mind derive from a narrative structure which activates a corresponding sequence of emotions. The concept of narrative interpretation applies equally well to the processing involved in both solving an insight problem and

understanding a joke. In two studies (one focusing on visuo-spatial insight problems, and the other on cartoons), we explored the applicability of the same set of categories in order to analyze the participants' choices of the most enjoyable or least enjoyable problems and cartoons. We do not wish to imply that these categories describe exactly the same aspect in the two contexts. Every time general categories are instantiated in different areas (and even in different individual cases within the same area, e.g., in our case, in specific cartoons or specific visuo-spatial problems), their meaning changes slightly. There is, however, still an element which is invariant. **Tables 1**, **4** show how we modulated the same general categories for the purposes of the two contexts. **Table 1** applies to visuo-spatial insight problem solving and **Table 4** to humor. The interpretations do not aspire to be definitive; rather, they represent an initial operational proposal derived from the general definitions provided by Kubovy (1999). The question was whether putting both activities under a common umbrella (as suggested by the TPM) might reveal something in common in terms of the relative underlying cognitive mechanisms. At the present state of the art, it was not possible to formulate a predictive hypothesis regarding the application of the abovementioned set of categories to two different cognitive activities. In fact, in the original paper (Kubovy, 1999), the application of the TPM to humor and problem solving was more hinted at than actually demonstrated analytically.

With all these premises in mind, we still consider the results of our research to be extremely encouraging and further testing would certainly be worthwhile. An evaluation of whether the results of the studies also offer useful feedback in terms of a theoretical elaboration of the theory which was assumed as a framework, that is the TPM, is beyond the scope of this paper. In this paper, we have shown that the mindset underlying the TPM supports the idea of re-conceptualizing many of the proposals which have been developed in research on the subject of problem solving and humor (sections Connections between the TPM Approach and More "Local" Theories Relating to the Emotions Elicited by Insight Problem Solving and Connections between the TPM Approach and More "Local" Theories on Humor). Furthermore, we have shown that a joint application of this set of common terms to both visuo-spatial insight problems (section Study 1: Factors Determining Enjoyment and Lack of Enjoyment in Insight Problem Solving) and cartoons (section Study 2: The Factors Determining Enjoyment or Lack of Enjoyment in Humor) revealed a varying prominence of the various categories.

In problem solving, Curiosity and Joy of verification were the most often referred to in relation to the problems which were judged to be more enjoyable. This means that being fascinated by a problem and then happy to discover that the solution is in fact correct (or nearly correct) both trigger a "pleasure of the mind" experience. Conversely, lack of enjoyment was more frequently linked to an a priori negative evaluation of the type of task (category Content type) than to any specific difficulties which had been encountered during the search phase. By investigating three conditions, two requiring the participants to engage in a search for a solution (i.e., 3 min engagement or 7 min engagement) and one requiring them to simply read the problems and their solutions, it was possible to verify that Virtuosity occurred in relation to the two most enjoyable problems significantly more frequently for those participants who had been engaged in the search for 7 min as compared to those who had not been engaged at all. Engaging in a search for the solution for a reasonable amount of time (i.e., long enough to try various strategies and therefore experience being "virtuous") thus seems to be a critical factor in terms of whether or not the problem solver experiences pleasure related to virtuosity in this kind of task. Participants who were not engaged in the search phase obviously could not experience feelings of virtuosity. Those who were engaged in the search phase for only 3 min were able to experience the positive emotions that, according to the TPM, typically characterize the early stages of processing in an enjoyable activity, namely Curiosity, but they were unable to experience the emotions characterizing the final stages (e.g., Virtuosity). In fact, this latter category was usually chosen as a reason for lack of enjoyment due to feelings of frustration, that is, for negative rather than positive reasons. Finally, realizing that a change in perspective was needed in order for the problem to be solved (i.e., Violation of expectations) turned out to be a clear source of enjoyment for some but a clear source of lack of enjoyment for others. In fact, Violation of expectations was one of the most frequently reported explanations in association with both the most and the least enjoyable problems. In particular, it was more frequently reported as the cause of lack of enjoyment by those participants who had failed to solve the problems as compared to those who had succeeded (and the same held for lack of Virtuosity and lack of Curiosity).

With regard to the cartoons, Joy of verification and Feeling of surprise turned out to be two essential categories. Indeed, absence of understanding (or of clear understanding) and absence of surprise were the two categories which were significantly associated with the cartoons which were judged to be the least enjoyable. Violation of expectations was another category which occurred frequently but, in contrast with the results of the problem solving study, it was only specifically mentioned as a reason for enjoyment (i.e., it was associated with the most enjoyable rather than the least enjoyable cartoons).

As things stand, it is not possible to ascertain whether the findings of our two studies are specific to the six problems and the six cartoons used or whether it is a generalizable outcome. Further studies extending the analysis to a different sample of problems and humorous stimuli would be required for this to be established. However, as already clarified, the ambition of this paper was in no way to be all inclusive or conclusive but rather to open a research path. The above findings paint a reasonable picture of the similarities and differences relating to people's experiences of pleasure of the mind resulting from these two activities. With regard to the differences, for example, the fact that Virtuosity played a major role in problem solving but not in understanding humor seems to be in line with the consideration that the incongruity which is a basic component of humor is noticed and resolved quickly in cartoons, whereas the re-organization of a problem that needs to be addressed in insight problem solving is neither fast nor without effort. This effort

is part of the process and, as the responses of our participants confirmed, also part of the pleasure. In contrast, finding humor difficult to understand is not experienced as a part of the process; as one of the participants in **Study 2** clearly said "Even after I had understood the humor in the cartoon when I read the explanation, I could *understand* what the point was but I only got it in my "head": I didn't *experience enjoyment*."

As a final consideration, we would like to focus on the major role of Violation of expectations which emerged in both studies. In a totally different context, i.e., a cognitive analysis of the reasoning mechanisms underlying problem solving and humor, it has been demonstrated that contrast is key to any exploration of alternative strategies in insight problem solving (Branchini et al., 2015, 2016), as well as in inductive (Gale and Ball, 2012) and deductive thinking (Augustinova, 2008), and it has also been argued that contrast is fundamental to the incongruity mechanism in humor (Colston, 2002; Canestrari and Bianchi, 2012, in press; Canestrari et al., 2017; see reviews in Keith-Spiegel, 1972; Martin, 2007; Larkin-Galiñanes, 2017). The results discussed in the present paper (in particular with respect to the Violation of expectations) suggest that contrast also represents a link between insight problem solving and humor in terms of the cognitive emotions triggering pleasures of the mind.

One of the aspects that we are aware our experimental design did not factor in is the perceived complexity of the problems which were presented to the participants. In Berlyne's aesthetic theory (1972), he used the "inverted U paradigm" to demonstrate how a stimulus of medium complexity elicits an intermediate level of arousal which impacts positively on the hedonic value of the stimulus. This paradigm has also been used within the literature on humor (e.g., (Berlyne, 1972; Wyer and Collins, 1992) and the references therein) to describe the relationship between the complexity of a joke and its entertainment value, whereas, to our knowledge it has not been used to describe difficulties experienced in problem solving and the pleasure derived from it. It was also extensively discussed in Kubovy's original paper (1999). This is an aspect, in addition to widening the range of insight problems (i.e., visuo-spatial insight problems) and humor stimuli (cartoons) to include other types, that future studies would need to address.

ETHICS STATEMENT

All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethical committees of the Department of Education, Cultural Heritage and Tourism, University of Macerata, Department of Humanities, University of Macerata, Department of Human Sciences, University of Verona.

AUTHOR CONTRIBUTIONS

CC, EB, IB, US, RB substantially contributed to the conception of the work, the design of the study, the drafting of the work, and the interpretation of the data. CC, IB, EB contributed to the acquisition of the data; CC and EB contributed to the coding of responses; RB and IB contributed to the analysis of the data. CC, IB, EB, RB, US approved the final version to be published and agree to be accountable for all aspects of the work in terms of the accuracy or integrity of any part of the study.

FUNDING

This research was supported by the Department of Education, Cultural Heritage and Tourism, University of Macerata (Italy), the Department of Humanities (Section Philosophy and Human Sciences), University of Macerata (Italy), and the Department of Human Sciences, University of Verona (Italy).

REFERENCES

Angeleri, R., and Airenti, G. (2014). The development of joke and irony understanding: a study with 3- to 6-year-old children. *Can. J. Exp. Psychol.* 68, 133–146. doi: 10.1037/cep0000011

Apter, M. J. (1982). *The Experience of Motivation: The Theory of Psychological Reversal*. London: Academic Press.

Attardo, S. (2017). "The general theories of verbal humor," in *The Routledge Handbook of Language and Humor*, ed S. Attardo (London: Routledge), 126–142.

Attardo, S., and Chabanne, J. C. (1992). Jokes as a text type. *Humor* 5, 165–176. doi: 10.1515/humr.1992.5.1-2.165

Attardo, S., and Raskin, V. (1991). Script theory revis(it)ed: joke similarity and representation model. *Humor* 4, 293–348. doi: 10.1515/humr.1991.4.3-4.293

Augustinova, M. (2008). Falsification cueing in collective reasoning: example of Wason selection task. *Eur. J. Soc. Psychol.* 38, 770–785. doi: 10.1002/ejsp.532

Ball, L. J., Marsh, J. E., Litchfield, D., Cook, R. L., and Booth, N. (2015). When distraction helps: evidence that concurrent articulation and irrelevant speech can facilitate insight problem solving. *Think. Reas.* 21, 76–96. doi: 10.1080/13546783.2014.934399

Bar, M. (2009). A cognitive neuroscience hypothesis of mood and depression. *Trends Cogn. Sci.* 13, 456–463. doi: 10.1016/j.tics.2009.08.009

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beattie, J. (1776). *Essay on Laughter and Ludicrous Composition*. Edinburgh: Printed for William Creech.

Bergen, D. (1998). "Developments of the sense of humor," in *The Sense of Humor. Explorations of a Personality Characteristics*, ed W. Ruch (Berlin: Mouton de Gruyter), 329–358.

Berlyne, D. E. (1972). "Humor and its kin," in *The Psychology of Humor. Theoretical Perspectives and Empirical Issues*, eds J. Goldstein and P. McGhee (New York, NY: Academic Press), 43–60.

Besemer, S. D., and Treffinger, D. (1981). Analysis of creative products: review and synthesis. *J. Creat. Behav.* 15, 158–178. doi: 10.1002/j.2162-6057.1981.tb00287.x

Bianchi, I., Canestrari, C., Roncoroni, A. M., Burro, R., Branchini, E., and Savardi, U. (2017). The effects of modulating contrast in verbal irony as a cue for giftedness. *Humor* 30, 383–415. doi: 10.1515/humor-2017-0028

Boden, M. (2004). *The Creative Mind: Myths and Mechanisms*. New York, NY: Basic Books.

Bonaiuto, P. (2006). Art, science, and humor: the study of humorous experience at the intersection between psychology and the art world. *Empir. Stud. Arts* 24, 3–41. doi: 10.2190/47F3-NKJ0-ER8L-KKVH

Bowden, E., and Jung-Beeman, M. (1998). Getting the right idea: semantic activation in the right hemisphere may help solve insight problems. *Psychol. Sci.* 9, 435–440. doi: 10.1111/1467-9280.00082

Bowden, E., and Jung-Beeman, M. (2007). Methods for investigating the neural components of insight. *Methods* 42, 87–99. doi: 10.1016/j.ymeth.2006.11.007

Bowden, E. M., Jung-Beeman, M., Fleck, J., and Kounios, J. (2005). New approaches to demystifying insight. *Trends Cogn. Sci.* 9, 322–328. doi: 10.1016/j.tics.2005.05.012

Branchini, E., Bianchi, I., Burro, R., Capitani, E., and Savardi, U. (2016). Can contraries prompt intuition in insight problem solving? *Front. Psychol.* 7:1962. doi: 10.3389/fpsyg.2016.01962

Branchini, E., Burro, R., Bianchi, I., and Savardi, U. (2015). Contraries as an effective strategy in geometrical problem solving. *Think. Reas.* 21, 397–430. doi: 10.1080/13546783.2014.994035

Brunyé, T. T., Gagnon, S. A., Paczynski, M., Shenhav, A., Mahoney, C. R., and Taylor, H. A. (2013). Happiness by association: breadth of free association influences affective states. *Cognition* 127, 93–98. doi: 10.1016/j.cognition.2012.11.015

Canestrari, C., and Bianchi, I. (2012). Perception of contrariety in jokes. *Discourse Process* 49, 539–564. doi: 10.1080/0163853X.2012.710524

Canestrari, C., and Bianchi, I. (in press). Perceptual opposites and contrast modulation in irony. *Rev. Cogn. Linguist*.

Canestrari, C., Bianchi, I., and Cori, V. (2017). De-polarizing verbal irony. *J. Cogn. Psychol*. doi: 10.1080/20445911.2017.1368525. [Epub ahead of print].

Canestrari, C., Dionigi, A., and Zuczkowski, A. (2014). Humor understanding and knowledge. *Lang. Dialogue* 4, 261–283. doi: 10.1075/ld.4.2.05can

Colston, H. L. (2002). Contrast and assimilation in verbal irony. *J. Pragm.* 34, 111–142. doi: 10.1016/S0378-2166(02)80008-X

Corcoran, R., Cahill, C., and Frith, C. D. (1997). The appreciation of visual jokes in people with schizophrenia: a study of mentalising ability. *Schizophr. Res.* 24, 319–327. doi: 10.1016/S0920-9964(96)00117-X

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, NY: Harper & Row.

Cunningham, W. A., and Derks, P. (2005). Humor appreciation and latency of comprehension. *Humor* 18, 389–403. doi: 10.1515/humr.2005.18.4.389

Danek, A. H., Fraps, T., and von Müller, A. (2013). Aha! experiences leave a mark: facilitated recall of insight solutions. *Psychol. Res.* 77, 659–669. doi: 10.1007/s00426-012-0454-8

Danek, A. H., Fraps, T., and von Müller, A. (2014a). Working wonders? Investigating insight with magic tricks. *Cognition* 130, 174–185. doi: 10.1016/j.cognition.2013.11.003

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., and Öllinger, M. (2014b). It's a kind of magic – what self-reports can reveal about the phenomenology of insight problem solving. *Front. Psychol.* 5:1408. doi: 10.3389/fpsyg.2014.01408

Danek, A. H., and Wiley, J. (2017). What about false insights? Deconstructing the Aha! experience along its multiple dimensions for correct and incorrect solutions separately. *Front. Psychol.* 7:2077. doi: 10.3389/fpsyg.2016.02077

De Bono, E. (1969). *The Mechanism of Mind*. New York, NY: Penguin.

Derks, P., Staley, R. E., and Haselton, M. G. (1998). "Sense" of humor: perception, intelligence, or expertise?" in *The Sense of Humor. Explorations of a Personality Characteristic*, ed W. Ruch (Berlin: De Gruyter), 143–158.

Dews, S., Caplan, J., and Winner, E. (1995). Why not say it directly? The social functions of irony. *Discourse Process* 19, 347–367. doi: 10.1080/01638539509544922

Dobson, A. J., and Barnett, A. (2008). *An introduction to Generalized Linear Models, 3rd Edn*. Abingdon: Chapman & Hall/CRC.

Dominowski, R. L., and Dallob, P. (1995). "Insight and problem solving," in *The Nature of Insight*, eds R. J. Sternberg and J. E. Davidson (Cambridge: MIT Press), 33–62.

Durso, F. T. F., Rea, C. C. B., and Dayton, T. (1994). Graph-theoretic confirmation of restructuring during insight. *Psychol. Sci.* 5, 94–97. doi: 10.1111/j.1467-9280.1994.tb00637.x

Ellwood, S., Pallier, G., Snyder, A., and Gallate, J. (2009). The incubation effect: hatching a solution? *Creat. Res. J.* 21, 6–14. doi: 10.1080/10400410802633368

Fagen, R. (1981). *Animal Play Behavior*. New York, NY: Oxford University Press.

Fedor, A., Szathmáry, E., and Öllinger, M. (2015). Problem solving stages in the five square problem. *Front. Psychol.* 6:1050. doi: 10.3389/fpsyg.2015.01050

Fleck, J. S., and Weisberg, R. W. (2013). Insight versus analysis: evidence for diverse methods in problem solving. *J. Cogn. Psychol.* 25, 436–463. doi: 10.1080/20445911.2013.779248

Forabosco, G. (1992). Cognitive aspects of the humor process: the concept of incongruity. *Humor* 5, 45–68. doi: 10.1515/humr.1992.5.1-2.45

Forabosco, G. (1998). "The ill side of humor: pathological conditions and sense of humor," in *The Sense of Humor*, ed W. Ruch (Berlin: Mouton de Gruyter), 271–292.

Forabosco, G. (2008). Is the concept of incongruity still a useful construct for the advancement of humor research? *Lodz Papers Pragm.* 4, 45–62. doi: 10.2478/v10016-008-0003-5

Fox, J. (2003). Effect displays in R for generalised linear models. *J. Stat. Softw.* 8, 1–9. doi: 10.18637/jss.v008.i15

Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression*. Thousands Oaks, CA: Sage Pubblications.

Gale, M., and Ball, L. J. (2012). Contrast class cues and performance facilitation in a hypothesis testing task: evidence for an iterative counterfactual model. *Mem. Cogn.* 40, 408–419. doi: 10.3758/s13421-011-0159-z

Gick, M. L., and Lockhart, R. S. (1995). "Cognitive and affective components of insight," in *The Nature of Insight*, eds R. J. Sternberg and J. E. Davidson (Cambridge, MA: MIT Press), 197–228.

Gilhooly, K. J., Georgiou, G. J., and Devery, U. (2013). Incubation and creativity: do something different. *Think. Reas.* 19, 137–149. doi: 10.1080/13546783.2012.749812

Gilhooly, K. J., Georgiou, G. J., Garrison, J., Reston, J. D., and Sirota, M. (2012). Don't wait to incubate: immediate versus delayed incubation in divergent thinking. *Mem. Cogn.* 40, 966–975. doi: 10.3758/s13421-012-0199-z

Giora, R. (1991). On the cognitive aspects of the joke. *J. Pragmat.* 16, 465–485. doi: 10.1016/0378-2166(91)90137-M

Giora, R. (2002). "Optimal innovation and pleasure," in *Processing of The April Fools' Day Workshop on Computational Humour*, eds O. Stock, C. Strapparva, and A. Nijholt (Trento: ITC-itst), 11–28.

Goodchilds, J. (1972). "On being witty: causes, correlates and consequence," in *The Psychology of Humor*, eds J. H. Goldstein and P. E. McGhee (New York, NY: Academic Press), 173–193.

Hedne, M. R., Norman, E., and Metcalfe, J. (2016). Intuitive feelings of warmth and confidence ininsight and non insight problem solving of magic tricks. *Front. Psychol.* 7:1314. doi: 10.3389/fpsyg.2016.01314

Hempelmann, C. F., and Samson, A. C. (2008). "Cartoons: drawn jokes?" in *The Primer of Humor Research,* ed V. Raskin (Berlin: De Gruyter), 609–640.

Hill, G., and Kemp, S. M. (2016). Uh-oh! What have we missed? A qualitative investigation into everyday insight experience. *J. Creat. Behav.* doi: 10.1002/jocb.142. [Epub ahead of print].

Hull, R., Tosun, S., and Vaid, J. (2016). What's so funny? Modelling incongruity in humor production. *Cogn. Emot.* 31, 484–499. doi: 10.1080/02699931.2015.1129314

Ivanova, A. M., Enikolopov, S. N., and Mitina, O. V. (2014). Sense of humor disorders in patients with schizofrenia and affective disorders. *Psychol. Russia State Art* 7, 146–157. doi: 10.11621/pir.2014.0114

Kahneman, D. (1988). "Experimental economics: a psychological perspective" in *Bounded Rational Behavior in Experimental Games and Markets: Proceedings of the Fourth Conference on Experimental Economics, Bielefeld, West Germany, September 21-25,* eds R. Tietz, W. Albers, and R. Selten (Berlin: Springer-Verlag), 11–18.

Kahneman, D. (1999). "Objective happiness," in *Well-Being: The Foundation of Hedonic Psychology*, eds D. Kahneman, E. Diener, and D. N. Schwarz (New York, NY: Russell Sage Foundation), 3–25.

Kahneman, D., Diener, E., and Schwarz, N. (eds.). (1999). *Well-Being: The Foundations of Hedonic Psychology*. New York, NY: Russel Sage Foundation.

Kaplan, C. A., and Simon, H. A. (1990). In search of insight. *Cogn. Psychol.* 22, 374–419. doi: 10.1016/0010-0285(90)90008-R

Katona, G. (1940). *Organizing and Memorizing: Studies in the Psychology of Learning and Teaching*. New York, NY: Columbia University.

Keith-Spiegel, P. (1972). "Early conception of humor: varieties and issues," in *The Psychology of Humor,* eds J. H. Goldstein and P. E. McGhee (New York, NY: Academic Press), 3–39.

Knoblich, G., Ohlsson, S., Haider, H., and Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 1534–1556. doi: 10.1037/0278-7393.25.6.1534

Knoblich, G., Ohlsson, S., and Raney, G. E. (2001). An eye movement study of insight problem solving. *Mem. Cogn.* 29, 1000–1009. doi: 10.3758/BF03195762

Koestler, A. (1964). *The Act of Creation*. London: Hutchinson.

Köhler, W. (1969). *The Task of Gestalt Psychology*. Princeton, NJ: Princeton University Press.

Korovkin, S. Y., and Nikiforova, O. S. (2010). Cognitive and affective mechanisms of creative problem solving facilitation by humor. *Exp. Psihol.* 7, 37–51.

Korovkin, S. Y., and Nikiforova, O. S. (2015). "Humor as a facilitator of insight problem solving," *Poster Presented at the EuroAsianPacific Joint Conference on Cognitive Science* (Torino, IT).

Kounios, J., Frymiare, J. L., Bowden, E. M., Fleck, J. I., Subramaniam, K., Parrish, T. B., et al. (2006). The prepared mind. Neural activity prior to problem presentation predicts subsequent solution by sudden insight. *Psychol. Sci.* 17, 882–890. doi: 10.1111/j.1467-9280.2006.01798.x

Kozbelt, A., and Nishioka, K. (2010). Humor comprehension, humor production, and insight: an exploratory study. *Humor* 23, 375–401. doi: 10.1515/humr.2010.017

Kreft, G. G., and De Leeuw, J. (1998). *Introducing Multilevel Modeling, Introducing Statistical Methods*. London: Sage Publications Ltd.

Kubovy, M. (1999). "On the pleasures of the mind," in *Well-Being: The Foundation of Hedonic Psychology*, eds D. Kahneman, E. Diener, and D. N. Schwarz (New York, NY: Russell Sage Foundation), 134–154.

Larkin-Galiñanes, C. (2017). "An overview of humor theory," in *The Routledge Handbook of Language and Humor*, ed S. Attardo (London: Routledge), 4–16.

Lenth, R. V. (2016). Least-squares means: the R Package lsmeans. *J. Stat. Softw* 69, 1–33. doi: 10.18637/jss.v069.i01

Maier, N. R. F. (1932). A Gestalt theory of humor. *Br. J. Psychol.* 23, 64–74.

Martin, R. A. (2007). *The Psychology of Humor: An Integrative Approach*. Burlington, MA: Elsevier Academic Press.

McGhee, P. E. (1974). Cognitive mastery and children's humor. *Psychol. Bull.* 81, 721–730. doi: 10.1037/h0037015

McGhee, P. E. (1979). *Humor: Its Origin and Development*. San Francisco, CA: Freeman.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* 69, 220–232. doi: 10.1037/h0048850

Öllinger, M., Jones, G., Faber, A. H., and Knoblich, G. (2013). Cognitive mechanisms of insight: the role of heuristics and representational change in solving the eight-coin problem. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 931–939. doi: 10.1037/a0029194

Öllinger, M., Jones, G., and Knoblich, G. (2006). Heuristics and representational change in two-move matchstick arithmetic tasks. *Adv. Cogn. Psychol.* 2, 239–253. doi: 10.2478/v10053-008-0059-3

Öllinger, M., Jones, G., and Knoblich, G. (2008). Investigating the effect of set on insight problem Solving. *Exp. Psychol.* 55, 270–282. doi: 10.1027/1618-3169.55.4.269

Öllinger, M., and Knoblich, G. (2009). "Psychological research on insight problem solving," in *Recasting Reality*, eds H. Atmanspacher and H. Primas (Berlin-Heidelberg: Springer), 275–300.

O'Quin, K., and Derks, P. (1997). "Humor and creativity: a review of the empirical literature," in *The Creativity Research Handbook*, ed M. A. Runco (Cresskill, NJ: Hampton Press), 227–256.

Ohlsson, S. (1984). Restructuring revisited: I. Summary and critique of the gestalt theory of problem solving. *Scand. J. Psychol.* 25, 65–78. doi: 10.1111/j.1467-9450.1984.tb01001.x

Ohlsson, S. (1992). "Information-processing explanations of insight and related phenomena," in *Advances in the Psychology of Thinking*, eds K. J. Gilhooly and M. T. Keane (London: Harvester-Wheatsheaf), 1–44.

Ormerod, T. C., MacGregor, J. N., and Chronicle, E. P. (2002). Dynamics and constraints in insight problem solving. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 791–799. doi: 10.1037/0278-7393.28.4.791

Pepiciello, W. J. (1989). Ambiguity in verbal and visual riddles. *Humor* 2, 207–215. doi: 10.1515/humr.1989.2.3.207

Pexman, P. M., Glenwright, M., Krol, A., and Tammy, J. (2005). An acquired taste: children's perceptions of humor and teasing in verbal irony. *Discourse Process* 40, 259–288. doi: 10.1207/s15326950dp4003_5

Pien, D., and Rothbart, M. K. (1980). Incongruity and resolution in children's humor: a reexamination. *Child Dev.* 47, 966–971.

Rothbart, M. K., and Pien, D. (1977). "Elephants and marshmallows: a theoretical synthesis of incongruity-resolution and arousal theories of humour," in *It's a Funny Thing, Humour*, eds A. J. Chapman and H. C. Foot (New York, NY: Pergamon), 37–40.

Rowe, G., Hirsh, J. B., and Anderson, A. K. (2007).Positive affect increases the breadth of attentional selection. *Proc. Natl. Acad. Sci. U.S.A.* 104, 383–388. doi: 10.1073/pnas.0605198104

Ruch, W. (1998). "Humor measurements tools," in *The Sense of Humor: Explorations of a Personality Characteristic*, ed W. Ruch (Berlin: Muton De Gruyter), 404–412.

Ruch, W., Köhler, G., and van Thriel, C. (1996). Assessing the "humorous temperament": construction of the facet and standard trait forms of the State-Trait-Cheerfulness-Inventory STCI. *Humor* 9, 303–339.

Ruch, W., Köhler, G., and van Thriel, C. (1997). To be in good or bad humor: construction of the state form of the State- Trait-Cheerfulness-Inventory STCI. *Pers. Individ. Dif.* 22, 477–491. doi: 10.1016/S0191-8869(96)00231-0

Salvi, C., Bricolo, E., Kounios, J., Bowden, E., and Beeman, M. (2016). Insight solutions are correct more often than analytic solutions. *Think. Reason.* 22, 443–460. doi: 10.1080/13546783.2016.1141798

Scheffler, I. (1991). "In praise of the cognitive emotions," in *In Praise of the Cognitive Emotions and Other Essays in the Philosophy of Education*, ed I. Scheffler (New York, NY: Routledge), 2–13.

Schiller, P. (1938). A configurational theory of puzzles and jokes. *J. Gen. Psychol.* 18, 217–234. doi: 10.1080/00221309.1938.9709976

Schooler, J. W., Ohlsson, S., and Brooks, K. (1993). Thoughts beyond words: when language overshadows insight. *J. Exp. Psychol. Gen.* 122, 166–183. doi: 10.1037/0096-3445.122.2.166

Searle, S. R., Casella, G., and McCulloch, C. E. (2008). *Variance Components*. Haboken, NJ: John Wiley & Sons.

Shen, W., Yuan, Y., Liu, C., and Luo, J. (2016). In search of the 'Aha!' experience: elucidating the emotionality of insight problem-solving. *Br. J. Psychol.* 107, 281–298. doi: 10.1111/bjop.12142

Sitton, S. C., and Pierce, E. R. (2004). Synesthesia, creativity and puns. *Psychol. Rep.* 95, 577–580. doi: 10.2466/pr0.95.2.577-580

Snijders, T. A. B., and Bosker, R. I. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: SAGE Publications Ltd.

Suls, J. M. (1972). "A two-stage model for the appreciation of jokes and cartoons," in *The Psychology of Humor,* eds J. H. Goldstein and P. E. McGhee (New York, NY: Academic Press), 81–100.

Suls, J. M. (1983). "Cognitive processes in humor appreciation," in *Handbook of Humour Research*, Vol. 1: *Basic Issues*, eds P. E. McGhee and J. Goldstein (New York, NY: Springer-Verlag), 39–57.

Trapp, S., Shenhav, A., Bitzer, S., and Bar, M. (2015). Human preferences are biased towards associative information. *Cogn. Emot.* 29, 1054–1068. doi: 10.1080/02699931.2014.966064

Tsakona, V. (2009). Language and image interaction in cartoons: towards a multimodal theory of humor. *J. Pragmat.* 41, 1171–1188. doi: 10.1016/j.pragma.2008.12.003

Vaid, J., Hull, R., Heredia, R., Gerkens, D., and Martinez, F. (2002). Getting a joke: the time course of meaning attivation in verbal humor. *J. Pragmat.* 35, 1431–1449. doi: 10.1016/S0378-2166(02)00184-4

Watts, R. J. (1989). Comic strips and theory of communication. *Word Image* 5, 173–180. doi: 10.1080/02666286.1989.10435400

Webb, M. E., Little, D. R., and Cropper, S. J. (2016). Insight is not in the problem: investigating insight in problem solving across task types. *Front. Psychol.* 7:1424. doi: 10.3389/fpsyg.2016.01424

Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.

World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013. 281053

Wyer, R. S., and Collins, J. E. (1992). A theory of humor elicitation. *Psychol. Rev.* 99, 663–688. doi: 10.1037/0033-295X.99.4.663

Ziv, A. (1984). *Personality and Sense of Humor*. New York, NY: Springer-Verlag.

Zweyer, K., Velkler, B., and Ruch, W. (2004). Do cheerfulness, exhilaration, and humor production moderate pain tolerance? A FACS study. *Humor* 17, 85–119. doi: 10.1515/humr.2004.009

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Canestrari, Branchini, Bianchi, Savardi and Burro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 December 2017
doi: 10.3389/fpsyg.2017.02236

Exposure to Political Disparagement Humor and Its Impact on Trust in Politicians: How Long Does It Last?

Andrés Mendiburo-Seguel[1], Salvador Vargas[1,2] and Andrés Rubio[3,4]*

[1] Facultad de Educación, Universidad Andrés Bello, Santiago, Chile, [2] Department of Psychology, University of Girona, Girona, Spain, [3] Facultad de Psicología, Universidad Diego Portales, Santiago, Chile, [4] Fundación Centro de Estudios Cuantitativos, Santiago, Chile

The experimental research that looks into the effects of political humor on an individual's attitudes toward politics and politicians does not evaluate its long-term effects. With this in mind, this study aims to determine the possible effects that being exposed to humor which belittles politicians may have on an ordinary citizen's trust in them, while at the same time it observes the possible effects that such exposure has on them and the time such effects last. Two hypotheses were tested. The first one was that humor involves less cognitive elaboration, which leads to a short-term impact on the perception of the individual. The second one was that the repetition of a message can augment the swing of such message. Also, a series of elements regarding disposition toward politicians and political affiliation were considered. Two experiments were designed. The first experiment, (N = 94), considered three groups: one exposed to political disparagement humor; one control group exposed to disparagement humor against non-politician subjects; and a control group exposed to a non-humorous political video. Trust in politicians was evaluated first at baseline, then immediately after the experimental manipulation, and once again a week after the experimental manipulation had happened. In the second experiment (N = 146), participants were randomly assigned to one experimental and two control groups. The trust in politicians of the three groups was estimated and they were sent political cartoons, non-political cartoons, and newspaper headlines regarding political topics twice a day for a week via WhatsApp. Trust in politicians among the three groups was assessed again after 1 week, and for a third time 1 week after that. As a result, it was observed that a one-off exposure to political disparagement humor affects trust in politicians negatively; however, the effect it attains is short-lived and can be explained through the political content of the item and not only humor. Also, being exposed to cartoons constantly for a week had no impact whatsoever on the way politics and politicians were perceived during the time the experiment was carried out. Possible explanations for these findings are discussed.

Keywords: political humor, trust in politicians, disparagement humor, elaboration likelihood model, disposition theory

OPEN ACCESS

Edited by:
Willibald Ruch,
University of Zurich, Switzerland

Reviewed by:
Jody Baumgartner,
East Carolina University, United States
Patrick Alan Stewart,
University of Arkansas, United States

****Correspondence:***
Andrés Mendiburo-Seguel
andres.mendiburo@unab.cl;
amendiburo@gmail.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *20 April 2017*
Accepted: *08 December 2017*
Published: *22 December 2017*

Citation:
Mendiburo-Seguel A, Vargas S and Rubio A (2017) Exposure to Political Disparagement Humor and Its Impact on Trust in Politicians: How Long Does It Last? Front. Psychol. 8:2236. doi: 10.3389/fpsyg.2017.02236

INTRODUCTION

The main aim of this research is to observe the possible effects that being exposed to political disparagement humor has on trust in politicians. This has been studied (e.g., Olson et al., 1999; Baumgartner et al., 2012), but the possible duration of effects has not been considered before, which provided the focus of the research reported here. Most of the empirical research on the topic of humor and politics has considered short-term effects when using experimental designs (e.g., Hobden and Olson, 1994; Olson et al., 1999; Holbert et al., 2007; Baumgartner and Morris, 2008; Kim and Vishak, 2008; Xenos and Becker, 2009; Becker, 2011, 2014; Becker and Haller, 2014) or non-experimental designs (e.g., Moy et al., 2005, 2006; Kenski and Stroud, 2006; Cao, 2008; Baumgartner, 2013) but the duration of the possible effects has not been studied using follow-up assessments after the initial post-exposure assessments. In this view, the two experiments presented here seek not only to assess the effect of political disparagement humor but also to observe its possible consequences 1 week after the exposure had occurred.

DISPARAGEMENT HUMOR

Disparagement humor (Zillmann, 1983) refers to the use of humor to denigrate a given target (Ferguson and Ford, 2008; for a review, see Wicker et al., 1980). According to Ferguson and Ford (2008, p. 283), "disparagement humor refers to remarks that (are intended to) elicit amusement through the denigration, derogation, or belittlement of a given target (e.g., individuals, social groups, political ideologies, material possessions)," enabling the expression and satisfaction of aggressive impulses in a socially acceptable way (Ferguson and Ford, 2008).

Disparagement humor is strongly related to prejudice (Ford and Ferguson, 2004), given that humor communication is not intended to be evaluated in a serious way. When a targeted group is disparaged, people will be less likely to be critical of the content of that message and will, consequently, adopt the attitudes implicit in the message (Nabi et al., 2007). This, in turn, could lead to a lower threshold to accept discrimination. This principle can be extended to the political arena, considering that when denigration is expressed in a humorous manner people will be in a good disposition to accept a negative description of politicians.

Although the use of this type of humor can be interpreted within the framework of either psychoanalytic or superiority theories, it has also been analyzed within social identity theory (Tajfel, 1974, 1979, 1982; Tajfel and Billig, 1974). According to this view, people construct their social identity through the comparison of the groups they belong to (in-groups) with other groups (out-groups). This comparison serves to achieve a positive distinctiveness, enhancing features that favor the in-group over the out-group arbitrarily. In this context, disparagement humor may be used as a way to obtain a positive distinctiveness, especially in the face of perceived identity threats from the out-groups, considering that people should be more amused when disparagement targets an out-group, as suggested in the literature (Wolff et al., 1934).

Another way of understanding the joy that disparagement humor causes can be found in Zillmann and Cantor's (1976) disposition theory of humor and disposition theory of mirth. Disposition theory of humor is a conceptual framework deriving from disparagement humor (Wicker et al., 1980), but it relates better to superiority theories than to social identity theories. According to disposition theory, the response to humorous stimuli depends on the affective disposition toward the targeted person or group (McGhee and Lloyd, 1981; Becker, 2014). This theory posits that people react affectively to any target in a continuum that ranges from extreme positivity to extreme negativity, through a neutral point. In that context, it is considered that the closer the targeted group is toward the negative pole, the more amusement, humor, or mirth will be perceived by the individuals (Zillmann and Cantor, 1976).

The literature suggests that humor in general, and disparagement humor in particular, can be enjoyed because it acts as a kind of "mental balm," which allows the sender to deliver information by bringing about "high spirits," thus creating greater possibilities for the messages to be received effectively (Sternthal and Craig, 1973; Kuiper et al., 1995). This would generate positive affect that would inhibit counterargument (Mackie and Worth, 1989), which can also be understood through the elaboration likelihood model.

The elaboration likelihood model posits that individuals are not always either thoughtful or mindless about messages (Cacioppo and Petty, 1984; Petty and Cacioppo, 1984a, 1986; Cacioppo et al., 1985). Instead, different factors influence the way in which people process the information they obtain from the environment. When these factors or sources enhance interest in the received message, the elaboration likelihood is higher, so people will be more likely to process and think carefully about the arguments proposed by the message. Conversely, when interest is lowered, the elaboration likelihood is also lower, which will lead to the opposite consequence. Therefore, messages are processed in two ways: a central route, where the message is as persuasive as the argument is adequate, and a peripheral route, which is affective and non-critical, implying less cognitive elaboration (Petty and Cacioppo, 1986).

When elaboration likelihood is high, people will prefer central routes of persuasion (Petty and Cacioppo, 1984b), meaning that they will evaluate the positive and negative arguments with some care. However, when elaboration likelihood is low, people will prefer peripheral routes. These are characterized by cues external to the actual message, such as the external features of the transmitter or the quantity of arguments instead of their quality. LaMarre et al. (2014) observed that messages based on humor tend to decrease the recipients' motivation to process the arguments underlying such messages, making it more likely for them to adopt the attitudes implicit in the message (Nabi et al., 2007), since exposure to humor implies a reduced willingness to argue against it (Baumgartner and Morris, 2008).

The elaboration likelihood model also posits that message repetition has an effect on persuasion, explained by a two-stage process (Cacioppo and Petty, 1979). When someone is exposed

to a message, the repeated presentation of it can enhance the ability to process arguments. However, this process can also lead to a second stage in which repetition can produce tedium or reactance, and therefore decreased message acceptance by, for example, acting as a negative affective cue.

It is particularly interesting to consider the elaboration likelihood model when talking about the possible effects of being exposed to political disparagement humor and its duration because attitudes formed or changed by the peripheral route are less persistent (Petty et al., 2005). Also, as moderate repetition can have positive effects on persuasion, it can be hypothesized that a constant exposure to political disparagement humor will have effects on trust in politicians and that these effects will not be as short-lived as the ones caused by a one-time exposure.

Finally, evidence supports the idea that humor is processed via the peripheral rather than the central route (Zhang, 1996; Young, 2004; Baumgartner and Morris, 2008). Zhang (1996) found that humor (in the form of humorous advertisement) was more effective in the case of people that were low in need of cognition (i.e., people who are not predisposed to scrutinize and evaluate messages) and less effective in the case of people high in need of cognition. In the case of political humor, it has been observed that when people are presented with a humorous message which criticizes a political party, they tend to challenge less than if the message was presented seriously (Young, 2004).

POLITICAL DISPARAGEMENT HUMOR AND ITS EFFECTS ON ATTITUDES

According to Paletz (1990) humor directed against authority can be subverting, involving disparagement of political figures or ideologies, and can shape the attitudes of those who are exposed to this type of humor (Zenner, 1970; La Fave and Mannell, 1976). If disparagement humor makes negative stereotypes more accessible, the same stereotypes can take a person to have specific perceptions about targeted groups (Olson et al., 1999).

Effects of political disparagement humor on attitudes of those who are exposed to it have been a subject of a range of empirical studies (e.g., Hobden and Olson, 1994; Olson et al., 1999; Moy et al., 2006; Baumgartner et al., 2012; Baumgartner, 2013; Becker, 2014; Becker and Haller, 2014), though this research has not proved completely conclusive. This means that while some studies have not found any effects of the exposure to disparagement humor on attitudes (Olson et al., 1999), others have done so. For example, Baumgartner et al. (2012) found evidence suggesting that the impersonation of Sarah Palin by Tina Fey did achieve changes in attitudes toward her candidacy as Vice President (people who saw the spoof had a higher probability of disapproving her choice). Similarly, Hobden and Olson (1994) observed that after reading disparaging jokes about lawyers, people expressed more negative evaluations about them, which could lead to dissonance and therefore changes in attitudes (Olson et al., 1999). To summarize, though the existing research is not completely conclusive, most of the literature tends to acknowledge the effects of humor, including disparagement humor, on attitudes.

AIMS AND HYPOTHESES

Two things can be concluded from the above review of the literature: firstly, political disparagement humor can have an effect on trust in politicians, as most of the previous research shows; secondly, this effect will be short-lived since humor probably implies less cognitive elaboration which leads to less persistent changes on attitudes, something that has been specifically addressed by other researchers such as Baumgartner and Morris (2008). However, moderate repetition of a message can help in reinforcing and changing attitudes. Thus, the following hypotheses guided this research:

- One-off exposure to political humor will have a negative effect on trust in politicians, but this effect will be short-lived (with the impact wearing off after a period of 1 week).
- Being exposed to political humor (through cartoons) on a daily basis will have a negative effect on trust in politicians, and this effect will not decrease after 1 week.

STUDY 1

The first study sought to determine the effect of political stand-up comedy on people's trust in politicians and whether the effect wears off over time. With this in mind, an experimental pretest–posttest control group design was created, with a first experimental group which was exposed to a video containing political disparagement humor, a second control group that was exposed to a video which showed instances of disparaging humor against regular, non-political citizens, and a third one exposed to a non-humorous political video. Trust in politicians was assessed in all the three groups at the baseline, immediately after the experimental manipulation, and once again a week later. As an incentive to take part in the experiment all participants had an equal chance to win a $50 gift voucher.

Method and Procedure

Procedure

The questionnaire was programmed on 25 computers in a university laboratory. During a 2-week period, laboratory sessions were held at the university campus. The experiment was explained to the participants in the campus in broad terms by two research assistants. After that, those who accepted to take part in the experiment were taken to the laboratory and were asked to read and sign an informed consent.

A baseline questionnaire which included the dependent variable, trust in politicians, along with disposition toward politicians, political affiliation, and assessment of sex and age was presented to the volunteers (Time 1). After having completed the baseline questionnaire, they were automatically and randomly assigned to one of the three conditions and were exposed to the respective stimuli.

Once they had watched each video, a second questionnaire was presented to them to be filled out containing the dependent variable, an assessment of cognitive elaboration, and of funniness and aversiveness (Time 2). Finally, 1 week later, they were sent

an electronic link with a third questionnaire containing the dependent variable (Time 3).

Sample

We used the G*Power software (Faul et al., 2009) to determine the minimum sample size required for obtaining a significant medium effect size (f = 0.25), given α = 0.05, and a statistical power of 0.80, assuming no correlation between measures. With this analysis, we estimated a minimum sample size of 69 individuals. One hundred and fifty-eight undergraduate students participated in Study 1, and were randomly assigned to each of the three groups. Sixty-two participants were dismissed from analyses because they either (a) failed in watching the video – which was inferred from the time they took in completing the study – or (b) did not follow the instructions appropriately (for example, used their phones, talked with other participants during the experiment, or opened web pages on the computer). Attrition followed a random pattern, given that no significant differences were found between those participants who were considered in the final sample and participants who were not, either by sociodemographic characteristics, such as age, $F(1,156) = 0.121$, $p > 0.05$, $\eta^2 = 0.001$, and sex, $\chi^2(1) = 3.599$, $p > 0.05$, or baseline trust, $F(1,154) = 2.564$, $p > 0.05$, $\eta^2 = 0.016$. Fifty-one percent of the participants were male, and the mean of the age was 20.96 (SD = 2.15). Descriptive statistics for the sample are presented in **Table 1**.

Stimuli

Three videos were used. For the videos containing humor (i.e., experimental group and the control group exposed to the disparagement humor against the non-politicians video) two edited stand-up comedy routines were used. Both were by the same comedian (Edo Caroe, a popular Chilean stand-up comedian and magician), which aired in 2015 and 2016 and presented on the Festival of Viña del Mar and the Festival del Huaso de Olmué, both Chilean festivals with live transmission to Latin America. The presentations were edited to have similar duration (12 min and 32 s for the experimental video and 13 min and 2 s for control video containing humor) and to ensure that their content would be in line with the aims of the study.

To assess the validity of the videos that we used, we asked four evaluators (university students) to rate four statements about the experimental video and three about the control video. The statements are displayed in **Table 2**.

TABLE 1 | Sex and age descriptive statistics for the three groups.

Group	*n*	Sex		Age (*SD*)
		Men	Women	
Political disparagement humor	31	51.60%	48.40%	20.90 (2.06)
Disparagement humor against non-politicians	34	46.90%	53.10%	20.75 (2.00)
Non-humorous political video	31	54.80%	45.20%	21.23 (2.40)

TABLE 2 | Statements presented to assess stimuli validity.

Political humor (experimental)	Non-political video (control)
It is funny	It is funny
It is political humor	It is political humor
There is disparagement	There is disparagement
It does not specially attack politicians of a political party, but instead criticizes transversally	–

To do this, the raters had to answer "yes" or "no" to each statement. In every case, each rater agreed on the same answer. In the case of the experimental video, the four raters answered "yes" to all the statements. In the case of the control video, the four raters answered "yes" to the statements "*it is funny*" and "*there is disparagement*" and "no" to the statement "*it is political humor.*"

Parts of the transcription of the video which used political disparagement humor are below:

> *For example Senator Pizarro. When his region most needed him, he traveled to England in order to attend a rugby match. Rugby has always been a gentleman's sport, what was he doing there?! If he wanted to see dirty people, he could have gone to La Moneda!*
> *Politicians in Chile are dumb. They were bought by big enterprises, write useless laws. For example, Jaime Orpis received bribes from Corpesca. Money, real money! Bribes in Chile are strange: I always thought that bribes involved two men with suits, sunglasses in a dark alley leaving a suitcase, or at a restaurant, passing a suitcase under the table. It is different in Chile. Here politicians give you a receipt. Let's be corrupt but keep things in order. Jaime Orpis is so stupid that he even wrote on the receipt "Bribe May 2015."*
> *Every time our politicians are on TV, no one thinks "oh, great, our politicians on TV, let's see what the new social advance is." No, it's "what did they do now?". That happens to me every time I see Gustavo Hasbún. Every time. Have you ever seen someone more stupid than Hasbún? He's so stupid that idiots refer to themselves as "Hasbún."*
> *Let's take a look at the example of Dávalos* (Michele Bachelet's son). *He got rich using his position and, not happy with that, he erased everything, all the evidence that was on his computer. You might even say he was something like Pinochet. He tried to eliminate the PC* (Note: in Spanish, PC can also mean "partido comunista" or "communist party"). *Let's be clear, the president has had bad luck. She has been a lousy manager but she has had bad luck. The other day I saw a black cat that was very scared because it met Bachelet. She never knows anything! One day she met Daddy Yankee and he told her "You know"* (Note: his catchphrase) *and the lady didn't know!*

Some parts of the transcription of the video using disparagement humor against non-politicians are below:

> *You can't imagine how nervous I am of being here. I wasn't this nervous since my wife gave birth. I hope Birth will be able to forgive us when she grows up.*

My mother always said that when you go somewhere where no one knows you, you must introduce yourself, so well, my name is Edo Caroe, I'm a comedian, I come from Temuco. Somebody from Temuco here? A horrible city, that's why I left. No, no, sorry. Just kidding. I'm proud of Temuco, I have always been.

I decided to become a comedian just to see my father smile. I then found out that I should have been a doctor, since he has a horrible facial paralysis.

My father just learned how to use WhatsApp and he spends the whole day sending nude pictures of naked women to me. He's a forensic expert.

I've always liked humor, maybe because it has always been difficult for me to be still and not move. My mother always remembers how I kicked her belly. Especially when she was pregnant with my sister.

I love my daughter. She is older now, she lost her first tooth yesterday. I apologized and promised I wouldn't drink again. Last night she went to our room exactly when my wife was having an orgasm. It was a very uncomfortable moment for me and my friends, but it was a good opportunity to teach her about teamwork.

I've always wanted to come to this city. My family was very happy for me, my grandmother who has diabetes was jumping on one leg. The other one had been amputated. But I was not sure if I should come. I thought it would be difficult, more difficult than playing Scrabble with a dyslexic kid. I decided to come because I like risks. I like risks so much, that if Johnny Herrera (Note: a football player involved in a car accident) *offers to give me a lift, I say yes. Really. I once bungee jumped from Lucho Jara's ego* (Note: a famous TV host). *And most people don't understand those that like taking risks. Loving risk is going to "Who wants to be a millionaire" and use "Ask a Friend" to call Arturo Longton.*

The second control group was exposed to a non-humorous political video. This was a video blog by the Chilean journalist Tomás Mosciatti. To assess its validity the same questions as the ones used with the humorous videos were used with the same four raters. They all agreed that it was not funny, that it was disparaging, that it had political content, and that it attacked targets across the political spectrum.

Instruments

Trust in politicians

A modified version of the Yamagishi and Yamagishi's General Trust Scale was used (Yamagishi and Yamagishi, 1994) replacing "people" by "politicians." It was assessed by means of a 100-point scale that ranged from total disagreement to total agreement and covered the following statements: "*Most politicians are essentially honest,*" "*Most politicians are essentially good and kind,*" "*Most politicians are trustworthy,*" and "*Most politicians will respond kindly when they are trusted by others.*" Scale reliability was high for the baseline questionnaire ($\alpha = 0.76$), the second questionnaire ($\alpha = 0.79$), and the third questionnaire ($\alpha = 0.82$).

Disposition toward politicians

It was assessed with the item "*How much would you say you like politicians?*" with responses ranging from 1 (*Do not like*) to 100 (*Like very much*).

Political affiliation

Participants were asked about their political ideology according to a left-right political spectrum, for which possible responses were *"Left wing," "Center-Left wing," "Center," "Center Right Wing," "Right Wing,"* or "*None of the above.*"

Funniness

It was assessed with one item that ranged from 1 (*not funny*) to 100 (*very funny*).

Aversiveness

It was assessed with one item that ranged from 1 (*no aversiveness*) to 100 (*high aversiveness*).

Cognitive elaboration

It was assessed using a modified version of the scale created by Igartua (2010), considering a 100-point scale that varied between total disagreement and total agreement to the following statements: *"I have reflected on the topic it dealt with,"* "*I have thought about the situation and the motivations of the characters,*" "*I have tried to see how the plot was related to other topics that interest me,*" and "*I have wanted to draw some conclusions about the topic addressed here."* Scale reliability was high ($\alpha = 0.81$).

Results and Discussion

Manipulation Checks

First, the experimental manipulation was checked. A univariate ANOVA revealed significant differences in funniness, $F(2,91) = 93.261$, $p < 0.001$, $\eta^2 = 0.672$. The Tukey *post hoc* test showed that the control group exposed to the non-humorous political video ($M = 14.00$, $SD = 19.689$) was different from both the experimental group ($M = 80.61$, $SD = 24.52$) and the control group exposed to the disparagement humor against non-politicians video ($M = 75.28$, $SD = 19.64$), $p < 0.001$ in both cases. In addition, there were no significant differences in aversiveness among the groups, $F(2,91) = 2.559$, $p > 0.05$, $\eta^2 = 0.053$. These results suggest that the manipulation through exposure to video was successful.

Means and standard deviations for the three groups regarding trust at times 1, 2, and 3, funniness, aversiveness, and cognitive elaboration can be found in **Table 3**.

Main Analyses

We controlled for disposition and political affiliation by using randomized groups. In this case, no differences between the groups regarding both variables were found.

The first hypothesis refers to the cognitive elaboration that each stimulus implied, so as to observe if less elaboration was being used in the case of humorous stimuli. A univariate ANOVA showed significant differences regarding this variable, $F(2,91) = 12.875$, $p < 0.000$, $\eta^2 = 0.223$. The main differences were observed contrasting the control group exposed to the video presenting disparagement humor against non-politicians

TABLE 3 | Means and standard deviations for trust times 1, 2, and 3, funniness, aversiveness, and cognitive elaboration.

		M	*SD*
Trust time 1	Political disparagement humor	30.23	13.72
	Disparagement humor against non-politicians	28.76	16.18
	Non-humorous political video	30.20	15.27
Trust time 2	Political disparagement humor	23.56	9.79
	Disparagement humor against non-politicians	30.78	15.37
	Non-humorous political video	23.25	17.38
Trust time 3	Political disparagement humor	29.06	13.03
	Disparagement humor against non-politicians	28.82	13.10
	Non-humorous political video	28.57	16.45
Funninness	Political disparagement humor	80.61	24.51
	Disparagement humor against non-politicians	75.00	19.73
	Non-humorous political video	14.00	19.69
Aversiveness	Political disparagement humor	30.00	24.80
	Disparagement humor against non-politicians	32.12	23.69
	Non-humorous political video	20.58	25.71
Cognitive Elaboration	Political disparagement humor	68.69	15.64
	Disparagement humor against non-politicians	46.10	19.34
	Non-humorous political video	63.48	21.98

(M = 45.30, SD = 19.58) to both the experimental group (M = 68.69, SD = 15.64) and the control group exposed to the non-humorous political video (M = 63.48, SD = 21.98), $p < 0.001$ in both cases.

The second hypothesis, and the core of Study 1, refers to the effects of political humor on political trust. A 3 (condition) by 3 (time) ANOVA with repeated measures was performed. The results showed a significant main effect of the measures of political trust, $F(2,182) = 6.344$, $p < 0.01$, $\eta_p^2 = 0.065$, but not of the group, $F(1,91) = 0.405$, $p > 0.05$, $\eta_p^2 = 0.009$. Nevertheless, and more importantly, the interaction between both factors was significant, $F(4,182) = 3.949$, $p < 0.01$, $\eta_p^2 = 0.080$. It is important to note that for the effects of the measures of political trust and the interaction between the factors, the observed power was high (0.896 and 0.900, respectively). When controlled by either funniness or aversiveness, a similar pattern of results was found, obtaining in both cases the same significant interaction term, $F(4,180) = 3.384$, $p < 0.01$, $\eta_p^2 = 0.070$, and $F(4,180) = 3.528$, $p < 0.01$, $\eta_p^2 = 0.073$, respectively.

Given these results, we contrasted the effects of group on political trust, for each measure separately. For the baseline, we found no significant differences by group, $F(2,91) = 0.007$, $p > 0.05$, $\eta^2 = 0.000$. In the first post-measure, we observed significant differences, $F(2,91) = 3.241$, $p < 0.05$, $\eta^2 = 0.066$. The *post hoc* analysis revealed that the control group exposed to the video presenting disparagement humor against non-politicians (M = 31.50, SD = 15.56) was marginally different from both the experimental (M = 23.56, SD = 9.79), $p < 0.1$, and the control groups exposed to the non-humorous political video (M = 23.25, SD = 17.38), $p < 0.1$. Finally, in the second post-measure, there were no significant differences between the groups, $F(2,91) = 1.815$, $p > 0.05$, $\eta^2 = 0.000$.

In sum, the overall pattern of results suggests that both groups exposed to political content declined in political trust immediately after viewing the video, but returned to the baseline levels 1 week later, as it is shown in **Figure 1** (each error bar is constructed using a 95% confidence interval of the mean).

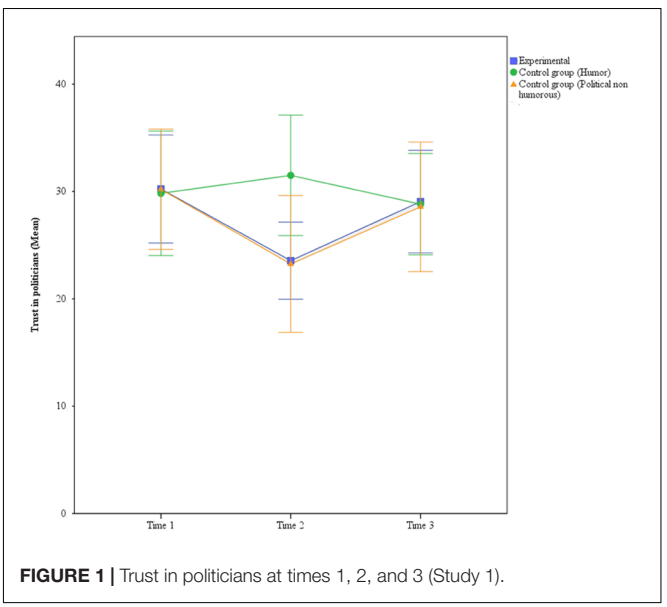


FIGURE 1 | Trust in politicians at times 1, 2, and 3 (Study 1).

It must be considered that these results are independent of either the perceived funniness or the perceived aversiveness, as was shown earlier. A supplementary analysis showed that funniness was not significantly related to trust in the baseline, the second measure, and the last measure for the experimental group, with similar results for those participants assigned to control group 1 and those assigned to control group 2. The same pattern of results was obtained analyzing the relationship between aversiveness and the three measures of trust for those assigned to the experimental

group, for those assigned to the first control group, and those assigned to the second control. In addition, there was no relationship between funniness and elaboration, neither for the experimental group nor for both the control group exposed to disparagement humor and the control group exposed to the non-humorous political video. This can be observed in **Table 4**.

Discussion

Results from study 1 show two key elements of this research. On the one hand, it was observed that the effect of political disparagement humor on individuals tends to be similar to the effect of political information that is non-humorous. This can be due to the fact that political humor implies more cognitive elaboration than non-political humor, even at the same level of political non-humorous information. On the other hand, it was also observed that the effects in both cases did not last long, being, as hypothesized, short-lived.

One topic is still open regarding whether constant presentation of a stimulus for a long period of time implies long-term effects. With the intention of addressing this, a second study was designed.

STUDY 2

The second experiment aimed to find evidence on the way that being exposed to political humor (in the form of cartoons) on a daily basis might impact trust of the individuals in politicians. For this purpose, a pretest–posttest control group design was used. Participants were university students at the university campus. They first received the baseline questionnaire which contained assessments of trust in politicians, political affiliation, disposition toward politicians, exposure to political humor, exposure to political information, sex, age, and WhatsApp number. After this, participants were randomly assigned either to an experimental or to one of two control groups, which received different stimuli via WhatsApp twice a day for 1 week. The experimental group received political cartoons; the first control group received non-political cartoons; and the second control group received newspaper headlines regarding political topics (such as conflicts of interests). Trust in politicians and attention paid to the stimuli among the three groups were assessed again after 1 week and a third time after 2 weeks via WhatsApp. As in study 1, as an incentive to take part in the experience, all participants had an equal chance to win three $50 gift vouchers.

Method and Procedure

Procedure

A research assistant contacted the participants on the university campus and explained the general aim of the study and the procedure. They were given an informed consent document that explained the study in detail. After agreeing to participate in the study, the participants were given a questionnaire with baseline questions (Time 1) containing the dependent variables (trust in politicians), political affiliation, disposition toward politicians, exposure to political humor, exposure to political information, sex, age, and a WhatsApp number. Starting

TABLE 4 | Correlations between considered variables.

		Trust time 1	Trust time 2	Trust time 3	Funninness	Aversiveness	Political affiliation	Disposition toward politicians
Political disparagement humor	Trust time 2	0.57**	1.00					
	Trust time 3	0.69**	0.60**	1.00				
	Funninness	−0.17	−0.20	−0.21	1.00			
	Aversiveness	−0.35	0.17	−0.01	−0.36*	1.00		
	Political affiliation	−0.03	0.02	−0.31*	−0.15	0.06	1.00	
	Disposition toward politicians	0.42**	0.69**	0.43**	−0.15	0.04	0.02	1.00
	Cognitive elaboration	−0.28	−0.45**	−0.38*	0.16	−0.35*	−0.14	−0.25
Disparagement humor against non-politicians	Trust time 2	0.80**	1					
	Trust time 3	0.75**	0.82**	1				
	Funninness	0.15	0.11	0.05	1			
	Aversiveness	−0.10	−0.07	0.03	−0.20	1		
	Political affiliation	−0.07	0.07	0.09	−0.03	−0.1	1	
	Disposition toward politicians	0.67**	0.70**	0.64**	0.07	−0.11	0.02	1
	Cognitive elaboration	0.15	−0.01	0.10	0.31*	0.13	−0.20	−0.08
Non-humorous political video	Trust time 2	0.70**	1.00					
	Trust time 3	0.72**	0.74**	1.00				
	Funninness	0.13	0.13	0.11	1.00			
	Aversiveness	−0.19	0.04	−0.09	0.09	1.00		
	Political affiliation	0.06	0.19	0.18	−0.13	0.27	1.00	
	Disposition toward politicians	0.60**	0.51**	0.44**	−0.10	−0.21	0.19	1.00
	Cognitive elaboration	0.06	−0.06	0.02	0.09	−0.06	−0.27	−0.15

*$p < 0.05$, **$p < 0.01$.

the next day and for 7 days, the stimuli were sent via WhatsApp to the participants who were randomly assigned to the experimental group (political cartoons), the first control group (non-political cartoons), and the second control group (newspaper headlines regarding political topics). After that week, the same questions assessing trust in politicians were sent to the experimental and control groups (Time 2). Finally, 1 week later, the three groups were sent the same questions (Time 3).

Sample

We used the GPower software (Faul et al., 2009) to determine the minimum sample size required, considering the effect size obtained in study 1 (f = 0.29), given α = 0.05, and a statistical power of 0.80, assuming no correlation between measures. With this analysis, we estimated a minimum sample size of 78 individuals. Three hundred and forty-seven students participated in the baseline (59.1% women, M_{age} = 20.90, SD = 1.73). One hundred and ninety-seven of them sent their responses back after 1 week (55.8% women, M_{age} = 20.83, SD = 1.69). Finally, 146 sent their responses back 1 week after that (50.7%women, M_{age} = 20.81, SD = 1.67). It can be established that there are no differences between those who were part of the final sample and those who were not, since attrition followed a random pattern. No significant differences were found between the two groups regarding age, $F(1,345) = 0.738$, $p > 0.05$, $\eta^2 = 0.002$, sex, $\chi^2(1) = 2.573$, $p > 0.05$, baseline trust, $F(1,343) = 2.762$, $p > 0.05$, $\eta^2 = 0.008$ and disposition toward politicians, $F(1,345) = 0.997$, $p > 0.05$, $\eta^2 = 0.003$. Descriptive statistics for the sample are presented in **Table 5**.

Stimuli

We used 14 political cartoons selected from image databases that implied criticism toward politicians in general, with no party-political bias. None referred to a politician or political figure identified by name or appearance. Two university students rated the 14 cartoons with complete agreement, evaluating disparagement ("*There is disparagement*" with response options being "yes" and "no"), if they were political ("*It is political humor*" with response options being "yes" and "no"), and if they were transversal ("*It does not specially attack politicians of a political party, but instead criticizes transversally*" with response options being "yes" and "no"). An example of a cartoon by the Chilean cartoonist Malaimagen is displayed in **Figure 2**. In the case of the 14 non-political cartoons and the 14 newspaper headlines there was also agreement.

TABLE 5 | Sex and age descriptive statistics for the three groups Study 2.

Group	*n*	Sex		Age (*SD*)
		Men	Women	
Political cartoons	54	57.4%	42.6%	20.85 (1.83)
Non-political cartoons	43	46.5%	53.5%	20.84 (1.59)
Newspaper headlines	49	42.9%	57.1%	20.73 (1.60)

FIGURE 2 | Example of cartoon used for study 2.

Instruments

Trust in politicians

We used the same adaptation used in study 1. It presented adequate internal consistency considering Cronbach's Alpha at times 1, 2, and 3 (0.83, 0.84 and 0.89, respectively).

Attention

We asked the participants to rate how much attention they pay to the stimuli after the 1st week (1 = No attention; 100 = Total attention).

Disposition toward politicians

It was assessed with the item "*How much would you say you like politicians?*" with responses ranging from 1 (*Do not like*) to 100 (*Like very much*).

Political affiliation

Participants were asked about their political ideology according to a left–right political spectrum, for which possible responses were "*Left wing,*" "*Center-Left wing,*" "*Center,*" "*Center Right Wing,*" "*Right Wing,*" or "*None of the above.*"

Exposure to political humor

We used the item "*How often do you watch shows or read websites or newspapers that make fun of politicians?*" (1: Almost never; 10: Always).

Exposure to political information

We used the item "*How often do you watch shows or read websites or newspapers that refer to politics?*" (1: Almost never; 10: Always).

Funniness and aversiveness were not assessed. This decision was made due to the characteristics of the design, and given that it would have involved asking participants to rate the stimuli twice a day for 7 days, which could have led to higher

attrition (57.9%). Considering that, we decided to assess trust and attention after the exposure to the stimuli, since funniness and aversion had already been rated by two raters. This is discussed in the limitations sections.

Results

The three groups were not different regarding sex, $\chi^2(2,146) = 2.368$, $p > 0.05$, age, $F(2,45) = 1.287$, $p > 0.05$, $\eta^2 = 0.057$, political affiliation, $F(2,143) = 0.071$, $p > 0.05$, $\eta^2 = 0.001$, or disposition toward politicians, $F(2,143) = 1.176$, $p > 0.05$, $\eta^2 = 0.016$. With this in mind, the decision was made to repeat the analysis of Study 1. In this case, a 3 (condition)-by-3 (time) ANOVA with repeated measures in the last factor was performed. The results showed no significant effect of either condition, $F(2,143) = 0.226$, $p > 0.05$, $\eta_p^2 = 0.003$, time, $F(2,286) = 2.078$, $p > 0.05$, $\eta_p^2 = 0.014$, or the interaction term, $F(4,286) = 0.153$, $p > 0.05$, $\eta_p^2 = 0.002$. The observed power for the three terms was low (0.085, 0.426, and 0.082, respectively). However, this should not be considered as a reason to discard this results since – as it will be discussed in the conclusions section – non-significant results can correspond to low observed power (Hoenig and Heisey, 2001). Means and confidence intervals for each condition in times 1, 2, and 3 can be observed in **Figure 3** (each error bar is constructed using a 95% confidence interval of the mean). Means and standard deviations for each group at times 1, 2, and 3 can be found in **Table 6**.

We performed three supplementary analyses including attention, exposure to political information, and exposure to political humor as covariates in separated models, but we obtained similar results. Specifically, when attention was included, we observed non-significant effects of either condition, $F(2,101) = 0.523$, $p > 0.05$, $\eta_p^2 = 0.010$, time, $F(2,202) = 2.666$, $p > 0.05$, $\eta_p^2 = 0.026$, attention, $F(1,101) = 0.002$, $p > 0.05$, $\eta_p^2 = 0.000$, the interaction term between time and condition, $F(4,202) = 0.128$, $p > 0.05$, $\eta_p^2 = 0.003$, and the interaction term between time and attention, $F(2,202) = 1.861$, $p > 0.05$, $\eta_p^2 = 0.018$. The observed power for the terms was 0.134, 0.525, 0.050, 0.076, and 0.385, respectively. When exposure to political information was included, there were no significant effects of either condition, $F(2,101) = 0.524$, $p > 0.05$, $\eta_p^2 = 0.010$, time, $F(2,202) = 2.885$, $p > 0.05$, $\eta_p^2 = 0.028$, exposure to political information, $F(1,101) = 0.117$, $p > 0.05$, $\eta_p^2 = 0.001$, the interaction term between time and condition, $F(4,202) = 0.157$, $p > 0.05$, $\eta_p^2 = 0.003$, and the interaction term between time and exposure to political information, $F(2,202) = 1.928$, $p > 0.05$, $\eta_p^2 = 0.019$. The observed power for the terms was 0.134, 0.560, 0.063, 0.083, and 0.397, respectively. Finally, when we included exposure to political humor, there were no significant effects of either condition, $F(2,101) = 0.389$, $p > 0.05$, $\eta_p^2 = 0.019$, time, $F(2,202) = 0.595$, $p > 0.05$, $\eta_p^2 = 0.006$, exposure to political humor, $F(1,101) = 2.980$, $p > 0.05$, $\eta_p^2 = 0.029$, the interaction term between time and condition, $F(4,202) = 0.250$, $p > 0.05$, $\eta_p^2 = 0.005$, and the interaction term between time and exposure to political humor, $F(2,202) = 0.987$, $p > 0.05$, $\eta_p^2 = 0.010$. The observed power for the terms was 0.211, 0.148, 0.401, 0.104, and 0.220, respectively.

TABLE 6 | Means and standard deviations for trust times 1, 2, and 3, attention, exposure to political humor and exposure to political information (Study 2).

		M	*SD*
Trust time 1	Political cartoons	28.97	16.79
	Non-political cartoons	27.32	16.98
	Newspaper headlines	26.40	18.06
Trust time 2	Political cartoons	29.20	16.79
	Non-political cartoons	28.59	15.79
	Newspaper headlines	27.42	17.12
Trust time 3	Political cartoons	29.93	16.39
	Non-political cartoons	29.70	16.53
	Newspaper headlines	27.96	17.47
Attention	Political cartoons	81.73	31.08
	Non-political cartoons	80.07	23.09
	Newspaper headlines	73.92	27.44
Exposure to political humor	Political cartoons	5.65	2.69
	Non-political cartoons	4.38	2.54
	Newspaper headlines	4.78	2.26
Exposure to political information	Political cartoons	5.13	1.98
	Non-political cartoons	5.21	2.48
	Newspaper headlines	5.03	2.46

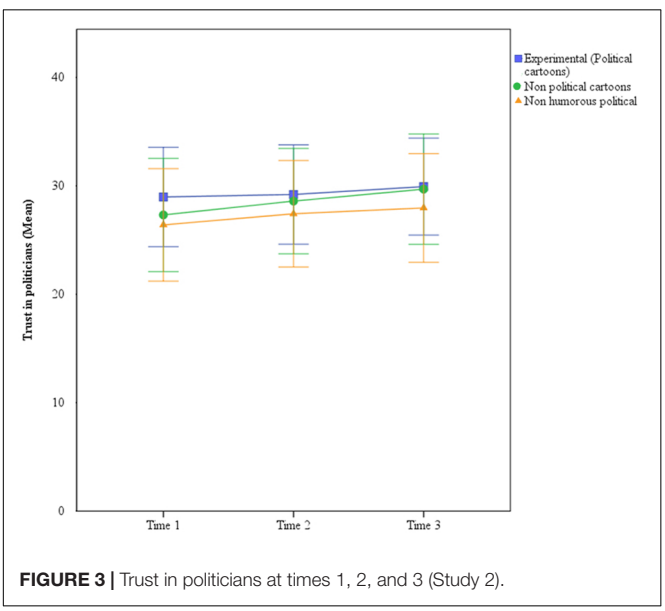


FIGURE 3 | Trust in politicians at times 1, 2, and 3 (Study 2).

CONCLUSION

In general terms, the obtained results point in the expected direction in most cases, but there are at least two elements that are worth considering. According to what was observed in study 1, political disparagement humor has an effect on trust in politicians. However, trust in politicians returns to the same level as the control groups in a second post-exposure measurement. It seems to be that humor can affect attitudes temporarily,

but does not change them permanently. These results are in accordance with earlier findings (Weinberger and Gulas, 1992; Olson et al., 1999). Although the result in the present study was expected, the explanatory pathways of the phenomenon are not clear.

On the one hand, it was hypothesized that the reason for this short-lived effect would be that humor is processed through the peripheral route, understood as less cognitive elaboration. Our results do not support this, since political disparagement humor and non-humorous disparagement political information did not show differences between them regarding the degree of cognitive elaboration. However, both of them showed higher cognitive elaboration than the non-political disparagement humor group. That is to say, humor did imply less cognitive elaboration, but disparagement political humor did not.

Therefore, it is not possible in this case to positively state that the limited durability of the effects of political disparagement humor on attitudes toward politicians can be explained because humor communicates through a peripheral route, decreasing the motivation to counter-argue against the message (Baumgartner and Morris, 2008).

On the other hand, the behavior of two of the groups of study 1 was almost identical. Both the experimental group and the control group exposed to the video with disparagement non-humorous political content showed decreases in the first post-exposure evaluation, being different from their previous measurements and the control group exposed to non-political disparagement humor.

The conclusion to these two aspects seems to be the same: it looks as if political humor is not different from other ways of communicating political content regarding its effects on trust in politicians. This, considering that all the groups were comparable in relation to political affiliation and disposition toward the politicians, would imply that although there could be a positive or a negative disposition toward politicians, disparagement political content has an effect in any form in which it is presented.

It is also necessary to refer to the results of study 2. In this case, there were no effects of political humor on trust in politicians, or any of the relationships that were observed in study 1. Two ideas may help explain these results.

The first one refers to the degree of control that experiments of these characteristics can have. This was not an experiment run in a laboratory, which makes it difficult to assure that participants pay proper attention to the stimuli, for example, even when in this case we did ask participants to rate how much attention they pay to the stimuli.

The second idea refers to theoretical implications of the results of study 2. The type of stimulus used in study 1 was audiovisual, whereas in study 2, only graphic stimuli were used. There may be something in the content and the form of a more complex stimulus that arouses more attention and could therefore generate effects on trust.

There is also the topic of interest in the exposure to political material. It could be thought that forcing a person to consume material daily without any particular motivation would have no effect. In other words, it could be expected that those people who are more interested in consuming political humor could have their attitudes affected (or changed) for a longer period because they would constantly be in contact with stimuli of this kind. As Baumgartner (2013) suggests, it is possible to think that those who are more interested in politics and politicians are not only going to be more interested in consuming information about it, but also would be more interested in consuming political humor and, within it, political disparagement humor recurrently. However, our results do not show an effect of any of these variables on trust in politicians.

This research has limitations. For example, we have considered a measure of cognitive elaboration, but there are other ways of assessing this variable, like thought listing tasks, that could help as a useful complement. Finding other ways of exposing participants to political disparagement humor for longer periods of time would also be useful and could help improving the design of similar experiments.

The validity of the stimuli is essential in an experiment. In this case, we tried to assess such validity by asking two students to rate different elements of the videos and images used in both studies. However, the rating involved dichotomous answers ("yes/no") which could imply not being able to have an idea of the magnitude of possible differences, even though there was complete agreement in every evaluation and the manipulation checks suggest that the stimuli worked properly. It is then possible that the final results could be caused by differences in this magnitude and not the exposition to different stimuli. However, the manipulation checks showed that all the evaluations of the stimuli were as expected and in an expected direction, which is an indicator of a good selection and that the observed effects were very probably caused by the independent variable.

Another evident limitation is not having assessed funniness and aversiveness in study 2. We were forced to make this decision because the design of our experiment would have involved asking participants to rate the stimuli twice a day for a week, which would, we consider, generate higher attrition. Aversiveness and funniness are two basic components of the response to humor, so not considering its impact on participants could involve two things: one, that what we supposed would be disparagement humor was not in fact disparagement (not eliciting more aversiveness than other stimuli) and two, that the stimuli would not be in fact considered funny. Both elements would have an impact on trust, considering our design and the aims of this study. In this case we still had an evaluation of the stimuli given that two raters evaluated them, but not having the participants rate both variables and not being able to control for them (as it was possible in study 1, with clearer results about the role of both variables in the relation between exposure and trust) is a limitation of study 2 that we had to accept.

Finally, a last possible limitation is the low observed statistical power of Study 2. The method used in both studies to determine sample sizes considering a power of 0.80 with GPower was good enough in study 1 but not in study 2. Nevertheless, we think our results are reliable, given that we exceed the minimum sample size when power was computed *a priori*. In addition to this, the *post hoc* procedure of power calculation has been criticized by different authors because it depends on the observed *p*-value

and non-significant *p*-values might correspond to low observed powers (Goodman and Berlin, 1994; Hoenig and Heisey, 2001).

We have seen that the effect is short-lived, but when exactly does disparagement humor stop affecting trust in politicians? Which other variables could help by amplifying or weakening that effect? This research also showed that disparagement political humor was not cognitively processed as non-political humor, which presents an interesting line of research. We think that this research is a step forward, not only considering its results, but also considering the questions that arise from it.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the ethics committee of the University of Santiago with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethics committee of the University of Santiago.

AUTHOR CONTRIBUTIONS

All the authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

AM-S thanks the Chilean Comisión Nacional de Investigación Científica y Tecnológica. Studies 1 and 2 are part of a series of papers funded by the Chilean Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt de Iniciación) Project N°11160661.

ACKNOWLEDGMENT

The authors would like to thank Rolando Zapata, Dr. Mike Hough and Ximena Seguel for their help and useful comments.

REFERENCES

Baumgartner, J. (2013). Internet political ads in 2012: Can humor mitigate unintended effects of negative campaigning? *Soc. Sci. Comput. Rev.* 31, 601–613. doi: 10.1177/0894439313490399

Baumgartner, J., and Morris, J. (2008). One "Nation," under Stephen? The effects of the colbert reporton American youth. *J. Broadcast. Electron. Media* 52, 622–643. doi: 10.1080/08838150802437487

Baumgartner, J., Morris, J., and Walth, N. (2012). The fey effect: young adults, political humor, and perceptions of sarah palin in the 2008 presidential election campaign. *Public Opin. Q.* 76, 95–104. doi: 10.1093/poq/nfr060

Becker, A. (2011). Political humor as democratic relief? The effects of exposure to comedy and straight news on trust and efficacy. *Atl. J. Commun.* 19, 235–250. doi: 10.1080/15456870.2011.622191

Becker, A. (2014). Humiliate my enemies or mock my friends? Applying disposition theory of humor to the study of political parody appreciation and attitudes toward candidates. *Hum. Commun. Res.* 40, 137–160. doi: 10.1111/hcre.12022

Becker, A., and Haller, B. (2014). When political comedy turns personal: humor types, audience evaluations, and attitudes. *Howard J. Commun.* 25, 34–55. doi: 10.1080/10646175.2013.835607

Cacioppo, J. T., and Petty, R. E. (1979). Effects of message repetition and position on cognitive response, recall, and persuasion. *J. Pers. Soc. Psychol.* 37, 97–109. doi: 10.1037/0022-3514.37.1.97

Cacioppo, J. T., and Petty, R. E. (1984). The elaboration likelihood model of persuasion. *Adv. Consum. Res.* 11, 673–675.

Cacioppo, J. T., Petty, R. E., and Stoltenberg, C. D. (1985). "Processes of social influence: the elaboration likelihood model of persuasion," in *Advances in Cognitive-Behavioral Research and Therapy*, ed. P. C. Kendall (San Diego, CA: Academic Press), 215–274.

Cao, X. (2008). Political comedy shows and knowledge about primary campaigns: the moderating effects of age and education. *Mass Commun. Soc.* 11, 43–61. doi: 10.1080/15205430701585028

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2009). GPower3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Ferguson, M. A., and Ford, T. E. (2008). Disparagement humor: a theoretical and empirical review of psychoanalytic, superiority, and social identity theories. *Humor* 21, 283–312. doi: 10.1515/HUMOR.2008.014

Ford, T. E., and Ferguson, M. A. (2004). Social consequences of disparagement humor: a prejudiced norm theory. *Pers. Soc. Psychol. Rev.* 8, 79–94. doi: 10.1207/S15327957PSPR0801_4

Goodman, S. N., and Berlin, J. A. (1994). The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann. Intern. Med.* 121, 200–206. doi: 10.7326/0003-4819-121-3-199408010-00008

Hobden, K., and Olson, J. (1994). From jest to antipathy: disparagement humor as a source of dissonance-motivated attitude change. *Basic Appl. Soc. Psychol.* 15, 239–249. doi: 10.1207/s15324834basp1503_2

Hoenig, J. M., and Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *Am. Stat.* 55, 19–24. doi: 10.1198/000313001300339897

Holbert, R., Lambe, J., Dudo, A., and Carlton, K. (2007). Primacy effects of the daily show and national TV news viewing: young viewers, political gratifications, and internal political self-efficacy. *J. Broadcast. Electron. Media* 51, 20–38. doi: 10.1080/08838150701308002

Igartua, J. J. (2010). Identification with characters and narrative persuasion through fictional feature films. *Communications* 35, 347–373. doi: 10.1515/comm.2010.019

Kenski, K., and Stroud, N. (2006). Connections between internet use and political efficacy, knowledge, and participation. *J. Broadcast. Electron. Media* 50, 173–192. doi: 10.1207/s15506878jobem5002_1

Kim, Y., and Vishak, J. (2008). Just laugh! you dont need to remember: the effects of entertainment media on political information acquisition and information processing in political judgment. *J. Commun.* 58, 338–360. doi: 10.1111/j.1460-2466.2008.00388.x

Kuiper, N., McKenzie, S., and Belanger, K. (1995). Cognitive appraisals and individual differences in sense of humor: motivational and affective implications. *Pers. Individ. Dif.* 19, 359–372. doi: 10.1016/0191-8869(95)00072-e

La Fave, L., and Mannell, R. (1976). Does ethnic humor serve prejudice? *J. Commun.* 26, 116–123. doi: 10.1111/j.1460-2466.1976.tb01914.x

LaMarre, H., Landreville, K., Young, D., and Gilkerson, N. (2014). Humor works in funny ways: examining satirical tone as a key determinant in political humor message processing. *Mass Commun. Soc.* 17, 400–423. doi: 10.1080/15205436.2014.891137

Mackie, D., and Worth, L. (1989). Processing deficits and the mediation of positive affect in persuasion. *J. Pers. Soc. Psychol.* 57, 27–40. doi: 10.1037//0022-3514.57.1.27

McGhee, P. E., and Lloyd, S. A. (1981). A developmental test of the disposition theory of humor. *Child Dev.* 52, 925–931. doi: 10.2307/1129096

Moy, P., Torres, M., Tanaka, K., and McCluskey, M. (2005). Knowledge or trust? Investigating linkages between media reliance and participation. *Commun. Res.* 32, 59–86. doi: 10.1177/0093650204271399

Moy, P., Xenos, M. A., and Hess, V. K. (2006). Priming effects of late-night comedy. *Int. J. Public Opin. Res.* 18, 198–210. doi: 10.1093/ijpor/edh092

Nabi, R., Moyer-Gusé, E., and Byrne, S. (2007). All joking aside: a serious investigation into the persuasive effect of funny social issue messages. *Commun. Monogr.* 74, 29–54. doi: 10.1080/03637750701196896

Olson, J., Maio, G., and Hobden, K. (1999). The (null) effects of exposure to disparagement humor on stereotypes and attitudes. *Humor Int. J. Humor Res.* 12, 195–219. doi: 10.1515/humr.1999.12.2.195

Paletz, D. (1990). Political humor and authority: from support to subversion. *Int. Polit. Sci. Rev.* 11, 483–493. doi: 10.1177/019251219001100406

Petty, R., and Cacioppo, J. (1986). The elaboration likelihood model of persuasion. *Adv. Exp. Soc. Psychol.* 19, 123–205. doi: 10.1016/S0065-2601(08)60214-2

Petty, R. E., and Cacioppo, J. T. (1984a). "Source factors and the elaboration likelihood model of persuasion," in *Advances in Consumer Research*, Vol. 11, ed. T. C. Kinnear (Provo, UT: Association for Consumer Research), 668–672.

Petty, R. E., and Cacioppo, J. T. (1984b). The effects of involvement on responses to argument quantity and quality: central and peripheral routes to persuasion. *J. Pers. Soc. Psychol.* 46, 69–81. doi: 10.1037/0022-3514.46.1.69

Petty, R. E., Cacioppo, J. T., Strathman, A. J., and Priester, J. R. (2005). "To think or not to think: exploring two routes to persuasion," in *Persuasion: Psychological Insights and Perspectives*, 2nd Edn, eds T. C. Brock and M. C. Green (Thousand Oaks, CA: Sage Publications), 81–116.

Sternthal, B., and Craig, C. (1973). Humor in advertising. *J. Mark.* 37, 12–18. doi: 10.2307/1250353

Tajfel, H. (1974). Social identity and intergroup behaviour. *Soc. Sci. Inf.* 13, 65–93. doi: 10.1177/053901847401300204

Tajfel, H. (1979). Individuals and groups in social psychology. *Br. J. Soc. Clin. Psychol.* 18, 183–190. doi: 10.1111/j.2044-8260.1979.tb00324.x

Tajfel, H. (1982). Social psychology of intergroup relations. *Annu. Rev. Psychol.* 33, 1–39. doi: 10.1146/annurev.ps.33.020182.000245

Tajfel, H., and Billig, M. (1974). Familiarity and categorization in intergroup behavior. *J. Exp. Soc. Psychol.* 10, 159–170. doi: 10.1016/0022-1031(74)90064-X

Weinberger, M., and Gulas, C. (1992). The impact of humor in advertising: a review. *J. Advert.* 21, 35–59. doi: 10.1080/00913367.1992.10673384

Wicker, F. W., Barron, W. L. III, and Willis, A. C. (1980). Disparagement humor: dispositions and resolutions. *J. Pers. Soc. Psychol.* 39, 701–709. doi: 10.1037/0022-3514.39.4.701

Wolff, H., Smith, C., and Murray, H. (1934). The psychology of humor: a study of responses to race-disparagement jokes. *J. Abnorm. Soc. Psychol.* 28, 341–365. doi: 10.1037/h0074853

Xenos, M., and Becker, A. (2009). Moments of zen: effects of the daily show on information seeking and political learning. *Polit. Commun.* 26, 317–332. doi: 10.1080/10584600903053569

Yamagishi, T., and Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motiv. Emot.* 18, 129–166. doi: 10.1007/bf02249397

Young, D. G. (2004). *The Counterargument-Disruption Model of Political Humor (CADIMO): An Experimental Exploration of the Effects of Late-Night Political Jokes on Cognitive Elaboration and the Conditional Effects of Partisanship. Annual meeting of the American Political Science Association*. Chicago, IL: The Palmer House Hilton.

Zenner, W. (1970). Joking and ethnic stereotyping. *Anthropol. Q.* 43, 93–113. doi: 10.2307/3316601

Zhang, Y. (1996). Responses to humorous advertising: the moderating effect of need for cognition. *J. Advert.* 25, 15–32. doi: 10.1080/00913367.1996.10673493

Zillmann, D. (1983). "Disparagement humor," in *Handbook of Humor Research*, eds P. McGhee and J. Goldstein (NewYork, NY: Springer-Verlag), 85–107.

Zillmann, D., and Cantor, J. R. (1976). "A disposition theory of humour and mirth," in *Humour and Laughter: Theory, Research and Applications*, eds A. J. Chapman and H. C. Foot (London: Wiley).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Mendiburo-Seguel, Vargas and Rubio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 26 April 2019
doi: 10.3389/fpsyg.2019.00604

The Social Life of Class Clowns: Class Clown Behavior Is Associated With More Friends, but Also More Aggressive Behavior in the Classroom

*Lisa Wagner**

Personality and Assessment, Department of Psychology, University of Zurich, Zurich, Switzerland

A dimensional rather than a typological approach to studying class clown behavior was recently proposed (Ruch et al., 2014). In the present study, four dimensions of class clown behavior (class clown role, comic talent, disruptive rule-breaker, and subversive joker) were used to investigate the associations between class clown behavior and indicators of social status and social functioning in the classroom in a sample of N = 300 students attending grades 6 to 9 (mean age: 13.09 years, 47.7% male). Participants and their teachers completed measures of class clown behavior, and peer nominations of peer acceptance, mutual friends as well as social behavior in the classroom (popular-leadership, aggressive-disruptive, sensitive-isolated, and prosocial behaviors) were collected. The results showed that overall, class clown behavior was positively related to peer acceptance, the number of mutual friends in the classroom and peer-perceived social status. Overall, it was also positively related to peer-rated popular-leadership and aggressive-disruptive behaviors, as well as negatively related to prosocial behaviors. When considering the four dimensions of class clown behavior, comic talent was particularly relevant for the relationship with social status and with popular-leadership behaviors, but also with aggressive-disruptive behaviors. Aggressive-disruptive behaviors were also particularly related to the class clown dimension disruptive rule-breaker. The results underline the significance of class clown behavior for the social status and functioning of students and may help further understand the phenomenon in its multidimensional nature.

Keywords: class clown, humor, school, adolescence, peer relationships, peer acceptance, likeability, disruptive behavior

OPEN ACCESS

Edited by:
Hsueh-Chih Chen,
National Taiwan Normal University, Taiwan

Reviewed by:
Po-Sheng Huang,
National Taiwan University of Science and Technology, Taiwan
Lynn A. Barnett,
University of Illinois at Urbana-Champaign, United States

***Correspondence:**
Lisa Wagner
l.wagner@psychologie.uzh.ch;
lisawagne@gmail.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *02 November 2018*
Accepted: *04 March 2019*
Published: *26 April 2019*

Citation:
Wagner L (2019) The Social Life of Class Clowns: Class Clown Behavior Is Associated With More Friends, but Also More Aggressive Behavior in the Classroom. Front. Psychol. 10:604. doi: 10.3389/fpsyg.2019.00604

INTRODUCTION

Relationships with peers impact well-being throughout the entire life span. However, the influence that peers exert on many areas of life seems to be most pervasive in early adolescence (see Parker et al., 2006). The impact of peer relationships spans from the development of cognitive and social skills, to maladaptive functioning and physical and psychological well-being (e.g., Hartup and Stevens, 1999; Parker et al., 2006; Rubin et al., 2015). Of particular importance for child adolescent

development are peer relationships and social functioning in the classroom (e.g., Berndt and Ladd, 1989). In terms of how to assess social functioning in the classroom, peer nomination procedures such as the Revised Class Play (Masten et al., 1985), have demonstrated their ability to predict important life outcomes, such as academic and job success, social and romantic competence as well as internalizing and externalizing symptoms, for time spans of up to 10 years (Gest et al., 2006). Four areas of social behavior are typically distinguished (e.g., Realmuto et al., 1997; Zeller et al., 2003), two of which are adaptive (popular-leadership and prosocial behavior) and two of which are maladaptive (sensitive-isolated and aggressive-disruptive).

Next to social behavior in the classroom as perceived by the classmates, also the social status, that is peer acceptance and the number of mutual friends in the classroom, is an important indicator of positive peer relationships in adolescence, that has demonstrated its relevance for many important life outcomes. Peer acceptance is a unilateral construct that describes being liked by one's classmates, whereas mutual friends are defined by a bilateral understanding, i.e., friendship nominations by both friends (Rubin et al., 2006; Waldrip et al., 2008; Bagwell and Schmidt, 2011). While many studies have focused on the impact of having versus not having a mutual friend, research also shows that it matters how many mutual friends an adolescent has. For instance, Gest et al. (2001) found the number of friends to uniquely predict prosocial behavior. Students who are highly accepted, that is well-liked, by their peers have been found to show a range of highly desirable characteristics, for example behaving appropriately, communicating well, and being perceived as helpful, cooperative, and good leaders by other students (Rubin et al., 2006). Peer acceptance was also found to be related to academic and athletic accomplishments (e.g., Asher and McDonald, 2009). In addition, being well-accepted by peers and having many friends has also been associated with individual differences in sense of humor (e.g., McGhee, 1989; Wanzer et al., 1996; Gest et al., 2001). In conclusion, both peer acceptance and the number of friends a student has in the class are indicators of social status that go along with a number of important outcomes.

Humor has been identified as a strength of character, and humans tend to find the expression of their signature (i.e., most characteristic) character strengths fulfilling (Peterson and Seligman, 2004). Individuals with the signature strength of humor tend to express humor in their behavior across different contexts and might thus earn a reputation for their humor in their social networks which might result in others using type nouns (like joker, wit, buffoon, or mocking bird) to refer to them (Craik and Ware, 2007). Often people using humor play a function in the institution they are in, such as the "organizational fool" (Kets de Vries, 1990) and the "class clown" (e.g., Damico and Purkey, 1978). Humor is a way of highlighting signs of hubris in leaders, of addressing taboo topics, and of relaxing strained situations. Thus, the institutional fool may satirize leaders and followers. As the truth is spoken in a fun manner, it can trespass on otherwise forbidden territory. Thus, the organizational fool can create a corrective force against the leadership in institutions and be a mediator between leader and followers.

There are comparable circumstances in the classroom where teachers typically socialize children into school culture. The teacher mostly decides who will speak, when, and about what, and also decides what he or she can conceive from the students. Oppositional students – and among these, students considered class clowns – might negotiate power in their classroom communities, and try to resist the set order during classroom lessons (McLaren, 1985; Radigan, 2001; Norrick and Klein, 2008). A class clown may become the opponent of the teacher; poke fun at the teacher's words and action and undermine his authority, in front of the teacher or behind their backs. From the teachers' perspective, students described as class clowns are mostly viewed as difficult students that require being disciplined (see e.g., Cohen and Fish, 1993; Hobday-Kusch and McVittie, 2002). Analyzing accounts of humorous situations in the classroom initiated by students toward teachers, Meeus and Mahieu (2009) found that while testing out, rebellion, and misbehavior were common motives identified, also positive motives, such as humor as atmosphere maker, played a role in these accounts.

The first significant study of class clowns (Damico and Purkey, 1976, 1978) identified 96 mostly male class clowns in a sample of 3,500 eighth graders. Compared to a control sample, class clowns were seen by teachers as significantly higher on asserting behaviors, attention seeking, unruliness, leadership, and cheerfulness, but lower in accomplishing, which was defined as "behaviors leading to successful completion of academic assignments" (p. 393). Relative to other students, the class clowns self-reported seeing school authorities, such as teachers or the principal, less positively, but there was no significant difference in class clowns' attitudes toward classmates, the school in general, and the self. In this study by Damico and Purkey (1978), only students that received 10 or more nominations by their peers were considered a class clown and those that received 25 or more nominations were "super class clowns." However, it must be acknowledged that there was a large variation in the frequency of nominations and more or less arbitrary cut-off points were used to identify the group of class clowns.

Going beyond previous conceptualizations of class clowns as a distinct "type" or categorical concept, Ruch et al. (2014) suggested an alternative approach to study class clown behavior, which is based on a variable-centered or dimensional view. That is, they assume a general dimension of class clown behavior, but also different related facets that can be used to describe class clown behavior further. These can be assessed using the Class Clown Behavior Survey (CCBS; Platt, 2012). The hierarchical model proposes a general factor of class clown behavior (as measured by the total score of the CCBS) as well as four lower-order dimensions of class clown behavior, which are positively correlated with class clown behavior. The first dimension, *class clown role*, consists of being labeled as the class clown by oneself and others (sample item of the CCBS: "My classmates would call me a class clown"). The second dimension, *comic talent*, describes being quick-witted, liking to entertain others with funny things, and spreading a good mood. It describes behavior that is not

necessarily directed against the teacher or questioning the rules at school (sample item: "During class it does not take long until something funny comes into my mind that I can share with the person next to me"). The third dimension, *disruptive rule-breaker*, is characterized by poking fun at the teachers, not taking school rules seriously, and directly challenging the authority (sample item: "Some rules in class I find stupid and I laugh at them"). Similarly, in the fourth dimension, *subversive joker*, the class clown behavior is also aimed at undermining the teachers' authority or directed against classmates. However, it is done behind the teacher's back instead of in a direct confrontation (sample item: "When my teacher turns away, I invent jokes that I write on paper to show it to my classmates").

Ruch et al. (2014) found that all four dimensions went along with higher scores in the character strength of humor, but lower scores in strengths related to restraint, such as self-regulation or prudence. In addition, the dimension *comic talent* went along with higher scores in the strengths of leadership, perspective, zest, social intelligence, bravery, hope, love, and creativity. Platt et al. (2016) added to these results by demonstrating that class clown behavior was associated with lower school satisfaction and lower GPA. When considering the four dimensions, it was only the dimension *disruptive rule-breaker* that related to lower school satisfaction, and only the dimensions *disruptive rule-breaker* and *class clown role* that were negatively related to GPA. On the other hand, the two dimensions *class clown role* and *comic talent* were associated with the experience of more positive emotions in the classroom. Also, while those with high scores on class clown behavior described the relationship with teachers as considerably worse than those with lower scores, the relationships with classmates did not seem to be affected negatively. The dimension *comic talent* had a small but positive correlation with positive relationships with classmates, though it failed to reach statistical significance. Overall, the results of these two studies suggest that different dimensions of class clown behavior can be distinguished, that class clown behavior has both upsides and downsides, and that looking beyond the question of whether or not someone is labeled as "class clown" provides the possibility to gain a deeper insight into the correlates of class clown behavior.

There is some evidence of a relationship between class clown behavior and social status. Damico and Purkey (1978) concluded that their adolescent class clowns were found to have many behaviors and personal assessments in common with adult wits, and among the list of attributes (e.g., being male, leaders, active, independent, creative, and having positive self-perceptions), class clowns are also described as more popular. Likewise, Suitor et al. (2004) report that for male adolescents in private schools, class clowning is a successful route to gain prestige (more so than clothes or car ownership). A recent study by Barnett (2018) looked at the consequences of younger children's playfulness in the classroom. The study found peer-rated social status to be unrelated with self-ratings of being a class clown (which had a rather low mean and variance) and positively related with peer-ratings of being a class clown. This study also underlined the relevance of sex differences: Although the relationships were present in boys and girls in the peer ratings, teachers had a stronger tendency to designate boys as "class clowns" than girls and teacher-rated class clown status was positively related to social status for girls, and negatively related to social status for boys.

THE PRESENT STUDY

The present study investigates the social functioning and social status of students showing class clown behavior. For this purpose, self- and teacher ratings of class clown behavior were considered as they reflect two important perspectives that were previously found to converge moderately at best (see e.g., Barnett, 2018). More concretely, the present study deals with the relationships between self- and teacher-reported class clown behavior and (a) peer acceptance, number of mutual friends in the classroom, and peer-perceived social status as well as (b) peer-perceived classroom behavior describing social functioning. It adds to the existing research on class clown behavior by being the first one to investigate the dimensional approach to class clown behavior in relation to various aspects of social relationships in the classroom, by comparing the contribution of both a "type" approach (considered a class clown or not) and a dimensional approach to predicted social status and social functioning in the classroom, and by also considering teacher ratings on the dimensions of class clown behavior and considering their convergence with self-ratings.

Based on the associations between social status and class clown behavior described previously (e.g., Damico and Purkey, 1978). It was expected that class clown behavior would be positively related to peer acceptance as well as to number of friends and peer-perceived social status. In particular the dimension *comic talent* was expected to show the strongest links with these three variables as previous studies on the dimensional approach (Ruch et al., 2014; Platt et al., 2016) hinted at its relevance for positive (peer) relationships. With regards to social functioning, it was expected that the behavioral dimension *popular-leadership* would positively relate to class clown behavior (again, based on the previously established associations between class clowning and both popularity and leadership behavior, see e.g., Damico and Purkey, 1978; Masten, 1986), and again in particular to the dimension of *comic talent*. Based on previous research demonstrating a negative association between teacher-rated social behavior in the classroom and class clown behavior (Platt et al., 2016), it was expected that the behavioral dimension *prosocial* would be negatively related to class clown behavior. Prosocial behavior as described in the RCP is consistent with other-directed strengths and strengths of restraint. In these areas, students displaying class clown behavior, in particular in the *disruptive rule-breaker* dimension, tended to score lower (Ruch et al., 2014), which supports the present hypothesis. Finally, a positive relationship between class clown behavior and the behavioral dimension of *aggressive-disruptive* was expected. In particular the class clown behavior dimension of *disruptive rule-breaker* was expected to be related to aggressive-disruptive classroom behavior as this dimension

most directly contains disruptive behaviors and it showed strong negative relationships with teacher-rated positive classroom behavior (Platt et al., 2016).

MATERIALS AND METHODS

Participants

A sample of 300 students (47.7% male) aged on average 13.09 years (*SD* = 1.12; ranging from 11 to 17 years) participated in the study. They attended the sixth (28.7%), seventh (38.7%), eighth (21.3%), or ninth (11.3%) grade in nine different schools in German-speaking Switzerland and Liechtenstein. The number of participating students in each classroom ranged between 11 and 26, with an average of 19.40 (*SD* = 3.44). All students in one classroom typically spent (almost) the entire school day together; as a consequence, they can be assumed to be highly familiar with their peers' behavior in the classroom and during different lessons.

Classroom teachers (N = 17) completed the teacher ratings for students in each classroom. These classroom teachers (41.2% male, 11.8% missing information) were on average 39.40 years old (*SD* = 11.31; ranging from 26 to 64 years) and had on average 13.07 years of teaching experience (*SD* = 10.39; ranging from 3 to 40 years). They knew the students they were rating for on average 3.67 semesters (*SD* = 1.29) and were teaching them regularly, for on average 14.67 lessons (*SD* = 8.72) a week.

Instruments

In order to address the research questions, self-reports and teacher ratings on class clown behavior and peer ratings of classroom behavior were collected. In addition, a nomination procedure to assess peer acceptance and the number of mutual friends was used.

Class Clown Behavior

The *Class Clown Behavior Survey* (CCBS; Platt, 2012) is a self-report instrument assessing different class clown behaviors using 18 items with a 6-point answer format (ranging from 1 = *totally disagree* to 6 = *totally agree*). The survey measures class clown behaviors both as total score and in the form of four subscales of *class clown role*, *comic talent*, *disruptive rule-breaker*, and *subversive joker* (encompassing 4 or 5 items each; see Ruch et al., 2014). In the present sample, the total score yielded an internal consistency of $\alpha = 0.94$ and the subscales yielded coefficients of $\alpha = 0.90$, $\alpha = 0.87$, $\alpha = 0.87$, and $\alpha = 0.82$, respectively. Besides analyzing the scores dimensionally, Ruch et al. (2014) suggested identifying class clowns by averaging the items 4 and 9 (i.e., "My classmates would call me a class clown." and "In my class I am the class clown."), and consider those to be class clowns that had scores between 4 (partially agree) and 6 (totally agree). In the present sample the two items correlated highly, $r = 0.84$ ($p < 0.001$), and 18.3% of the participants had scores reaching the "partially agree" cut-off point for being classified as a class clown. In the following analyses, this score was used as *class clown status index*, with values below "partially agree" indicating not being considered a class clown and values above the cut-off point indicating being considered a class clown. Those considered class clowns were 29.4% of the boys and 8.3% of the girls.

For the teacher ratings of class clown behavior, four items consisting of descriptions of each of the four dimensions of class clown behavior as assessed by the CCBS were provided. Classroom teachers rated the extent to which they agreed that these items described a student's typical behavior in the classroom on a 6-point scale (1 = *totally disagree* to 6 = *totally agree*). The total score across the four items yielded an internal consistency of $\alpha = 0.81$. Similar to the self-reports, a class clown status index was computed for the teacher ratings using the item relating to class clown role ("The student would consider him/herself a class clown and is also referred to as class clown by his/her classmates"). Those students who received ratings of 4 (partially agree) or higher were considered class clowns (n = 58, i.e., 19.3% of the participants) in the analyses using the teacher ratings.

Peer Acceptance and Number of Friends

To assess peer acceptance, students were presented with a list of all their classmates and were asked to select all classmates they liked. The instructions read: "Please select those classmates that you like. These could be those that you like spending your breaks with or that you enjoy sitting next to." They were also instructed that they could choose not to nominate any of their classmates. To determine a students' peer acceptance, the number of received nominations was then adjusted for classroom size (i.e., divided by the number of participating students in the classroom minus 1). Consequently, the index for peer acceptance could range from 0 to 1. To determine the number of mutual friends in the classroom, students were asked to nominate their peers in the classroom whom they considered their friends by selecting them from a list of all students in the class, with a maximum of five nominations. A friendship was considered to be mutual when both friends had nominated each other. Thus, the number of mutual friends could range between 0 and 5.

Social Functioning in the Classroom

The *Revised Class Play* (Masten et al., 1985) was used to assess social behavior in the classroom. Students were presented with short behavior descriptions (e.g., someone who helps others) and were asked to nominate those students in their class that would be best suited to play this role in a hypothetical play. They did not have to nominate anyone and could nominate as many students as they wanted. To take different class sizes into account, the peer nominations a student received in each classroom were standardized. A German version of the instrument was developed using a standard translation-backtranslation procedure (Brislin, 1970). Four scales were computed according to Zeller et al. (2003): Popular-leadership (10 items), aggressive-disruptive (4 item; due to a negative corrected item-total correlation, the item "teases others" was deleted), sensitive-isolated (6 items), and prosocial

(3 items; due to negative corrected item-total-correlations, the items "waits turn" and "is trustworthy" were deleted), yielded internal consistency coefficients of $\alpha = 0.90$, $\alpha = 0.85$, $\alpha = 0.71$, and $\alpha = 0.85$, respectively. The dimension *popular-leadership* contains items referring to sociable behavior with peers (e.g., someone who makes new friends easily) as well as to leadership skills (e.g., someone everyone listens to). *Aggressive-disruptive* behavior includes items in relation to disruption in the peer group (e.g., someone who fights a lot) and aggression toward others (e.g., someone who picks on others). The dimension *sensitive-isolated* consists of items that relate to withdrawn behavior characterized by difficulty interacting with peers (e.g., someone who is often left out). Finally, the dimension *prosocial* includes items that describe good manners with peers (e.g., someone who helps others). In addition to these four dimensions, a scale called *peer-perceived social position* was also considered as an additional indicator of social position, as suggested by Gest et al. (2001). It includes the items "has many friends," "everyone likes to be with," "has trouble making friends" (reverse-scored), and "is often left out" (reverse-scored) and yielded an internal consistency of $\alpha = 0.78$ in the present sample.

Procedure

Data for this study were collected in a classroom setting using school computers to complete the questionnaires that were presented online. Trained research assistants oversaw the completion of the questionnaires. Questionnaires were presented in two blocks, one block contained the nomination procedure for the assessment of peer acceptance and number of friends and the second block contained the Revised Class Play and the CCBS. The two blocks were presented in a randomized order. Classroom teachers completed the ratings at the same time or shortly after the student data had been collected also using an online questionnaire. The data presented here were collected as a part of a larger project and overlap with the sample used in Wagner (2018), which covers different variables and research questions. In total, students took between two and three lessons (i.e., between 90 and 135 min), including breaks, to complete all questionnaires.

RESULTS

Preliminary Analyses

In preliminary analyses, it was tested whether the variables of interest were related to sex and age. **Table 1** shows the descriptive statistics, both for the total sample and separately for boys and girls, as well as the correlations with age. **Table 1** also shows the results of independent samples *t*-tests to test whether boys and girls differed in the studied variables.

In line with previous findings (Ruch et al., 2014), boys had higher scores on all of the four self-rated class clown dimensions, with effect sizes ranging between medium and large effects (see **Table 1**). This result was also found for the teacher ratings. Participants' age was negatively related to peer acceptance, i.e., younger students were nominated more frequently as being liked by their classmates. As a consequence of these differences, sex and age were controlled for in the main analyses.

The teacher ratings of class clown behavior converged moderately with the self-reported CCBS scales ($\rho = 0.42$, $\rho = 0.42$, $\rho = 0.36$, and $\rho = 0.35$; all $p < 0.001$). The mean score across the four ratings correlated highly with the CCBS total score ($\rho = 0.53$; $p < 0.001$). The dimension *popular-leadership* was positively related to peer acceptance, r (296) = 0.68, and to the number of mutual friends, r (292) = 0.46, both $p < 0.001$, when controlling for influences of sex and age.

TABLE 1 | Descriptive statistics and correlations with sex and age for study variables.

	Total sample			Boys		Girls		Comparison	
	M	*SD*	r_{Age}	*M*	*SD*	*M*	*SD*	*t*	*d*
Class Clown Behavior Survey (Self-report)									
Total score	2.56	1.05	0.04	2.86	1.08	2.29	0.95	4.90***	0.57
Class clown role	2.24	1.25	− 0.04	2.65	1.35	1.87	1.02	5.61***	0.69
Comic talent	3.46	1.29	0.10	3.74	1.26	3.21	1.27	3.68***	0.43
Disruptive rule breaker	2.24	1.18	0.05	2.52	1.26	1.98	1.06	3.95***	0.47
Subversive joker	2.12	1.04	− 0.03	2.37	1.10	1.90	0.93	4.04***	0.47
Class Clown Teacher Rating									
Total score	2.05	1.08	0.07	2.35	1.17	1.78	0.91	4.72***	0.58
Class clown role	1.97	1.41	0.07	2.40	1.59	1.59	1.09	5.10***	0.65
Comic talent	2.72	1.54	0.08	3.08	1.67	2.39	1.34	3.91***	0.47
Disruptive rule breaker	1.76	1.18	0.07	1.91	1.22	1.62	1.13	2.10*	0.24
Subversive joker	1.75	1.20	0.01	2.01	1.35	1.50	1.00	3.70***	0.46
Nomination procedure									
Peer acceptance	0.37	0.15	− 0.16**	0.37	0.15	0.37	0.15	0.13	0.02
Mutual friends	2.83	1.44	0.02	2.92	1.51	2.75	1.36	1.00	0.12

*N = 296–300 (total sample). Boys: n = 141–143. Girls: n = 155–157. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.*

The dimensions *aggressive-disruptive* and *sensitive-isolated* were negatively related to peer acceptance, r (296) = –0.19 and r (296) = –0.52, both $p < 0.001$, and the number of mutual friends, *aggressive-disruptive*: r (292) = –0.15, $p = 0.009$, *sensitive-isolated:* r (290) = –0.35, $p < 0.001$. The dimension *prosocial* was positively related to both aspects, r (296) = 0.30, $p < 0.001$, and r (292) = 0.17, $p = 0.003$ (all correlations controlled for sex and age). These findings replicate the results reported by Zeller et al. (2003), supporting the validity of the version of the measure used in the present study.

Relationships of Class Clown Behavior With Social Status

Since the data had a nested structure (students nested in classrooms and schools), the lme4 package (Bates et al., 2015) in R (R Core Team, 2013) was used to compute multilevel random coefficients models. We tested the expected relationships in three-level random-intercept models, which means that a different intercept was estimated for every classroom (within every school). The standardized coefficients for the four class clown behavior scales (both in self-and teacher-ratings) predicting the different indicators of social status in the classroom (peer acceptance, number of mutual friends, and peer-perceived social position) while controlling for sex and age are displayed in **Table 2**.

As shown in **Table 2**, both the total score of the self-rating of class clown behavior and the total score of the teacher rating of class clown behavior were positively associated with all three indicators of social status (all $p < 0.05$). On the level of subscales, peer acceptance was positively related with all self-rated class clown behavior dimensions and with teacher-rated comic talent. The number of mutual friends was positively correlated with the two self-rated subscales of *identified as a class clown* and *comic talent* as well as the teacher-rated *comic talent* and *subversive joker*. Peer-perceived social position yielded positive correlations across all class clown behavior dimensions in both self- and teacher ratings.

However, the associations found with class clown status index (see **Table 2**) suggest that some of the relevant information is already explained by the fact whether someone sees him- or herself as a class clown or not. Class clowns tended to have higher peer acceptance, a higher number of mutual friends, and a higher peer-perceived social status than those that do not identify as class clowns. To determine the unique contribution of each class clown behavior dimension above the binary variable class clown status index and the other dimensions all predictors were entered simultaneously (separately for self-ratings and teacher-ratings of class clown behavior). As predictors, the covariates sex and age, the class clown status index (0 = not considered a class clown, 1 = considered a class clown) and the dimensions of class clown behavior were entered. The results of these analyses are presented in **Table 3**.

To avoid problems with multicollinearity, the self-reported dimension of class clown role was excluded from these analyses since two of the four items forming the scale were also used to build the class clown status index resulting in a high correlation between the dimension class clown role and the class clown status index, $r(300) = 0.76$. As shown in **Table 3**, only the dimension of comic talent uniquely predicted peer acceptance, number of mutual friends and peer-perceived social position. The class clown status index as well as the other dimensions of class clown

TABLE 2 | Results of multilevel models predicting peer acceptance, number of mutual friends, and peer-perceived social status (controlled for age and sex): Standardized coefficients for the fixed effects of the predictors (Class Clown Behavior Survey and Class Clown Teacher Rating) of the respective random-intercept models for each predictor entered separately.

	Peer Acceptance	Number of Mutual Friends	Peer-Perceived Social Status
Class Clown Behavior Survey (Self-report)			
Total score	0.22***	0.16**	0.36***
Class clown role	0.20***	0.18**	0.34***
Comic talent	0.22***	0.17**	0.36***
Disruptive rule breaker	0.17**	0.11	0.26***
Subversive joker	0.18**	0.09	0.030***
Class clown status index (no = 0/yes = 1)	0.13*	0.15*	0.18***
Class Clown Teacher Rating			
Total score	0.17**	0.16**	0.34***
Class clown role	0.10	0.08	0.23***
Comic talent	0.26***	0.23***	0.36***
Disruptive rule breaker	0.06	0.05	0.23***
Subversive joker	0.08	0.12	0.23***
Class clown status index (no = 0/yes = 1)	0.05	0.04	0.16**

*N = 296–300. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.*

TABLE 3 | Results of multilevel models predicting peer acceptance, number of mutual friends, and peer-perceived social status (controlled for age and sex): Standardized coefficients for the fixed effects of the predictors of the respective random-intercept models for each block of predictors (self- and teacher-ratings of class clown behavior) entered simultaneously.

	Peer Acceptance	Number of Mutual Friends	Peer-Perceived Social Status
Class Clown Behavior Survey (Self-report)			
Class clown status index (no = 0/yes = 1)	0.05	0.11	0.04
Comic talent	0.17*	0.18*	0.28***
Disruptive rule breaker	0.00	0.00	– 0.02
Subversive joker	0.05	– 0.07	0.11
Class Clown Teacher Rating			
Class clown status index (no = 0/ yes = 1)	– 0.10	– 0.09	– 0.09
Comic talent	0.30***	0.25**	0.34***
Disruptive rule breaker	0.00	– 0.07	0.10
Subversive joker	0.01	0.11	0.07

*N = 296–300. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.*

behavior did not show any significant relationships with the outcomes in these analyses.

It was also tested whether the presented relationships were moderated by sex by including an interaction term between the class clown dimension and sex as additional predictor. No moderation effects were observed for any of the combinations of class clown dimensions and indicators of social status (all $p > 0.05$).

Relationships of Class Clown Behavior With Social Functioning in the Classroom

Table 4 shows the standardized coefficients for the fixed effects in the random-intercept models predicting the four dimensions of social behavior in the classroom (popular-leadership, aggressive-disruptive, sensitive-isolated, and prosocial) as nominated by peers from self- and teacher-reported class clown behavior.

The pattern of associations displayed in **Table 4** shows that, the class clown behavior *total score* was positively related to the scales *popular-leadership* and *aggressive*. It was unrelated to the scale *sensitive-isolated* and negatively related to *prosocial*. The class clown behavior dimensions *comic talent* and *subversive joker* showed consistent positive correlations to *popular-leadership* across both self- and teacher ratings. *Aggressive-disruptive* classroom behavior was consistently related to all dimensions of class clown behavior. The *sensitive-isolated* scale was unrelated to the class clown behavior dimensions. There were medium-sized negative correlations between all class clown dimensions and *prosocial* classroom behavior.

Again, the *class clown status index* seemed to carry some of the relevant variance. Considering oneself to be a class clown yielded higher scores in peer-rated popular and aggressive classroom behavior and lower scores in prosocial behavior. To determine the unique contribution of each class clown behavior dimension above the binary variable class clown status index and the other dimensions, all predictors were entered simultaneously (excluding the dimension of self-rated class clown role) for each of the classroom behavior dimensions. The results of these analyses are presented in **Table 5**.

Table 5 shows that when considering self-reported class clown behavior, comic talent uniquely predicted the dimension of *popular-leadership*. *Aggressive-disruptive* was predicted by both the class clown status index and comic talent. *Sensitive-isolated* was not predicted by class clown behavior, and *prosocial* was uniquely negatively related to the dimension of disruptive rule-breaker. When considering teacher-rated class clown behavior, comic talent uniquely predicted *popular-leadership*. Class clown status index and disruptive rule-breaker both predicted *aggressive-disruptive* behavior. *Sensitive-isolated* was uniquely negatively related with comic talent, and *prosocial* was negatively related with both class clown status index and disruptive rule-breaker.

To get a clearer picture of the relationships of class clown behavior with aggressive-disruptive classroom behavior, an additional exploratory analysis was performed. In this analysis, the extent to which the class clown status index and the class clown behavior dimensions predicted the individual items of the RCP scale was inspected by multilevel models parallel to those performed on the full scale. These analyses showed that using the self-rated class clown behavior dimensions, all items were predicted by the class clown status index (all $p < 0.05$). When using the teacher-rated class clown behavior dimensions, all items were predicted by the class clown status index and the dimension of disruptive rule-breaker (all $p < 0.05$). However, the item "too bossy" was additionally predicted by the dimension of comic talent in both self- and teacher ratings ($p < 0.05$).

It was also tested whether these relationships were moderated by sex by including an interaction term between the class clown dimension and sex as predictors for all four dimensions of classroom behavior as criteria. Overall, only two combinations of class clown dimensions and classroom behavior dimensions yielded a moderating effect of sex ($p < 0.05$); the positive relationship between *comic talent* and aggressive-disruptive behavior was stronger for boys than for girls and the positive relationship between *disruptive rule-breaker* and popular-leadership was stronger for girls than for boys (all analyses controlling for age).

DISCUSSION

The present study used different data sources (self-reports, peer nominations, and teacher ratings) to investigate how different dimensions of class clown behavior relate to the social status and social functioning of the students habitually displaying such behaviors. Overall, the results underline the relevance of class clown behavior for social functioning in the classroom. Like in prior studies (Ruch et al., 2014; Platt et al., 2016), the different dimensions of class clown behaviors (in particular comic talent and aggressive-disruptive) differentially affected outcome measures.

With regard to the first hypothesis, as expected, class clown behavior both from the perspectives of the students and the teachers generally went along with higher social status, that is being well-liked and having many friends as well as being perceived as well-liked and having many friends by one's classmates. When considering the overlap between the different dimensions of class clown behavior, *comic talent* was the most relevant one carrying the strongest associations with social status. Those who like to entertain their classmates with funny things and are quick-witted are well-accepted in the classroom, have many friends, and have a reputation for being well-liked and having many friends. As expected in the second hypothesis, class clown behavior was also related to higher scores on the dimensions popular-leadership. Again, the dimension of *comic talent* uniquely predicted the classroom behavior dimension of popular-leadership when considering the overlap between the dimensions and thus seems to be most relevant when predicting this behavior. Those who express humor in the classroom by sharing funny things with their classmates also tend to be considered leaders in the classroom.

Also the third and fourth hypotheses were confirmed, namely class clown behavior was associated with higher scores on

the classroom behavior dimension of aggressive-disruptive, and lower scores on the dimension prosocial. As predicted, the dimension of *disruptive rule-breaker* was most relevant when explaining differences in aggressive-disruptive behavior (for teacher-rated class clown behavior) and in (low) prosocial behavior. Those students who like to mock the teachers and to poke fun at the school rules are characterized as showing aggression in the classroom and as not being polite and helpful (cf. Platt et al., 2016). These findings are corroborated by the generally high convergence between the analyses using self- and those using teacher-rated class clown behavior.

Taken the results concerning the four hypotheses together, class clown behavior went along with positive (higher scores on popular-leadership), the absence of positive (lower scores on prosocial), and negative (higher scores on aggressive-disruptive) aspects of social behavior in the classroom. In light of the large amount of research showing links between social status and different forms of aggression (for a review, see e.g., Heilbron and Prinstein, 2008), this co-occurrence is not surprising. It might be interesting to look at class clown behavior in more detail as an example of a type of behavior that seems to contribute to both social status and aggressive behavior in the classroom.

Exploratory analyses were conducted on each of the items of the aggressive-disruptive scale of the RCP to generate ideas for a more detailed understanding of the relationships between the different dimensions of class clown behavior and the display of aggressive behavior in the classroom. These results suggest that while the class clown status is related with aggressive-disruptive behavior in general, the dimensions predicted the items differentially and different kinds of class clown behavior seem to involve different kinds of aggressive behaviors – the comic talents are perceived as dominant and self-opinionated ("too bossy") and the disruptive rule-breakers are perceived as verbally and/or physically aggressive ("picks on others" and "gets into fights"). In future studies, it might be worthwhile to look at different forms of aggressive behavior in more detail to gain a deeper understanding of these relationships.

The current results clearly support the usefulness of a dimensional approach when compared to a typological approach in studying class clown behavior. Within the dimensional approach, it seems that the dimensions of comic talent and disruptive rule-breaker showed clearly different patterns of associations, whereas the dimension of subversive joker did not emerge as unique predictor of any of the variables in the present study. Future research will be needed to critically examine whether it can predict other variables beyond the other dimensions. When comparing the two approaches, the dichotomous variable *class clown status index*, which categorized students into "class clowns" and "not-class clowns," also showed relations with the studied variables. This finding underlines that it matters whether or not a student perceives him- or herself as a class clown. However, when entered together with the dimensions of class clown behavior, the dimensions – in particular *comic talent* and *disruptive rule-breaker* – mostly outperformed the class clown status index in predicting the outcomes of interest. This shows that while the label "class clown" does have some relevance, the more powerful distinction is which kind of class clown behavior a student shows. For future research, it seems to be promising to move beyond studying "class clowns" as compared to "not-class clowns," which also requires somewhat arbitrary cut-offs, and to consider class clown behavior as a dimensional *and* multidimensional phenomenon.

With respect to sex differences, the present study replicates previous findings regarding the higher prevalence of class clown behavior among boys compared to girls. Regarding our substantive research questions, there was little evidence of moderating effects of sex in the studied relationships. The very few sex differences are nonetheless in line with previous

TABLE 4 | Results of multilevel models predicting dimensions of classroom behavior, as assessed by the revised class play (controlled for age and sex): Standardized coefficients for the fixed effects of the predictors (Class Clown Behavior Survey and Class Clown Teacher Rating) of the respective random-intercept models for each predictor entered separately.

	Popular-leadership	Aggressive-disruptive	Sensitive-isolated	Prosocial
Class Clown Behavior Survey (Self-report)				
Total score	0.23***	0.26***	– 0.09	– 0.27***
Class clown role	0.22***	0.26***	– 0.04	– 0.26***
Comic talent	0.23***	0.24***	– 0.09	– 0.21***
Disruptive rule breaker	0.16**	0.22***	– 0.06	– 0.26***
Subversive joker	0.17**	0.17**	– 0.10	– 0.21***
Class clown status index (no = 0/ yes = 1)	0.16**	0.27***	0.05	– 0.21***
Class Clown Teacher Rating				
Total score	0.23***	0.42***	0.05	– 0.39***
Class clown role	0.11	0.36***	0.08	– 0.30***
Comic talent	0.25***	0.28***	– 0.06	– 0.21***
Disruptive rule breaker	0.04	0.36***	0.09	– 0.42***
Subversive joker	0.14*	0.33***	0.08	– 0.36***
Class clown status index (no = 0/yes = 1)	0.07	0.31***	0.09	– 0.26***

*N = 300. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.*

TABLE 5 | Results of multilevel models predicting dimensions of classroom behavior, as assessed by the revised class play (controlled for age and sex): Standardized coefficients for the fixed effects of the predictors of the respective random-intercept models for each block of predictors (self- and teacher-ratings of class clown behavior) entered simultaneously.

	Popular-leadership	Aggressive-disruptive	Sensitive-isolated	Prosocial
Class Clown Behavior Survey (Self-report)				
Class clown status index (no = 0/yes = 1)	0.08	0.21**	0.12	– 0.10
Comic talent	0.21*	0.17*	– 0.07	– 0.07
Disruptive rule breaker	– 0.02	0.09	0.03	– 0.20*
Subversive joker	0.02	– 0.10	– 0.12	0.04
Class Clown Teacher Rating				
Class clown status index (no = 0/yes = 1)	– 0.06	0.11	0.12	– 0.04
Comic talent	0.27***	0.11	– 0.16*	– 0.02
Disruptive rule breaker	– 0.13	0.17*	0.06	– 0.31***
Subversive joker	0.15	0.13	0.05	0.13

*N = 300. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.*

findings (e.g., Barnett, 2018): Boys showed stronger associations of class clown behavior with negative outcomes and weaker associations with positive outcomes than girls did. This might be due to teachers and peers perceiving humorous behavior in the classroom more negatively in boys than in girls as suggested by Barnett (2018). It has to be noted though that a comparison to the findings by Barnett (2018) is hampered by the use of different age groups. The sample in Barnett's study was on average 9 years old at the last data collection, while the present sample consisted of adolescents who were on average 13 years old. It can be assumed that class clown behavior itself, as well as its perception by teachers and peers and its correlates, changes with age and also by the type of school a student attends. Studies using large samples from different age groups as well as additional longitudinal studies are needed to enhance our understanding of these processes.

Similarly, the perception of classroom behavior and its consequences seems to vary depending on the perspective (self, teacher, or peer). In the present study, self- and teacher-ratings of class clown behavior were considered. They converged moderately and also showed a generally similar pattern of results, even though the convergence was not perfect. Several reasons for this are conceivable. First, some of the behaviors are addressed toward the peers and might thus be less visible for the teacher. Second, in particular in the case of the class clown behavior dimension of subversive joker, students poke fun at the teachers behind their backs – so it should be more difficult for them to observe the behavior. Third, even though teachers in the present study were teaching students for a significant amount of lessons in a week, there were in most cases also other teachers who were teaching in the respective classroom, so one teacher would not be able to observe behavior in all lessons and with all teachers. In future studies, it would be interesting to also assess peer ratings on class clown behavior. In general, it seemed that in the teacher ratings the distinction between "positive" (comic talent) and "negative" class clown behavior was amplified. For instance, teacher-rating on the dimension *disruptive rule-breaker* were not related to peer acceptance or *popular-leadership* classroom behavior, while there were positive relationships with self-ratings of this dimension. Taken together with the observation that the means of the teacher ratings were generally lower than those of the self-ratings, it might be the case that teachers had a higher threshold for noticing or describing class clown behaviors, in particular disruptive ones, and thus their ratings might be more sensitive for more extreme behaviors. Future research might benefit systematically comparing self-, teacher- and peer-reports of class clown behavior, but also from extending beyond those perspectives. One approach could be observing distinct behaviors perceived as class clown behavior instead of generalized dimensions. The study of such distinct behaviors might lead to a clearer understanding what is perceived as class clown behavior, how different behaviors are appreciated, and how classmates and teachers react to different kinds of class clown behavior.

When interpreting the present results, it might also be useful to consider the different profiles of character strengths that have been found to be associated with the different dimensions of class clown behavior (Ruch et al., 2014). Character strengths as described in the VIA classification (Peterson and Seligman, 2004) represent a family of positive traits that can contribute to a "good life," and with that, they are related to a number of positive outcomes, including positive relationships. Recently, Wagner (2018) identified a number of character strengths as most relevant to adolescents' peer relationships and friendships in the classroom. The class clown behavior dimension of comic talent was found to be associated with a number of character strengths that overlap with those identified as most relevant for peer relationships; most notably perspective, love, social intelligence, leadership, and (naturally) humor (Ruch et al., 2014). While Ruch et al. (2014) found that the other three dimensions of class clown behavior were also associated with the character strength of humor (though to a lesser degree), they were not associated with most of the strengths mentioned and even displayed some negative correlations with strengths that have been identified as instrumental for social functioning in the classroom, such as honesty or teamwork. Future research might also aim to understand which additional individual differences underlie the different dimensions of class clown behavior. For instance, does

being a "comic talent" and showing quick-witted humor behavior in fact go along with high (verbal) intelligence (cf. Masten, 1986)?

Some limitations of the present study need mentioning. Firstly, the reported results are cross-sectional associations, not allowing for any conclusions regarding directionality or causality. Secondly, the measure used to assess teacher perceptions of class clown behavior was based on single items, limiting the reliability of the assessment. Thirdly, the psychometric properties of the Revised Class Play scales were not consistently desirable. The German translation used has not been validated previously, and thus these results need to be interpreted with some caution. Finally, while there is initial evidence on the validity of the CCBS (Ruch et al., 2014; Platt et al., 2016), further work is needed to corroborate its validity and to also test it systematically in different age groups.

CONCLUSION

The present study underlines that displaying humor in the classroom in the form of class clown behavior has both upsides add downsides for the individual. In general, humor is related not only to a life of pleasure, but also to an orientation to positive relationships (Wagner et al., 2019). Our results show that this is also true for class clown behavior – in particular one kind of class clown behavior, the comic talent. This dimension was found to be uniquely related with being well accepted by one's classmates and having many friends in the classroom, which are important aspects of positive peer relationships in adolescence. Teachers confronted with the expression of humor in the classroom might thus benefit from focusing on its positive effects (i.e., on relationships in the classroom and on the atmosphere, see also Meeus and Mahieu, 2009). There are, however, also clear downsides to class clown behavior with respect to social functioning in the classroom, in the form of aggressive behavior or low prosocial behavior.

The present results support a dimensional approach looking beyond assuming versus not assuming the role of a class clown. While the dimension of disruptive rule-breaker is related to various downsides and the dimension of comic talent seems to have many upsides when it comes to social functioning in the classroom, comic talent also went along with more aggressive behavior. Class clown behavior and its relationship with social functioning in the classroom, it seems, is not a black and white issue.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Swiss Psychological Association and with the Declaration of Helsinki. All participants provided written informed consent. In line with local guidelines, participants under the age of 14 years also provided the written informed consent of a parent or legal guardian. The local ethics committee at the University of Zurich also approved the procedures (including the consent procedures) before the start of the study.

AUTHOR CONTRIBUTIONS

LW designed the study, supervised data collection, conducted the data analysis, and wrote the manuscript.

ACKNOWLEDGMENTS

The author would like to thank Willibald Ruch for his help in designing the study. Andrea Meier and Benedikt Meier for their help with collecting the data. Willibald Ruch, Fabian Gander, and Mara Stewart for helpful comments on earlier versions of the manuscript as well as all participating students and their teachers.

REFERENCES

Asher, S. R., and McDonald, K. L. (2009). "The behavioral basis of acceptance, rejection, and perceived popularity," in *Social, Emotional, and Personality Development in Context. Handbook of Peer Interactions, Relationships, and Groups*, eds K. H. Rubin, W. M. Bukowski, and B. Laursen (New York, NY: Guilford Press), 232–248.

Bagwell, C. L., and Schmidt, M. E. (2011). *Friendships in Childhood and Adolescence*. New York, NY: Guilford Press.

Barnett, L. (2018). The education of playful boys: class clowns in the classroom. *Front. Psychol.* 9:232. doi: 10.3389/fpsyg.2018.00232

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Berndt, T. J., and Ladd, G. W. (eds) (1989). *Peer Relationships in Child Development*. Oxford: John Wiley & Sons.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *J. Cross Cult. Psychol.* 1, 185–216. doi: 10.1177/135910457000100301

Cohen, J. J., and Fish, M. C. (1993). *Handbook of School-based Interventions: Resolving Student Problems and Promoting Healthy Educational Environments*. San Francisco, CA: Jossey-Bass.

Craik, K. H., and Ware, A. P. (2007). "Humor and personality in everyday life," in *The Sense of Humor. Explorations of a Personality Characteristic*, ed. W. Ruch (Berlin: Mouton de Gruyter), 63–94.

Damico, S. B., and Purkey, W. W. (1976). "The class clown phenomenon among middle school students," in *Paper presented at the Annual Meeting of the American Educational Research Association*, (San Francisco, CA), 19–23.

Damico, S. B., and Purkey, W. W. (1978). Class clowns: a study of middle school students. *Am. Educ. Res. J.* 15, 391–398. doi: 10.3102/00028312015003391

Gest, S., Sesma, A. Jr., Masten, A., and Tellegen, A. (2006). Childhood peer reputation as a predictor of competence and symptoms 10 years later. *J. Abnorm. Child Psychol.* 34, 507–524. doi: 10.1007/s10802-006-9029-8

Gest, S. D., Graham-Bermann, S. A., and Hartup, W. W. (2001). Peer experience: common and unique features of number of friendships, social network centrality, and sociometric status. *Soc. Dev.* 10, 23–40. doi: 10.1111/1467-9507.00146

Hartup, W. W., and Stevens, N. (1999). Friendships and adaptation across the life span. *Curr. Dir. Psychol. Sci.* 8, 76–79. doi: 10.1111/1467-8721.00018

Heilbron, N., and Prinstein, M. J. (2008). A review and reconceptualization of social aggression: adaptive and maladaptive correlates. *Clin. Child Fam. Psychol.* 11, 176–217. doi: 10.1007/s10567-008-0037-9

Hobday-Kusch, J., and McVittie, J. (2002). Just clowning around: classroom perspectives on children's humour. *Can. J. Educ.* 27, 195–210. doi: 10.2307/1602220

Kets de Vries, M. (1990). The organizational fool: balancing a leader's hubris. *Hum. Relat.* 43, 751–770. doi: 10.1177/001872679004300804

Masten, A. S. (1986). Humor and competence in school-aged children. *Child Dev.* 57, 461–473. doi: 10.2307/1130601

Masten, A. S., Morison, P., and Pellegrini, D. S. (1985). A revised class play method of peer assessment. *Dev. Psychol.* 21, 523–533. doi: 10.1037/0012-1649.21.3.52

McGhee, P. E. (1989). Chapter 5: the contribution of humor to children's social development. *J. Child. Cont. Soc.* 20, 119–134. doi: 10.1300/J274v20n01_09

McLaren, P. L. (1985). The ritual dimensions of resistance: clowning and symbolic inversion. *J. Educ.* 167, 84–97. doi: 10.1177/002205748516700208

Meeus, W., and Mahieu, P. (2009). You can see the funny side, can't you? Pupil humour with the teacher as target. *Educ. Stud.* 35, 553–560. doi: 10.1080/03055690902880224

Norrick, N. R., and Klein, J. (2008). Class clowns: talking out of turn with an orientation toward humor. *Lodz Papers Pragmatics* 4, 83–107. doi: 10.2478/v10016-008-0002-6

Parker, J. G., Rubin, K. H., Erath, S. A., Wojslawowicz, J. C., and Buskirk, A. A. (2006). "Peer relationships, child development, and adjustment: a developmental psychopathology perspective," in *Developmental Psychopathology*, eds D. Cicchetti and D. J. Cohen (Hoboken, NJ: John Wiley & Sons, Inc), 419–493.

Peterson, C., and Seligman, M. E. P. (2004). *Character Strengths and Virtues: A Handbook and Classification*. New York, NY: Oxford University Press.

Platt, T. (2012). *Class Clown Behavior Survey*. Unpublished Research Instrument, University of Zurich, Switzerland.

Platt, T., Wagner, L., and Ruch, W. (2016). The association between class clown dimensions, school experiences and accomplishment. *Learn. Individ. Differ.* 51, 327–332. doi: 10.1016/j.lindif.2016.08.036

R Core Team (2013). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria*. Available at: http://www.R-project.org/

Radigan, J. (2001). The class clown and negotiated epistemologyin a classroom. *Crit. Ethnogr. Educ.* 5, 61–88. doi: 10.1016/S1529-210X(01)80005-X

Realmuto, G. M., August, G. J., Sieler, J. D., and Pessoa-Brandao, L. (1997). Peer assessment of social reputation in community samples of disruptive and nondisruptive children: utility of the revised class play method. *J. Clin. Child Psychol.* 26, 67–73. doi: 10.1207/s15374424jccp2601_7

Rubin, K. H., Bukowski, W. M., and Bowker, J. C. (2015). "Children in peer groups," in *Handbook of Child Psychology and Developmental Science: Ecological Settings and Processes*, Vol. 4, 7th Edn, eds M. H. Bornstein, T. Leventhal, and R. M. Lerner (Hoboken, NJ: John Wiley & Sons, Inc.), 175–222. doi: 10.1002/9781118963418.childpsy405

Rubin, K. H., Bukowski, W. M., and Parker, J. G. (2006). "Peer interactions, relationships, and groups," in *Handbook of Child Psychology (Volume 3: Social, Emotional, and Personality Development)*, ed. N. Eisenberg (Hoboken, NJ: John Wiley & Sons, Inc).

Ruch, W., Platt, T., and Hofmann, J. (2014). The character strengths of class clowns. *Front. Psychol.* 5:1075. doi: 10.3389/fpsyg.2014.01075

Suitor, J. J., Powers, R. S., and Brown, R. (2004). Avenues to prestige among adolescents in public and religiously affiliated high schools. *Adolescence* 39, 229–241.

Wagner, L. (2018). Good character is what we look for in a friend: character strengths are positively related to peer acceptance and friendship quality in early adolescents. *J. Early Adolesc.* doi: 10.1177/0272431618791286 [Epub ahead of print].

Wagner, L., Gander, F., Proyer, R. T., and Ruch, W. (2019). Character strengths and PERMA: investigating the relationships of character strengths with a multidimensional framework of well-being. *Appl. Res. Qual. Life* doi: 10.1007/s11482-018-9695-z [Epub ahead of print].

Waldrip, A. M., Malcolm, K. T., and Jensen-Campbell, L. A. (2008). With a little help from your friends: the importance of high-quality friendships on early adolescent adjustment. *Soc. Dev.* 17, 832–852. doi: 10.1111/j.1467-9507.2008.00476.x

Wanzer, M. B., Booth-Butterfield, M., and Booth-Butterfield, S. (1996). Are funny people popular? An examination of humor orientation, loneliness, and social attraction. *Commun. Q.* 44, 42–52. doi: 10.1080/01463379609369999

Zeller, M., Vannatta, K., Schafer, J., and Noll, R. B. (2003). Behavioral reputation: a cross-age perspective. *Dev. Psychol.* 39, 129–139. doi: 10.1037/0012-1649.39.1.129

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wagner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 April 2017
doi: 10.3389/fpsyg.2017.00606

Studying Irony Detection Beyond Ironic Criticism: Let's Include Ironic Praise

Richard Bruntsch[1,2]* *and Willibald Ruch*[1,2]

[1] Department of Psychology, Section Personality and Assessment, University of Zurich, Zurich, Switzerland, [2] Department of Psychology, Distance Learning University Switzerland, Brig, Switzerland

Studies of irony detection have commonly used ironic criticisms (i.e., mock positive evaluation of negative circumstances) as stimulus materials. Another basic type of verbal irony, ironic praise (i.e., mock negative evaluation of positive circumstances) is largely absent from studies on individuals' aptitude to detect verbal irony. However, it can be argued that ironic praise needs to be considered in order to investigate the detection of irony in the variety of its facets. To explore whether the detection ironic praise has a benefit beyond ironic criticism, three studies were conducted. In Study 1, an instrument (Test of Verbal Irony Detection Aptitude; TOVIDA) was constructed and its factorial structure was tested using $N = 311$ subjects. The TOVIDA contains 26 scenario-based items and contains two scales for the detection of ironic criticism vs. ironic praise. To validate the measurement method, the two scales of the TOVIDA were experimentally evaluated with $N = 154$ subjects in Study 2. In Study 3, $N = 183$ subjects were tested to explore personality and ability correlates of the two TOVIDA scales. Results indicate that the co-variance between the ironic TOVIDA items was organized by two inter-correlated but distinct factors: one representing ironic praise detection aptitude and one representing ironic criticism detection aptitude. Experimental validation showed that the TOVIDA items truly contain irony and that item scores reflect irony detection. Trait bad mood and benevolent humor (as a facet of the sense of humor) were found as joint correlates for both ironic criticism and ironic praise detection scores. In contrast, intelligence, trait cheerfulness, and corrective humor were found as unique correlates of ironic praise detection scores, even when statistically controlling for the aptitude to detect ironic criticism. Our results indicate that the aptitude to detect ironic praise can be seen as distinct from the aptitude to detect ironic criticism. Generating unique variance in irony detection, ironic praise can be postulated as worthwhile to include in future studies—especially when studying the role of mental ability, personality, and humor in irony detection.

Keywords: cheerfulness, confirmatory factor analysis, corrective humor, intelligence, ironic praise, irony, personality, STCI

OPEN ACCESS

Edited by:
Marcel Zentner,
University of Innsbruck, Austria

Reviewed by:
Roger J. Kreuz,
University of Memphis, USA
Ursula Beermann,
University of Innsbruck, Austria

***Correspondence:**
Richard Bruntsch
r.bruntsch@psychologie.uzh.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *05 December 2016*
Accepted: *03 April 2017*
Published: *21 April 2017*

Citation:
Bruntsch R and Ruch W (2017) Studying Irony Detection Beyond Ironic Criticism: Let's Include Ironic Praise. Front. Psychol. 8:606. doi: 10.3389/fpsyg.2017.00606

INTRODUCTION

Ironic criticism and ironic praise can be distinguished as two basic types of verbal irony (cf. Kreuz and Link, 2002). The two types are structurally similar to each other as both involve mock evaluations of circumstances with a valence opposite to the speaker's true appraisal. As the characteristic difference between the two, ironic praise is characterized by a negative valence in what

is said and a positive valence in the speaker's true appraisal of circumstances while in ironic criticism the converse is true[1].

When we use irony, we typically utter something different from what we want to express, i.e., typically the opposite of our true appraisal of circumstances. Characteristically, we expect the listener to recognize our overt dissimulation by seeing through the counterfactual nature of our utterance and to eventually detect the intended meaning of what we say nonetheless (Groeben and Scheele, 2003). However, this is not always the case, as listeners may not detect the irony for certain reasons. For example, imperfect irony detection rates were found as a function of the ambiguity of the context of ironic utterances. Accordingly, Ackerman (1983) reports considerable average error rates in his irony detection task (ranging from 5.6 to 24.1% depending on the difficulty of the stimuli) in a control group consisting of college students. Furthermore, individuals differ in their aptitude to detect verbal irony, which results in systematic variance in irony detection performance (e.g., Winner et al., 1998; see Bruntsch et al., 2016, for an overview).

In the studies investigating irony detection, a plethora of tasks and *ad-hoc* test has been used to assess individuals' aptitude to detect verbal irony. However, most of these studies did not utilize both *ironic criticism* (as a mock positive evaluation of negative circumstances) and *ironic praise* (as a mock negative evaluation of positive circumstances). Rather, the stimuli used in the existing studies on irony detection mostly rely on ironic criticisms (such as in the form of sarcasm[2]), whereas ironic praise is not represented (e.g., Ackerman, 1983; Happé, 1993; McDonald and Pearce, 1996; Mitchley et al., 1998). This is somewhat puzzling, as ironic praise can be found as counterbalanced with ironic criticism in the stimuli used in studies targeting different aspects of irony processing, such as when investigating processing times (i.e., response latencies) of ironic stimuli vs. their literal counterparts (Schwoebel et al., 2000). Likewise, there are studies investigating perceived speaker's intent in "ironic insults" (matching the definition of *ironic criticism* we adhere to; cf. Kreuz and Link, 2002) and "ironic compliments" (matching the definition of *ironic praise*) vs. direct insults and direct compliments, respectively (e.g., in terms of ratings of mocking and politeness, i.e., Pexman and Olineck, 2002).

However, studies investigating irony *detection* have largely neglected the sampling of ironic praise stimuli. This may be owed to the view that ironic praise can be seen as the less prevalent and less "prototypically ironic" type of irony (cf. Kreuz and Link, 2002). However, a study by Langdon et al. (2002) demonstrated that stimuli containing ironic praise led to different results than ironic criticism stimuli. Langdon et al. (2002) used both, ironic criticism (labeled as *sarcasm*) and ironic praise (labeled as *banter*), and distinguished them in separate scores for their investigation of irony detection in schizophrenic patients vs. normally functioning control subjects[3]. As Langdon et al. (2002) report, ironic praise was harder to detect than ironic criticism, especially in the group of patients with schizophrenia. Thus, it can be hypothesized that ironic praise may be the very type of irony that is affected by impaired or unusual cognitive and affective functioning. More generally, it may be suggested that ironic praise leads to meaningful interindividual variance in irony detection tasks beyond the one found for ironic criticism.

IRONIC CRITICISM VS. IRONIC PRAISE

As detailed below, we argue that the two types of irony can be distinguished considering at least three aspects: (a) they have different purposes and functions in communication, (b) in irony detection ironic praise may depend on individuals' expression of certain traits more than ironic criticism, and (c) in irony detection they demand different cognitive and affective processes in individuals.

(A) One may characterize that ironic praise is typically used for different purposes (for example good-natured "ironic teasing;" Keltner et al., 2001) than ironic criticism (for example aggressive ridicule). In the form of teasing, ironic praise may be reasoned to be a way to humorously apprise the recipient of social norms when harmless transgressions occur—such as when using it as a playful provocation in socializing, flirting, or entertaining. In contrast, ironic criticism may be employed for the purpose of apprising the recipient of social norms when more severe transgressions occur—such as when resolving conflicts by aggressive ridicule (cf. Norrick, 1994; Keltner et al., 2001). Furthermore, as ironic praise is typically used in the face of positive circumstances, one may reason that ironic praise is more

[1]To illustrate: imagine that a circle of friends is watching a sports match and some of the attendees support Team A while other attendees support Team B. An example of ironic criticism would be if one of the supporters of Team A said "Terrific shot! You're handing us a resounding defeat!" when a player of Team B tries but fails to score a goal in the match (for example when the speaker wants to ridicule the arrogant prediction made by one of the supporters of Team B that "their" Team B would win at a canter). In contrast, if one of the supporters of Team A said "Terrible shot! We don't stand the slightest chance!" when a player of Team A scores a goal, this would be an example of ironic praise (for example when the speaker wants to ridicule one of the supporters of Team B for his or her arrogant prediction that Team A would lose the match in a sad spectacle of defeat).

[2]The terms "irony" and "sarcasm" are sometimes used interchangeably (e.g., Pexman and Olineck, 2002) and there is an ongoing debate as to whether sarcasm and irony are essentially the same thing (cf. Attardo, 2000). However, we wish to adhere to a demarcation between irony and sarcasm in terms of two naturally overlapping but conceptually distinct phenomena. For example, in the *Merriam-Webster Dictionary* sarcasm is defined as "a sharp and often satirical or ironic utterance designed to cut or give pain." Irony can be seen as related to sarcasm because phrasing a criticism ironically was found to enhance the degree of perceived condemnation (as compared to phrasing it literally, i.e., Colston, 1997). In the present paper we will stick to the term *irony* (even if sarcasm is involved in a specific instance of irony), foremost because if any of the studies in our literature review use the term *sarcasm*, they originally refer to *ironic* sarcasm (rather than *non-ironic* sarcasm).

[3]Langdon et al. (2002) use the term *banter* when labeling the category of their stimuli containing a negative statement being used in a positive context (which corresponds to the definition of ironic praise we adhere to) without the intention to harm or to criticize, whereas their *sarcasm* stimuli were characterized by a positive statement used in a negative context (which corresponds to the definition of ironic criticism) with the intention to harm or to criticize. It is necessary to mention here that elsewhere bantering irony was conceptualized to occur not only in situations in which the speaker intends to ironically praise (i.e., in terms of *kind banter*) but also when *ironically criticizing* (i.e., in terms of *sarcastic banter*; cf. Anolli et al., 2002). However, as far as we can tell from their report, Langdon et al. (2002) used banter only in the form of *kind banter* (i.e., they did not include *sarcastic banter* involving ironic criticism) in their banter stimuli.

suitable than ironic criticism (which in turn is typically used in the face of adverse circumstances) for certain of the discourse goals found for verbal irony, such as to be funny or witty, to be humorous, and to play or to be silly (cf. Kreuz et al., 1991).

(B) The different functional aspects of the two types of irony (such as different utilities in social interaction) may affect the detection of ironic criticism and ironic praise differently, depending on individuals' expression of certain traits, such as the sense of humor. As the notion that humor is a function of irony is pervasive in the literature (cf. Bruntsch et al., 2016), the sense of humor (which can be defined as relatively stable interindividual differences in the tendency to react to humor and to produce humor, and a serene attitude toward life; see Ruch, 1998) can be assumed to go along with the readiness to detect or mis-detect verbal irony. Certain *facets* of the sense of humor may come into play more evidently in the detection of ironic praise than in the detection of ironic criticism. Furthermore, looking at ironic praise as a playful and light-hearted figure of speech, its detection may be facilitated by cheerfulness (e.g., Ruch et al., 1996) more than this is the case for ironic criticism. This may be the case because highly cheerful individuals may process cues signaling playfulness more readily, which helps to reject the uttered negative evaluation and detect the more positive implication of ironic praise. Importantly, this may not hold true for ironic criticism, which may be seen as less playful and less jocular than ironic praise.

(C) It can be argued that the norm violation that irony typically alludes to and criticizes (e.g., Utsumi, 2000; Garmendia, 2014) is harder to recognize in the case of ironic praise: it may be more obvious and hence easier to understand why ironic criticism is used. This may be because people generally have positive expectations (e.g., successful players in professional sports; cf. Kreuz and Link, 2002). Thus, the detection of ironic praise may require a more complex mental representation of the background of the ironic remark and a more effortful cognitive search for the antecedent event that ironic remarks typically refer to (Kreuz and Glucksberg, 1989), as compared to the detection of ironic criticism. In line with this consideration, intelligence may be more relevant for the detection of ironic praise than for the detection of ironic criticism. If the role of intelligence truly was more evident in the detection of ironic praise, ironic praise should be included in irony research when mental abilities as well as mental impairments are targeted.

AIMS OF THE PAPER

The current paper has three main aims. Firstly, a test for the assessment of irony detection with two different scales (i.e., ironic criticism vs. ironic praise) will be developed, opting for an indirect measurement format (Study 1). It is aimed to use two testing modes with different degrees of *irony alertness*: hiding the measurement intention from participants (i.e., *irony non-alert* mode) vs. making irony salient (*irony alert* mode). Using confirmatory factor analysis, the two-factor structure (corresponding to the distinction between ironic criticism and ironic praise) will be tested. Secondly, in Study 2 we will validate the soundness of the stimuli and the indirect measurement by (a) using an experimental approach (i.e., comparing four testing conditions: irony alert testing, irony non-alert testing, forced ironic interpretation, and forced literal interpretation), (b) testing whether there is a convergence between the test scores and direct irony-ratings, and (c) comparing direct irony-ratings between ironic items and non-ironic distractor items (which should differ from each other). Thirdly, Study 3 will explore ability and personality correlates of the two scales. It is expected that ironic praise detection scores are at least as strongly related—if not even more strongly related—to (a) intelligence, (b) the ability to distinguish irony from a lie, (c) different facets of the sense of humor, and (d) traits constituting the temperamental foundation of the sense of humor (e.g., cheerfulness), as this is the case for the detection of ironic criticism.

STUDY 1: DEVELOPMENT OF THE TEST OF VERBAL IRONY DETECTION APTITUDE (TOVIDA)

It is assumed that there is meaningful interindividual variance in irony detection performance in terms of an irony detection aptitude. It is hypothesized that this aptitude comprises two facets: the aptitude to detect ironic criticism and the aptitude to detect ironic praise. After selecting those items with the most acceptable psychometric features, a confirmatory factor analysis will be employed to investigate whether the two predefined concepts used in the instrument (ironic criticism and ironic praise) are represented by two different structural components. A first sample will be used to determine psychometric properties under *irony non-alert* testing conditions, as this unobtrusive method can be reasoned to reflect individuals' everyday mode of dealing with irony (i.e., usually, we do not deliberately watch out for irony). Then, a second sample will be used for cross-validation to see whether the fit of a two-factor model (i.e., ironic criticism vs. ironic praise) can be confirmed under *irony alert* testing conditions. Maximizing irony alertness can be reasoned to reduce systematic noise in the interindividual variance. To specify: as some individuals may be more biased not to anticipate irony in a psychological survey than others, *irony non-alert* testing presumably would lead to artificial co-variance between the items. Furthermore, as the shared variance between items systematically depends on the interindividual variance that makes co-variance arise in the first place, this method can be seen as a source of data accommodating a more conservative test of the assumed model.

Methods

Participants

Participants were recruited via university mailing lists, social platforms, and leaflets. Two independent samples were used. Sample 1 consisted of 152 German-speaking subjects (40 males [35.7%]). Age in Sample 1 ranged from 18 to 51 years with a mean of 22.8 ($SD = 5.8$). Sample 2 consisted of 159 German-speaking subjects (39 males [32.5%]). Age in Sample 2 ranged from 18 to 67 years with a mean of 24.1 ($SD = 7.3$).

Materials

Test of Verbal Irony Detection Aptitude-40 (TOVIDA-40)

To develop a test for the assessment of irony detection aptitude, 30 scenarios containing ironic target utterances (among which 20 contained ironic criticism and 10 contained ironic praise) and 10 scenarios with non-ironic target utterances were written using a rational construction procedure. Irony detection was defined as the comprehension of the true meaning of ironic target utterances as opposite to the literal meaning in ambiguous situations short of distinct information. Each scenario consists of a short story about two or more people and culminates in a final utterance (the target utterance) made by one of the protagonists. Target utterances contain either verbal irony or literal speech. When generating the stimuli, irony was designed as follows: in the ironic utterances used in the *ironic criticism* stimuli, speakers (i.e., the story characters making the target utterance) use a choice of words which, when used non-ironically, denotes a positive appraisal—while ironically implying an opposite (i.e., negative) appraisal. Conversely, as the characteristic feature of the utterances found in the *ironic praise* stimuli, speakers use a choice of words which, when used non-ironically, denote a negative appraisal of circumstances—while ironically implying an opposite (i.e., positive) appraisal. In the ironic criticism stimuli, speakers comment on a negative circumstance described in the short story (with a mock positive evaluation). In contrast, in the ironic praise stimuli, speakers comment on a positive circumstance (with a mock negative evaluation). In the TOVIDA-40, ironic utterances typically involve meta-messages indirectly implied by the speaker, such as when mocking the addressee's overly self-critical or self-effacing attitude[4].

The scenarios are designed as ambiguous in order to warrant sufficient psychometric item difficulty, i.e., to avoid ceiling effects. This is why the stories still make some sense when irony is not detected in the ironic items (i.e., in the case of false negative detection) and when irony is falsely detected in the non-ironic items (i.e., in the case of false positive detection). Accounting for ambiguity in the process of irony detection, Utsumi (2000) points out that irony is distinguished from non-irony by assessing the degree to which a given utterance resembles prototypical irony. That is, not every ironic utterance unambiguously fulfills the constituting criteria of irony. Rather, the listener detects irony by assessing the similarity between a given utterance and a prototype of irony. Hence, ambiguity can be seen as a typical feature of real-life situations involving irony. However, in the scenarios of the TOVIDA-40 there are unobtrusive cues signaling the preconditions for the ironic utterance, i.e., hints to a reason for the speaker to express a negative attitude via ironic criticism or ironic praise (cf. Utsumi, 2000; Garmendia, 2014)[5]. In order to assess whether participants chose a literal or an ironic interpretation of target utterances, participants have to judge scenarios along statements about factual aspects of the situation or actors' emotional states as causes or consequences of target utterances. A person detecting the irony correctly appraises the situation differently from a person not detecting the irony. The TOVIDA-40 was designed as an unobtrusive test that can be optionally administered without any mention of irony and distracts test-takers from its true measurement intention. Six statements are provided for the appraisal of the situation (to be rated on a four-point scale ranging from 1 = "does not apply at all" to 4 = "fully applies"), among which three are indicative of irony detection (see Appendix). The other three appraisal statements are designed to distract from the intention of the task. For example, there is a statement asking whether the protagonists behave like a typical male or female (according to his or her gender) provided for every scenario. A high item score in the ironic items indicates correct positive irony detection, i.e., the comprehension of the true meaning of ironic target utterances as opposite to the literal meaning. The ironic items are administered alternating with the non-ironic distractor items.

Procedure

Participants were tested individually using an online-survey. They were randomly assigned to one of two groups (labeled here as Sample 1 and Sample 2). They either were instructed without any mention of irony (Sample 1: *irony non-alert* testing) or provided with a definition of verbal irony and instructed to watch out for irony in the stimuli, i.e., they were instructed that some of the scenarios they were about to appraise contain verbal irony whereas others do not (Sample 2: *irony alert* testing). Participants completed the TOVIDA-40 after they filled in questions about their demographic features and German language proficiency.

Preliminary Analyses

In order to arrive at more reliable items scores, two of the three indicative statements were selected for every item applying a scale reliability criterion: inter-correlations between the three indicators were computed using Sample 1. The two indicators with the highest inter-correlation were selected and averaged to generate the item scores. In order to attain a more economic form of the TOVIDA-40, corrected item-total correlations (CITCs) were computed and considered as a selection criterion. Ironic criticism and ironic praise items were analyzed separately in this step. For selection purposes, only Sample 1 was used. For each of the two sub-scales eight items showed CITCs of $r_{cit} \geq 0.45$ and were selected to build two scales to be analyzed in the

[4]As speakers say something different from what they actually want to express, irony classifies as an indirect speech act, cf. (Holtgraves, 1997). What makes irony different from other forms of indirect speech acts is that ironic speech acts are characterized by an overt insincerity. That is, ironic speakers achieve indirectness by engaging in an evident dissimulation when inversing the valence of their true appraisal in the verbatim utterance (i.e., especially by using a choice of words denoting *the opposite* of their true appraisal of circumstances, cf. Attardo, 2000, 2001).

[5]As Garmendia (2014) argues, irony is always negative in terms of a critical attitude. That is, also in the case of ironic praise, which—as a meta-message—can be described to typically involve a hint to the transgression of (sometimes unwritten) rules, for example the norm of not to be vain, not to boast, not to be arrogant, not to be overly modest, not to make false promises, and so on. That does not mean that there is not another meta-message on a higher level that can be characterized as benevolent and more positive. For example, ironic teasing can be corrective and bonding at the same time, as the teaser implies that he or she thinks that the relationship with the teased person is strong and close enough to make playful provocation possible without risking a serious social damage (cf. Norrick, 1994; Boxer and Cortés-Conde, 1997).

further steps of Study 1. The 16 selected ironic items and the 10 non-ironic distractor items taken from the TOVIDA-40 will be referred to as the *TOVIDA* in the following sections.

Results

Internal consistencies of the two resulting sub-scales were sufficiently high. Cronbach's alpha was 0.83 (0.76) for the ironic criticism scale and 0.83 (0.77) for ironic praise scale in Sample 1 and Sample 2 (values for Sample 2 in brackets).

Within the irony alert sample (Sample 2), the fit of two different structural equation models was estimated. In the *assumed model*, two inter-correlating factors were modeled: one factor was defined by ironic criticism items and the other factor by ironic praise items. In the *control model*, a single factor was modeled defined by both ironic criticism and ironic praise items. As it turned out, the assumed two-component model had acceptable fit ($c^2 = 153.296$, $df = 103$; Bentler Comparative Fit Index [CFI] = 0.906; root mean square error of approximation [RMSEA] = 0.056 [90% CI: 0.036; 0.073]; standardized root mean square residual [SRMR] = 0.0643). In contrast, the control model did not show acceptable model fit ($\chi^2 = 227.025$, $df = 104$; CFI = 0.771; RMSEA = 0.078 [90% CI: 0.071; 0.102]; SRMR = 0.0840). The path coefficients for the assumed two-factor model are given in **Figure 1**. As **Figure 1** shows, the ironic criticism scale and the ironic praise factors were substantially intercorrelated.

Discussion

The selection from the two types of items resulted in two scales with sufficient internal consistency. This indicates that there is an underlying irony detection aptitude creating shared variance in the items. Furthermore, the two-factorial structure could be affirmed, implying that ironic praise generated unique variance in the TOVIDA. Hence, the findings of Study 1 support the assumption that the aptitude to detect ironic praise is worth distinguishing from the aptitude to detect ironic criticism.

STUDY 2: EXPERIMENTAL EVALUATION OF THE TOVIDA

The stimuli employed in the TOVIDA were designed as ambiguous in order to warrant sufficient psychometric item difficulty, i.e., to avoid ceiling effects. Furthermore, irony detection is assessed indirectly in order to make a testing mode feasible in which subjects are non-alert to the occurrence of irony in the stimuli. So it was deemed necessary to validate that the stimuli of the TOVIDA truly contain irony, and if so, that high (vs. low) test scores truly indicate high (vs. low) irony detection performance. The aim of Study 2 was to address these questions.

Four criteria were defined to evaluate whether the TOVIDA allows for the assessment of irony detection: firstly, participants in the *irony alert* group are expected to have higher scores than participants in a *forced literal appraisal* group (i.e., participants instructed to view all items as non-ironic). This criterion reflects the consideration that there must be a group consensus among participants who know about the intention of the test that differs from a forced appraisal opposite to the designed ironic content. Secondly, participants in the *irony alert* group are expected to have higher scores than the ones in the *irony non-alert* group. The rationale of this criterion is that irony detection is facilitated when participants are instructed to watch out for irony (vs. being not informed about the possible occurrence of irony). Thirdly, a *forced ironic appraisal* group (i.e., participants instructed to view all items as ironic) is expected to have higher scores than the *forced literal appraisal* group. This criterion aims at ensuring that the appraisals used for the indirect measurement (and hence the item scores) are sensitive to irony detection. As a fourth criterion, the item scores within the *irony alert* group are expected to be positively correlated with direct appraisals (i.e., explicit ratings) of ironic content (these were assessed only in this group).

Methods

Participants

Participants were recruited in university lectures, via university mailing lists, social platforms, and leaflets. The sample consisted of 154 German-speaking subjects (26 male [16.9%]). Participants' age ranged from 18 to 56 years with a mean of 24.8 years ($SD = 7.8$). They were randomly assigned to one of four testing conditions and the groups did not differ significantly as to age [$F_{(3, 150)} = 1.69$, $p = 0.17$], nor gender [$F_{(3, 150)} = 0.085$, $p = 0.97$].

Instruments

The *Test of Verbal Irony Detection Aptitude* (*TOVIDA; see* Study 1 for description and Appendix for an example item). Item scores were computed following the method of Study 1.

Procedure

In an online-survey, participants were randomly assigned to one of four test conditions: (1) one group was given a definition of verbal irony, was briefed that some of the scenarios they were about to see contain verbal irony whereas others do not, and instructed to take all target utterances as *ironic* when appraising the scenarios along the predefined statements (*forced ironic appraisal*), (2) one group was given a definition of verbal irony, was briefed that some of the scenarios they were about to see contain verbal irony whereas others do not, and instructed to take all target utterances as *literal* while appraising the scenarios (*forced literal appraisal*), (3) another group was given a definition of verbal irony, was briefed that some of the scenarios they were about to see contain verbal irony whereas others do not, and instructed to watch out for irony when appraising the scenarios according to their own interpretation (*irony alert*), and (4) the last group was instructed to appraise the scenarios according to their own interpretation without any mention of irony (*irony non-alert*). More specifically, the experimental instructions in the *forced ironic appraisal* group and the *forced literal appraisal* group briefed participants (a) to willfully view the last sentence in each of the situations as ironic or non-ironic, respectively, and (b) to respond to all of the concerned questions *as if* the last sentence was truly ironic or non-ironic, respectively. In the irony alert group, participants were requested to make direct appraisals (i.e., explicit ratings) of ironic content in addition to the standard appraisal. These explicit ratings of ironic content were assessed via a four-point Likert-type scale (1 = "not ironic," 2 = "rather

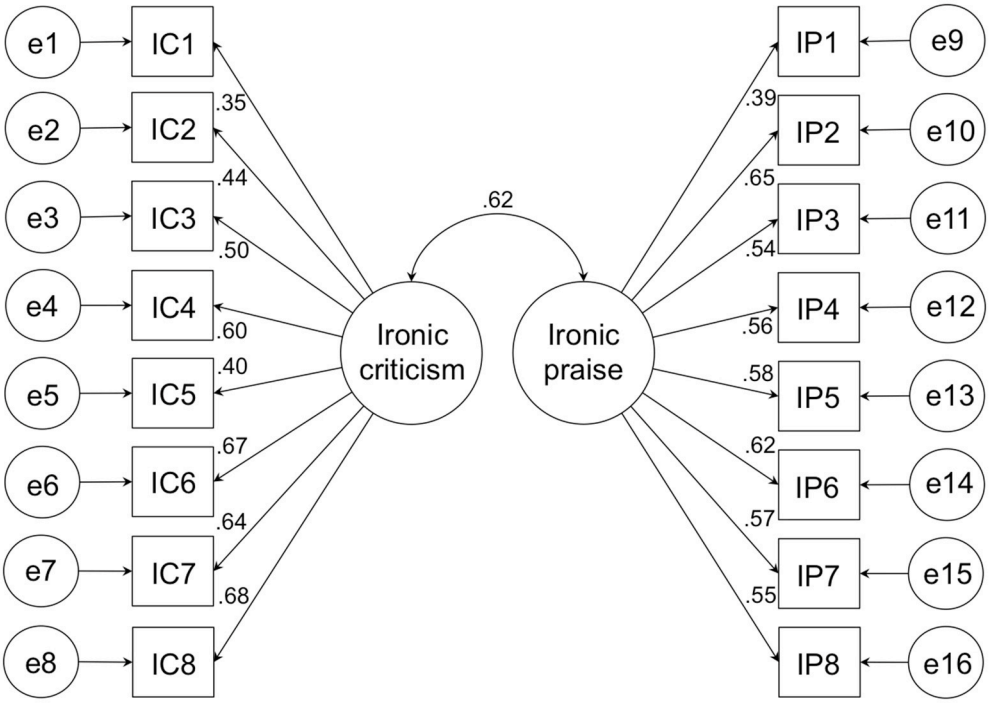


FIGURE 1 | Estimates for path coefficients in the two-factor model that was confirmed in Study 1.

not ironic," 3 = "rather ironic," 4 = "ironic"), accounting for the ambiguous nature of the scenarios. Participants in the alert group were considered lay judges for this purpose (for the use of laypersons for validation purposes see Legree, 1995). The *irony alert* group was randomly over-sampled in order to warrant sufficient sample size for the planned correlational analyses.

Results

Do the Stimuli of the TOVIDA Contain Irony?

Group means of item scores are given in **Table 1**. As **Table 1** shows, all items met the criterion to verify that they contain irony. More precisely, in line with the expectations, the *forced literal appraisal* group had lower means than the *irony alert* group with medium to large effect sizes, indicating that generally irony is detected in ironic items. Furthermore, in the *irony alert* group item scores were generally higher than in the *irony non-alert* group with small to large effect sizes (however, only 10 out of 16 of the comparisons yielded significant differences). In line with the expectation, being alert to irony facilitated irony detection.

Next, the direct appraisals of ironic content were examined to find out whether ironic items are viewed as more ironic than the non-ironic items. The frequencies of the single ratings were considered, given in **Table 2**. As **Table 2** shows, ironic criticism items and the ironic praise items had numerically higher appraisals of being ironic ("rather ironic" and "ironic" answers) than non-ironic control items. It is noteworthy that the distributions of the proportions of ironic appraisals had a contact point: the ironic item with the lowest frequency of ironic appraisals (IC1) was judged about just as ironic as the non-ironic control item with the highest frequency of ironic appraisals (NC08). However, these two items can be seen as outliers in their group and as there was still a fair amount of judges consenting that the ironic items in question contain irony. Thus, they can be considered as difficult items but still containing irony. To test whether ironic criticism and ironic praise items were rated as more ironic than the non-ironic control items in the direct appraisals of ironic content, a mean of ratings over the eight items per scale was computed as well as the mean of ratings for the 10 non-ironic control items. These scores were compared with paired sample *t*-tests. It turned out that the non-ironic control items were rated as less ironic ($M = 1.59$, $SD = 0.38$) than the ironic criticism items [$M = 2.95$, $SD = 0.58$, $t_{(63)} = -14.47$, $p < 0.001$] and the ironic praise items [$M = 3.23$, $SD = 0.54$, $t_{(63)} = -18.16$, $p < 0.001$], indicating large effect sizes (i.e., $d = 2.77$ and $d = 3.51$, respectively)[6].

Is Irony Detection Reflected in the Item Scores of the TOVIDA?

As **Table 1** shows, the item score means of the *forced ironic appraisal* group were higher than item score means of the *forced*

[6]An exploratory analysis indicated that ironic praise items were appraised as somewhat more ironic than the ironic criticism items with a medium effect size ($d = 0.50$). This is important to point out, as the direct appraisals are substantially correlated with the item scores (i.e., the indirect appraisals). The irony in ironic praise items hence can be seen as less difficult to detect than ironic criticism. It is not clear whether this is owed to the fact that the ironic praise items used in the present set of studies are less ambiguous than the ironic criticism items or whether ironic praise *per-se* is easier to detect than ironic criticism.

TABLE 1 | Descriptive and test statistics of group scores in Study 2.

Item	Test instruction (group)								Group comparisons					
	1 Non-ironic		2 Ironic		3 Alert		4 Non-alert		1 vs. 2		1 vs. 3		3 vs. 4	
	M	*SD*	M	*SD*	M	*SD*	M	*SD*	$t_{(63)}$	*d*	$t_{(90)}$	*d*	$t_{(87)}$	*d*
IC1	2.17	0.57	3.18	0.54	2.54	0.67	2.47	0.72	−7.25*	1.83	−2.59*	0.58	0.45	0.10
IC2	1.60	0.70	3.25	0.62	3.04	0.70	2.76	0.89	−10.09*	2.52	−9.06*	2.06	1.55	0.37
IC3	1.36	0.42	3.23	0.86	2.82	0.88	2.00	0.89	−10.63*	2.65	−8.34*	1.89	3.92*	0.93
IC4	1.37	0.55	3.50	0.72	2.64	1.05	2.13	1.06	−13.11*	3.26	−6.06*	1.37	2.05*	0.48
IC5	1.90	0.64	2.88	0.59	2.46	0.57	2.35	0.75	−6.35*	1.60	−4.19*	0.95	0.80	0.18
IC6	1.69	0.67	3.36	0.64	3.02	0.68	2.67	0.94	−10.23*	2.56	−8.62*	1.96	1.95	0.46
IC7	1.56	0.74	3.49	0.72	3.07	0.80	2.16	1.04	−10.52*	2.65	−8.54*	1.93	4.42*	1.04
IC8	2.08	0.81	3.25	0.67	2.89	0.56	2.43	0.82	−6.37*	1.60	−5.48*	1.26	3.03*	0.72
IP1	2.21	0.55	3.71	0.49	3.52	0.58	2.87	0.69	−11.59*	2.90	−10.13*	2.29	4.51*	1.06
IP2	1.57	0.84	3.31	0.77	3.18	0.61	2.67	0.98	−8.66*	2.17	−10.32*	2.34	2.96*	0.70
IP3	1.77	0.57	3.34	0.78	2.77	0.69	2.48	0.69	−8.93*	2.25	−6.66*	1.52	1.78	0.42
IP4	1.38	0.48	3.34	0.93	2.95	0.88	2.38	1.01	−10.18*	2.55	−8.88*	2.01	2.61*	0.62
IP5	2.01	0.52	3.05	0.69	2.94	0.51	2.57	0.54	−6.65*	1.67	−7.99*	1.81	3.02*	0.71
IP6	1.90	0.67	3.43	0.52	3.22	0.55	3.01	0.78	−10.39*	2.60	−9.86*	2.24	1.40	0.34
IP7	1.44	0.44	3.33	0.84	2.83	0.86	2.15	0.82	−10.87*	2.71	−8.09*	1.83	3.43*	0.80
IP8	1.77	0.67	3.00	0.91	2.60	0.70	2.21	0.57	−6.03*	1.51	−5.28*	1.20	2.47*	0.59

*IC1–IC8, ironic criticism items; IP1–IP8, ironic praise items. Non-ironic, forced literal appraisal (n = 28); Ironic, forced ironic appraisal (n = 37); Alert, irony alert testing (n = 64); Non-alert, irony non-alert testing (n = 25); d, Cohen's d coefficient of effect size. *p < 0.05.*

literal appraisal group, with large effect sizes. This indicates that a person will score high in all items if he or she detects the irony and score low if this is not the case. Finally, as expected, the direct appraisals (i.e., explicit ratings) of ironic content in the *irony alert* group correlated significantly with the respective item scores in all items with a mean of $r_{(63)} = 0.72$, indicating good convergence between direct and indirect appraisals. This finding indicates that the TOVIDA test scores reflect the degree to which participants considered the stimuli as ironic.

Discussion

The results support the claim that, the ironic criticism and ironic praise stimuli used by the TOVIDA contain irony. Firstly, item scores were higher the group instructed to watch out for irony (i.e., the *irony alert* group) than in the group with experimentally induced minimal irony detection (i.e., in the *forced literal appraisal* group). This finding indicates that irony can generally be detected in the items of the TOVIDA (with a fair amount of interindividual variance, as shown by substantial standard deviations in *irony alert* and *irony-non alert* individuals' detection scores). Secondly, alertness to the ironic content of the stimuli fostered irony detection as the *irony-alert* group had higher item scores than the *irony non-alert* group in the majority of the items. Thirdly, the direct appraisals of the ironic content indicate that the ironic items were viewed as more ironic than the non-ironic items. There is also support for the claim that test scores reflect iron detection. Firstly, this was evident in terms of considerable differences between a group with experimentally induced minimal irony detection (i.e., in the *forced literal appraisal* group) and a group with experimentally induced maximal irony detection (i.e., in the *forced ironic appraisal* group). Secondly, the item scores corresponded well with direct appraisals (i.e., explicit ratings) of ironic content. These findings indicate that the items of the TOVIDA assess irony detection performance and that the stimuli—although they were designed as ambiguous—were consented as containing verbal irony to an acceptable degree.

STUDY 3: EXPLORING THE USEFULNESS OF IRONIC PRAISE IN A STUDY OF IRONY DETECTION CORRELATES

Study 3 aimed at exploring whether ironic praise stimuli have a benefit in the investigation of ability and personality correlates of irony detection. Among the preexisting studies assuming an individual differences perspective in irony research, Ivanko et al. (2004) explored the possibility to explain interindividual variance in an irony interpretation task (i.e., in terms of participants' ratings of speaker's intent, such as *sarcasm*, *mocking*, and *politeness*) by means of participants' scores in "conversational indirectness" (i.e., the tendency to phrase one's remarks indirectly and the extent to which a person looks for indirect meanings in the remarks of others, cf. Holtgraves, 1997). The present study aims to extend this and other previous work (e.g., Blouin and McKelvie, 2012) by (a) looking at irony *detection* (rather than irony *comprehension* as the interpretation of speaker's attributes in ironic utterances) and (b) including intelligence and a broad range of personality traits as individual differences variables.

As one of the hypothesized correlates, it may be argued that trait cheerfulness has a relevance especially to the detection of

TABLE 2 | Direct irony appraisal using explicit irony ratings for the single items of the TOVIDA (Study 2).

Items	Rating scale steps							
	"Not ironic"		"Rather not ironic"		"Rather ironic"		"Ironic"	
	f	(%)	*f*	(%)	*f*	(%)	*f*	(%)
IC1	19	(29.7)	19	(29.7)	12	(18.8)	14	(21.9)
IC2	3	(4.7)	6	(9.4)	22	(34.4)	33	(51.6)
IC3	6	(9.4)	17	(26.6)	15	(23.4)	26	(40.6)
IC4	15	(23.4)	10	(15.6)	17	(26.6)	22	(34.4)
IC5	10	(15.6)	16	(25.0)	20	(31.3)	18	(28.1)
IC6	7	(10.9)	3	(4.7)	25	(39.1)	29	(45.3)
IC7	6	(9.4)	8	(12.5)	17	(26.6)	33	(51.6)
IC8	5	(7.8)	13	(20.3)	14	(21.9)	32	(50.0)
IP1	5	(7.8)	6	(9.4)	8	(12.5)	45	(70.3)
IP2	3	(4.7)	2	(3.1)	4	(6.3)	55	(85.9)
IP3	10	(15.6)	20	(31.3)	16	(25.0)	18	(28.1)
IP4	7	(10.9)	5	(7.8)	23	(35.9)	29	(45.3)
IP5	2	(3.1)	3	(4.7)	25	(39.1)	34	(53.1)
IP6	2	(3.1)	3	(4.7)	13	(20.3)	46	(71.9)
IP7	8	(12.5)	12	(18.8)	14	(21.9)	30	(46.9)
IP8	11	(17.2)	12	(18.8)	23	(35.9)	18	(28.1)
NC01	33	(51.6)	21	(32.8)	7	(10.9)	3	(4.7)
NC02	49	(76.6)	14	(21.9)	1	(1.6)	0	(0)
NC03	34	(53.1)	18	(28.1)	4	(6.3)	8	(12.5)
NC04	37	(57.8)	20	(31.3)	2	(3.1)	5	(7.8)
NC05	47	(73.4)	9	(14.1)	6	(9.4)	2	(3.1)
NC06	34	(53.1)	23	(35.9)	6	(9.4)	1	(1.6)
NC07	56	(87.5)	6	(9.4)	1	(1.6)	1	(1.6)
NC08	18	(28.1)	20	(31.3)	13	(20.3)	13	(20.3)
NC09	46	(71.9)	15	(23.4)	2	(3.1)	1	(1.6)
NC10	34	(53.1)	19	(29.7)	8	(12.5)	3	(4.7)
DESCRIPTIVE STATISTICS								
M_{IC}	8.88	13.9	11.50	18.0	17.75	27.7	25.88	40.4
SD_{IC}	5.49	8.6	5.68	8.9	4.33	6.8	7.24	11.3
M_{IP}	6.00	9.4	7.88	12.3	15.75	24.6	34.38	53.7
SD_{IP}	3.55	5.5	6.27	9.8	7.55	11.8	13.41	20.9
M_{NC}	38.80	60.6	16.50	25.8	5.00	7.8	3.70	5.8
SD_{NC}	10.84	16.9	5.48	8.6	3.80	5.9	4.03	6.3

N = 64. IC1–IC8, ironic criticism items; IP1-IP8, ironic praise items; NC01–NC10, non-ironic control items; M_{IC}/ SD_{IC}, mean/standard deviation for ironic criticism item ratings; M_{IP}/SD_{IP}, mean/standard deviation for ironic praise item ratings; M_{NC}/SD_{NC}, mean/standard deviation for non-ironic control item ratings.

ironic praise as cheerful individuals may have a more positive outlook on themselves and others and hence be more inclined to expect jolly and jovial interactions involving playful ironic teasing rather than hostile and negative interaction involving serious ridicule, such as in the form of ironic criticism. Furthermore, certain facets of the sense of humor may be more relevant to the detection of ironic praise than to the detection of ironic criticism. According to Ruch and Heintz (2016), the sense of humor includes also two virtue-related facets, i.e., *benevolent humor* and *corrective humor*. As an accepting way of dealing with negative circumstances (e.g., human weaknesses), benevolent humor may be relevant especially to ironic criticism (typically occurring in the face of negative circumstances) but not as relevant to ironic praise (typically occurring in the face of positive circumstances). That is, individuals prone to use, enjoy, seek, and understand benevolent humor may have a higher aptitude to detect ironic criticism. The other facet is characterized by tendencies to wittily ridicule those who deserve it from a moral stance in terms of *corrective humor*. Importantly, irony is listed as one of the ways in which corrective humor manifests itself in speech. It can be argued that by exposing transgressions of social rules in a witty and playful way, corrective humor is conceptually more related to ironic praise than to ironic criticism, which in turn can be seen as the more serious and less ingenious form of irony. Hence, individuals who are prone to use, enjoy, seek, and understand corrective humor, may have a higher readiness to detect irony

in the case of ironic praise more than in the case of ironic criticism.

Furthermore, irony detection can be related to mental abilities—and presumably especially so in the case of ironic praise. According to previous studies (e.g., Mitchley et al., 1998) intelligence can be seen as a prerequisite for the detection of ironic criticism. Under the presupposition that the detection of ironic praise poses a different cognitive challenge to the individual than the detection of ironic criticism, there may be a unique relationship between the detection of ironic praise and mental abilities. Hence, a test for the assessment of general mental ability (i.e., intelligence) will be employed. To include a measure of an ability more specific to irony detection, a task by Winner et al. (1998) will be jointly administered that was designed to assess the ability to discriminate between irony and lies among patients with brain damage. Simultaneously, by testing its convergence with the detection of ironic criticism and ironic praise, the convergent validity of the TOVIDA will be explored.

Accordingly, we expect that there are associations between ironic praise detection and individual differences variables that are robust beyond the influence of the variance the detection of ironic praise shares with the detection of ironic criticism. Moreover, it is expected that both of the two scales of the TOVIDA correlate positively with the irony/lie discrimination task, as the ability to distinguish irony from a lie can be seen as relevant to ironic praise to the same extent as to ironic criticism.

As a secondary aim, the association between the two scales of the TOVIDA and the Big Five personality traits will be explored to learn more about the discriminant value of the irony detection measure. It is expected that the Big Five as broad personality dimensions distal to the sense of humor and distinct from mental ability are largely unrelated to irony detection scores. For exploratory purposes, again two testing modes will be employed with different degrees of *irony alertness*: hiding the measurement intention from participants (i.e., *irony non-alert* mode) vs. making irony salient (*irony alert* mode). As there are no comparable previous studies on personality and ability correlates of irony detection, it was preferred to include both the irony non-alert and the irony alert mode of testing in order to safeguard the investigation against a selective method bias.

Methods

Participants

Participants were recruited in university lectures, and by means of university mailing lists, social platforms, and leaflets. Two independent quasi-experimental groups were tested. The first group (irony non-alert testing mode) consisted of 103 German-speaking subjects (28 male [22.0%]). Age in Group 1 ranged from 18 to 38 years with a mean of 21.6 ($SD = 3.5$). Group 2 (irony alert testing mode) consisted of 80 German-speaking subjects (16 males [17.6%]). Age in this group ranged from 18 to 46 years with a mean of 22.7 ($SD = 5.5$).

Instruments

Test of Verbal Irony Detection Aptitude (TOVIDA)

The *Test of Verbal Irony Detection Aptitude* (*TOVIDA*) was used for the assessment of irony detection performance (see Study 1 for description/Appendix). Item scores were computed following the method of Study 1. The scores of the eight ironic criticism items and the eight ironic praise items were averaged to build an ironic criticism detection score and an ironic praise detection score, respectively. The internal consistencies of the two scales were comparable to those found in Study 1. Cronbach's alpha was 0.81 (0.74) for the ironic criticism scale and 0.83 (0.79) for the ironic praise scale in the *irony non-alert* group and the *irony alert* group, respectively (values for the *irony alert* group in brackets).

Achievement Measurement System 2 (LPS-2 [Leistungsprüfsystem 2]; Kreuzpointner et al., 2013)

The LPS-2 is a performance test for the assessment of general mental ability. It employs 11 subtests that are allocated to four of the eight dimensions proposed by Carroll's (1993) model of intelligence, namely "crystallized intelligence" (e.g., solving anagrams), "fluid intelligence" (e.g., reasoning), "visual perception" (i.e., the ability to generate and process mental representations of spatial objects, to visualize, and to detect spatial patterns, e.g., mental rotation), and "cognitive speed" (e.g., arithmetic). A general IQ score is derived by aggregating the four subscales. Internal consistencies for subtests and the four dimensions are satisfactory in the norm sample with Cronbach's alpha ranging from 0.72 to 0.95. The total internal consistency for form A (form B) is high in the norm sample, $\alpha = 0.96$ ($\alpha = 0.97$). Split-half reliability of subtests ranges from sufficient ($r_{tt} = 0.81$) to high ($r_{tt} = 0.93$). Validity is confirmed in terms of concurrence with a range of other tests of mental ability. Furthermore, the targeted dimensional structure of the test is confirmed. The LPS-2 can be administered in groups and takes around 60 min to complete.

Irony/lie discrimination task (Winner et al., 1998)

This task measures the capacity to attribute second-order mental state and the ability to distinguish between ironic statements and lies. Subjects are required to read 15 short stories and to identify whether the final assertion is a lie or an ironic joke. There are eight stories involving a lie and seven stories implicating irony (in terms of intentionally and overtly uttering a counterfactual statement to a person known to be aware of the true circumstances). According to the characterization given by Winner et al. (1998), each story describes a context in which one person witnesses another individual breaking a rule sneakily (e.g., stealing food). The main difference between the two story types is that in the lie stories, the protagonist does not know that he or she had been seen doing the "sneaky action" and utters a lie to the witness to avoid getting caught. In the ironic stories, the protagonist knows he or she has been seen during the transgression and thereupon utters an ironic comment (i.e., a joke) to conceal his or her shame of being caught. For each story type (i.e., "joke" stories and lie stories), a separate score is generated by summing up participants' individual false negative decisions (i.e., the discrimination errors).

State-Trait Cheerfulness Inventory (STCI; Ruch et al., 1996)

The STCI is a questionnaire measure for the components of exhilaratability as the temperamental basis of the sense of humor. The trait version (STCI-T) encompasses three scales

assessing *cheerfulness* (e.g., "I have a 'sunny' nature."), *seriousness* (e.g., "I prefer people who communicate with deliberation and objectivity."), and *bad mood* (e.g., "Even if there is no reason, I often feel ill-humored."). In current study a 60-item short form of the STCI-T was used. The questionnaire assesses the endorsements of statements on a four-point scale (ranging from 1 = "strongly disagree" to 4 = "strongly agree"). Internal consistencies in the present sample were comparable to the ones in the construction sample reported by Ruch et al. (1996) with Cronbach's alpha ranging from 0.80 (seriousness) to 0.95 (bad mood).

Statements of Benevolent and Corrective Humor (BenCor; Ruch and Heintz, 2016)

The BenCor is a list of statements assessing two virtue-related facets of the sense of humor. *Six* statements are used for benevolent humor (e.g., "Even when facing unpleasant events I can keep my distance and discover something amusing or funny in it") and corrective humor (e.g., "I caricature my fellow humans' wrongdoings in a funny way to gently urge them to change"), each. They were answered on a 7-point Likert scale ranging from 1 ("strongly disagree") to 7 ("strongly agree"). Internal consistencies in the present sample were sufficient: Cronbach's alpha was 0.75 for benevolent humor and 0.78 for corrective humor.

Inventory of Minimal Redundant Scales

Inventory of Minimal Redundant Scales (*MRS-25* [Inventar Minimal Redundanter Skalen], Ostendorf, 1990; 25-item short form developed by Schallberger and Venetz, 1999). The MRS-25 is a list of 25 bipolar adjectives pairs for the assessment of the Big Five personality dimensions *extraversion* (e.g., impulsive vs. restrained), *agreeableness* (e.g., affirmative vs. oppositional), *conscientiousness* (e.g., diligent vs. lazy), *emotional stability* (e.g., robust vs. vulnerable), and *culture* (e.g., inventive vs. conventional). Answers are given on a six-point scale (*very—quite—rather—rather—quite—very*). Schallberger and Venetz (1999) report high internal consistencies of the scales and evidence for the validity of the MRS-25. Internal consistencies in the present sample were satisfactory with Cronbach's alpha ranging from 0.72 (agreeableness) to 0.86 (conscientiousness and emotional stability).

Procedure

Participants were tested in two consecutive sessions. In Session 1, groups up to 30 persons completed the LPS-2 as the first part of a larger assessment battery also including measures that were unrelated to the present study in the laboratory, quasi-randomly assigned to form A or Form B, depending on their seating position (as to avoid influence by neighboring participants). Due to time constraints, all other measures were included in an online survey. Participants were assigned an individual code and provided with an invitation containing an URL directing them to the online survey (Session 2). Within 7 days after Session 1, participants logged in and indicated their personal code for matching purposes. In Session 2, participants first completed the TOVIDA quasi-randomly assigned to one of two conditions: Half of the groups tested in Session 1 were given a definition of verbal irony and were instructed to watch out for irony, i.e., they were told that some of the scenarios they were about to appraise contain verbal irony whereas others do not (*irony alert* condition). The other half took the test naïve to its true intention (*irony non-alert* condition), i.e., there was no mention of the possible occurrence of verbal irony. Subsequently, STCI-T, the Big Five measure (MRS-25), the sense of humor measure (i.e., the BenCor), and the irony/lie discrimination task by Winner et al. (1998) were completed.

Results

Is the Detection of Ironic Criticism and Ironic Praise Associated with Abilities and Traits?

The correlations between the two subscales of the TOVIDA and the other measures are given in **Table 3**, for the irony non-alert and the irony-alert group separately. As **Table 3** shows, the ironic criticism scale was correlated substantially with the ironic praise scale but not correlated significantly with the other measures in the *irony non-alert* group. However, there was a trend for an association between the ironic criticism scale and

TABLE 3 | Correlations between irony detection scores and the personality and ability measures (Study 3).

Personality and ability measures	TOVIDA test instruction					
	Irony non-alert			Irony alert		
	IC	IP	IP_p	IC	IP	IP_p
TOVIDA IP	0.52*	–	–	0.46*	–	–
INTELLIGENCE (LPS-2)						
Crystallized intelligence	0.03	0.04	0.03	0.00	0.10	0.11
Fluid intelligence	0.05	0.25*	0.26*	−0.04	0.17	0.21
Visual perception	0.15	0.16	0.09	0.07	0.28*	0.28*
Cognitive speed	−0.07	0.09	0.14	0.00	0.14	0.16
General IQ	0.06	0.22*	0.22*	0.03	0.23*	0.24*
IRONY/LIE DISCRIMINATION						
Irony (joke stories)	−0.17	−0.27*	−0.23*	−0.24*	−0.30*	−0.22
Non-irony (lie stories)	0.09	0.01	−0.04	0.10	0.02	−0.03
BIG FIVE						
Agreeableness	0.11	0.04	−0.02	0.17	0.02	−0.06
Conscientiousness	0.05	0.02	−0.01	0.05	0.12	0.12
Emotional stability	0.04	0.03	−0.01	0.17	0.24*	0.18
Extraversion	0.06	0.05	−0.03	0.01	0.17	0.20
Culture	0.14	0.19	0.14	0.04	0.01	0.03
TEMPERAMENTAL TRAITS						
Cheerfulness	−0.03	−0.08	−0.07	0.20	0.29*	0.23*
Seriousness	−0.08	−0.04	−0.01	−0.07	−0.04	−0.01
Bad mood	−0.03	0.01	0.04	−0.34*	−0.35*	−0.23*
SENSE OF HUMOR						
Benevolent humor	0.11	0.07	0.03	0.24*	0.30*	0.22
Corrective humor	−0.05	−0.06	−0.06	0.09	0.26*	0.25*

*n = 97–103 irony non-alert individuals. n = 80 irony alert individuals. IC, ironic criticism scale of the TOVIDA; IP, ironic praise scale of the TOVIDA; Sense of Humor, scales of the BenCor; IP_p, partial correlations with IP controlling for the influence of IC. *p < 0.05 (two-tailed).*

the visual perception dimension of the LPS-2 (i.e., spatial ability), the performance in the ironic items (i.e., the joke stories) of the irony/lie discrimination task by Winner et al. (1998), and culture. In the *irony alert* group, again the ironic criticism scale was correlated substantially with the ironic praise scale. Furthermore, as expected, there was an association between the ironic criticism scale and the performance in the ironic items of the irony/lie discrimination task by Winner et al. (1998). Furthermore, there was also a trend for an association between the ironic criticism scale and emotional stability. In line with the expectations, among the self-report measures, bad mood and benevolent humor showed a significant relation to the ironic criticism scale and there was a trend for an association with cheerfulness. Furthermore, there was also a trend for ironic criticism detection showing an association with agreeableness and emotional stability.

As expected, the ironic praise scale was significantly correlated with intelligence in terms of fluid intelligence and with the performance in the ironic items of the irony/lie discrimination task in the *irony non-alert* group. Furthermore, there was a trend for an association with visual perception and culture for the ironic praise scale. In the *irony alert* group, the ironic praise scale was associated with intelligence in terms of the LPS-2 dimension visual perception (and there was also a trend for an association with the fluid intelligence dimension). Furthermore, the ironic praise scale again was negatively correlated with the number of errors made in the ironic items of the irony/lie discrimination task by Winner et al. (1998). Among the scales of the self-report measures, emotional stability, cheerfulness, bad mood, benevolent humor, and corrective humor showed significant correlations with the ironic praise scale. Furthermore, there was also a trend for an association with extraversion for the ironic praise scale in this group.

Are There Unique Correlates for Ironic Praise Beyond Ironic Criticism?

Next, it was tested whether in the study of irony detection correlates ironic praise generates meaningful variance that contributes a surplus value over the meaningful variance found for ironic criticism. Therefore, partial correlations were computed between the ironic praise detection scale and the external variables while controlling for individuals' ironic criticism detection scores. The partial correlations are given in **Table 3**. As can be seen in **Table 3**, in the *irony non-alert* group, ironic praise correlated positively with *fluid intelligence* and negatively with the error rate in the irony items of the irony/lie discrimination task even beyond the influence of the variance shared with ironic criticism detection. In the *irony alert* group ironic praise correlated positively with the *visual perception* dimension of the intelligence test, trait cheerfulness, trait bad mood (in a negative direction), and corrective humor over and above the variance that the ironic criticism scale shared with ironic praise and these variables.

Discussion

The findings of Study 3 indicate that assessing the detection of ironic praise can provide a surplus value over the detection of ironic criticism. Ironic praise detection can be seen as more challenging than the detection of ironic criticism in terms of numerically higher associations as well as significant partial correlations with the intelligence measure when the influence of the aptitude to detect ironic criticism was controlled for[7]. Hence, ironic praise detection appears to be dependent on mental ability to a certain degree, which is in line with previously reported findings on the role of intelligence in irony detection (e.g., Mitchley et al., 1998). However, considering the numerical size of the correlations, ironic praise detection aptitude can be seen as distinct from intelligence. Furthermore, as expected, it was found that the detection of ironic praise was uniquely associated with corrective humor, while ironic criticism was related only to benevolent humor. Also, cheerfulness played a unique role in the detection of ironic praise. Possibly increasing the readiness to process humorous meta-messages or playful cues in ironic teasing, a cheerful temperament hence can be assumed to facilitate the detection of irony, foremost in the form of ironic praise.

The Big Five personality traits were largely unrelated to irony detection scores except for a correlation between the ironic praise scale and *emotional stability*. It can be assumed that emotionally stable individuals have a higher readiness to reject the uttered criticism in what is literally said and recognize the more benevolent nature of what is ironically implied in the ironic praise items, compared to individuals low in emotional stability (who in turn may not "get over" the criticism or insult uttered in ironic praise). Although there was also a trend for an association between the irony detection scores on the one hand and culture and agreeableness on the other, the Big Five can be seen as less relevant for irony detection than narrower and more humor-related traits. Moreover, participants' scores in the TOVIDA converged with their scores in the ironic items of the irony/lie discrimination task, indicating convergent validity of the TOVIDA.

Do Ability and Personality Variables Interact in Irony Detection?

As an exploratory analysis complementing our correlational analyses, we wish to address the possibility that ability and personality variables interact in irony detection. To illustrate, although intelligence was found as positively related to irony detection, there might be highly intelligent individuals who still perform poorly in irony detection because they lack the requisite personality traits facilitating irony detection. Guided by the findings displayed in **Table 3**, we explored the data from Study 3 to see whether interactions between intelligence and personality could be found to predict irony detection beyond the main effects of the separate variables. Indeed, this assumption was found to hold true in one of the cases that we studied: in the irony-alert

[7]Differential associations between the two scales of the TOVIDA and the intelligence variables could be explained by differences in average item difficulty. As ironic praise items were more frequently appraised as ironic than ironic criticism items in Study 2, it is possible that the lack of association between the ironic criticism scale and intelligence hence might be an artifact created by higher ambiguity of the materials. This may be the case because intelligence may foster irony detection only when items have a low ambiguity.

sample the interaction between the spatial ability dimension of the LPS-2 (i.e., *visual perception*) and *benevolent humor* predicted ironic praise detection significantly by explaining incremental variance beyond the main effects of the single predictors.

A hierarchical regression analysis with two steps was computed with the ironic praise detection score as the criterion. In Step 1, *visual perception* ($\beta = 0.25$) and *benevolent humor* ($\beta = 0.26$) were significant predictors, $F_{(2, 77)} = 6.70$, $p = 0.002$. As it turned out, the interaction term (computed as the simple multiplication of *visual perception* and *benevolent humor* scores) explained a significant increment of criterion variance when added to the equation in Step 2, $F_{(3, 76)} = 7.27$, $p < 0.001$; $\Delta R^2 = 0.075$, $p = 0.008$. As a possible interpretation of this finding, intelligence could be seen as a necessary but not sufficient condition for irony detection, as irony detection may be facilitated by individuals' cognitive ability only if individuals have enough sense of humor to successfully deal with irony. The inverse may also be true: the sense of humor may only manifest itself in irony detection performance if individuals have the necessary ability to successfully deal with its cognitive demands.

GENERAL DISCUSSION

Our findings support the assumption that the detection of ironic criticism and the detection of ironic praise can be found as two intercorrelated but still discriminant facets of irony detection aptitude. Furthermore, our findings substantiate the assumption that ironic praise is useful beyond ironic criticism: applied in an investigation of ability and personality correlates, the detection of ironic praise was found to be uniquely associated with certain variables (i.e., intelligence, trait bad mood, trait cheerfulness, and the corrective facet of the measure of the sense of humor), beyond the influence of ironic criticism detection aptitude.

Extrapolating our findings, we may propose assumptions as to *why* more intelligent individuals high in cheerfulness and low in bad mood with high scores in benevolent and corrective humor may have a higher readiness to detect the irony in ironic praise. Maybe they are more able or ready to (a) reason and infer the meta-message of an ironic praise (i.e., *fluid intelligence*), (b) generate an easily interpreted mental "image" of the background of an ironic remark (i.e., *visual perception* as the ability to generate mental representations, to visualize, and to detect patterns), (c) take into account playful and humorous communicative intentions in terms of the processing of exhilarant stimuli (i.e., high trait cheerfulness and low trait bad mood), (d) have a smiling attitude toward the imperfections of life (e.g., human weakness) and know how to deal with them by using *benevolent humor* (i.e., in terms of the principle "it takes one to know one"), and (e) expose transgressions of morally valued social rules by using irony with satirical meta-messages in order to educate and better social others (i.e., the tendency to produce, to enjoy, and to make sense of *corrective humor*).

The Role of Irony Alertness

There was an irregularity in the findings of Study 3 (which, however, occurred in a quite constant fashion): in the *irony non-alert* group, the association between the personality variables and irony detection was not evident compared to the *irony alert* group. As a possible explanation for this finding, participants in the *irony non-alert* sample may have been biased toward expecting a bona fide communication mode, as in the given psychological assessment situation a serious state of mind may have been induced. This consideration may have an implication for the assessment of irony detection in general terms, as in many of the pre-existing measurement procedures for the assessment of irony detection irony alertness is reduced by not mentioning to participants that the stimuli they are about to encounter contain irony and by using indirect measurement (i.e., not asking the participants directly whether they think that there is irony in a stimulus)[8]. At least as far as the study of personality and ability correlates of irony detection is concerned, it can be seen as worthwhile to further explore the benefit of maximizing irony alertness and using direct testing.

Is the TOVIDA Too Difficult?

In the construction of the TOVIDA we assumed that, in order to tap into the variance in irony performance among normally functioning adults, psychometrically difficult items need to be employed (as to avoid ceiling effects). Notably, there is a trade-off between item difficulty (i.e., ambiguousness of the stimuli) and test-takers' consensus as to the ironic nature of the stimuli. Certainly, the items should not be too difficult to allow for a sufficient consensus among test takers as to whether irony is present in the stimuli or not. However, a fair amount of variance (i.e., an imperfect consensus) can be argued to be admissible as this variance (a) must be expected when conceptualizing irony detection aptitude as an approximately normally distributed variable, and (b) is rooted in the nature of the construct when dealing with phenomena involving an inherent uncertainty, which—apart from irony—can also be found for example in certain knowledge domains. Accordingly, Legree (1995) for example argues in favor of a Likert-based assessment of social intelligence because of the level of uncertainty involved in the stimuli. He characterizes the challenge of assessing knowledge of ambiguous relationships when he states that "situational judgment scales attempt to simulate everyday problem situations but cannot allow the formulation of unambiguously "correct" solutions. This ambiguity partially reflects real-world interpersonal interactions, which are often ambiguous [...]" (Legree, 1995, p. 249).

The Possible Role of Self-involvement

In the TOVIDA, test-takers have to make sense of situations containing verbal irony from an observer's perspective (i.e., with low *self-involvement*). It would also be thinkable to test irony detection performance using self-involving situations, such as when instructing test takers to place themselves into the respective situation as if they would encounter them in real life.

[8]In previous studies indirect measurement of irony detection was operationalized for example by resorting to fact questions (e.g., Ackerman, 1983; Happé, 1993), using questions targeting mental states of the speaker and emotions of the target of the ironic utterance (e.g., McDonald and Pearce, 1996), asking whether it made any sense for the speaker to make the target utterance (e.g., Langdon and Coltheart, 2004), or rewording the use of irony as "joking" (e.g., Winner et al., 1998).

Importantly, this may lead to certain variables coming into play more prominently as correlates of irony detection performance. For example, self-involvement may accentuate the association between ironic praise detection and *emotional stability*. If a specific instance of ironic praise is an interpersonal evaluation, emotionally unstable individuals may be more attached to the negative interpersonal valence of the verbatim utterance (which can occur in the form of a mock critical offense) and hence may be less prone to reject the literal interpretation of the ironic remark—and importantly so this mechanism may be accentuated as self-involvement in the assessment of irony detection increases. This consideration may also apply to certain other traits, such as self-esteem or the fear of being laughed at (i.e., gelotophobia; cf. Ruch et al., 2014). For example, because of their general belief to be inherently ridiculous and deficient, gelotophobes may be sensitive to derisive ironic criticism especially when self-involvement is high. Accordingly, future studies investigating traits relevant to derisive criticism or offense in irony detection should explore the benefit of self-involving test stimuli and instructions.

CONCLUSIONS

Ironic criticism and ironic praise can be seen as separate scales in irony detection. The two types of irony were differently related to ability and personality variables, as ironic praise detection showed unique associations with intelligence and certain traits. Hence,—at least as far as the stimuli used in our investigation are concerned—ironic praise can be postulated to generate variance with surplus meaning beyond the variance generated by ironic criticism in irony detection. Consequently, ironic praise as the less "prototypical" and formerly neglected type of irony and can be postulated as especially important to include when studying the role of ability, personality, and humor in irony detection.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Psychological Research Ethics Committee of the University of Zurich with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Psychological Research Ethics Committee of the University of Zurich.

AUTHOR CONTRIBUTIONS

RB: Data collection, data analysis, drafting manuscript. WR: Data analysis, drafting manuscript.

ACKNOWLEDGMENTS

The authors thank Jasmine Fong for her help during data collection for Study 2 and Jenny Hofmann for her comments on an earlier version of this manuscript.

REFERENCES

Ackerman, B. P. (1983). Form and function in children's understanding of ironic utterances. *J. Exp. Child Psychol.* 35, 487–508. doi: 10.1016/0022-0965(83)90023-1

Anolli, L., Ciceri, R., and Infantino, M. G. (2002). From "blame by praise" to "praise by blame": Analysis of vocal patterns in ironic communication. *Int. J. Psychol.* 37, 266–276. doi: 10.1080/00207590244000106

Attardo, S. (2000). Irony as relevant inappropriateness. *J. Pragmat.* 32, 793–826. doi: 10.1016/S0378-2166(99)00070-3

Attardo, S. (2001). "Humor and irony in interaction: From mode adoption to failure of detection," in *Say Not to Say: New Perspectives on Miscommunication,* eds L. Anolli, R. Ciceri, and G. Riva (Amsterdam: IOS Press), 159–179.

Blouin, P. S., and McKelvie, S. J. (2012). Postformal thinking as a predictor of creativity and of the identification and appreciation of irony and metaphor. *N. Am. J. Psychol.* 14, 39–50. Available online at: http://najp.us/north-american-journal-of-psychology-index

Boxer, D., and Cortés-Conde, F. (1997). From bonding to biting: conversational joking and identity display. *J. Pragmat.* 27, 275–294. doi: 10.1016/S0378-2166(96)00031-8

Bruntsch, R., Hofmann, J., and Ruch, W. (2016). Virgin soil in irony research: personality, humor, and the "sense of irony." *Transl. Issues Psychol. Sci.* 2, 25–34. doi: 10.1037/tps0000054

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge: Cambridge University Press.

Colston, H. L. (1997). Salting a wound or sugaring a pill: the pragmatic functions of ironic criticism. *Discourse Process.* 23, 25–45. doi: 10.1080/01638539709544980

Garmendia, J. (2014). The clash: humor and critical attitude in verbal irony. *Humor* 27, 641–659. doi: 10.1515/humor-2014-0094

Groeben, N., and Scheele, B. (2003). "Produktion von Ironie und Witz [Production of irony and wit]," in *Enzyklopädie der Psychologie: Sprachproduktion [Encyclopedia of Psychology: Speech Production], 1st Edn.,* Vol. 3, Hrsg.[Eds.] eds T. Herrmann and J. Grabowski (Göttingen: Hogrefe), 733–763.

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition* 48, 101–119. doi: 10.1016/0010-0277(93)90026-R

Holtgraves, T. (1997). Styles of language use: individual and cultural variability in conversational indirectness. *J. Pers. Soc. Psychol.* 73, 624–637. doi: 10.1037/0022-3514.73.3.624

Ivanko, S. L., Pexman, P. M., and Olineck, K. M. (2004). How sarcastic are you? Individual differences and verbal irony. *J. Lang. Soc. Psychol.* 23, 244–271. doi: 10.1177/0261927X04266809

Keltner, D., Capps, L., Kring, A. M., Young, R. C., and Heerey, E. A. (2001). Just teasing: a conceptual analysis and empirical review. *Psychol. Bull.* 127, 229–248. doi: 10.1037/0033-2909.127.2.229

Kreuz, R. J., and Glucksberg, S. (1989). How to be sarcastic: the echoic reminder theory of verbal irony. *J. Exp. Psychol. Gen.* 118, 374–386. doi: 10.1037/0096-3445.118.4.374

Kreuz, R. J., and Link, K. E. (2002). Asymmetries in the use of verbal irony. *J. Lang. Soc. Psychol.* 21, 127–143. doi: 10.1177/02627X02021002002

Kreuz, R. J., Long, D. L., and Church, M. B. (1991). On being ironic: pragmatic and mnemonic implications. *Metaphor Symbolic Activ.* 6, 149–162. doi: 10.1207/s15327868ms0603_1

Kreuzpointner, L., Lukesch, H., and Horn, W. (2013). *LPS-2. Leistungsprüfsystem 2 [Achievement Measurement System 2]*. Göttingen: Hogrefe.

Langdon, R., and Coltheart, M. (2004). Recognition of metaphor and irony in young adults: the impact of schizotypal personality traits. *Psychiatry Res.* 125, 9–20. doi: 10.1016/j.psychres.2003.10.005

Langdon, R., Davies, M., and Coltheart, M. (2002). Understanding minds and understanding communicated meanings in schizophrenia. *Mind Lang.* 17, 68–104. doi: 10.1111/1468-0017.00190

Legree, P. J. (1995). Evidence for an oblique social intelligence factor established with a Likert-based testing procedure. *Intelligence* 21, 247–266. doi: 10.1016/0160-2896(95)90016-0

McDonald, S., and Pearce, S. (1996). Clinical insights into pragmatic theory: frontal lobe deficits and sarcasm. *Brain Lang.* 53, 81–104. doi: 10.1006/brln.1996.0038

Mitchley, N. J., Barber, J., Gray, J. M., Brooks, D. N., and Livingston, M. G. (1998). Comprehension of irony in Schizophrenia. *Cogn. Neuropsychiatry* 3, 127–138. doi: 10.1080/135468098396206

Norrick, N. R. (1994). Involvement and joking in conversation. *J. Pragmat.* 22, 409–430. doi: 10.1016/0378-2166(94)90117-1

Ostendorf, F. (1990). *Sprache und Persönlichkeitsstruktur. Zur Validität des Fünf-Faktoren-Modells der Persönlichkeit [Language and personality. On the Validity of the Five-Factor Model of Personality]*. Regensburg: Roderer.

Pexman, P. M., and Olineck, K. M. (2002). Does sarcasm always sting? Investigating the impact of ironic insults and ironic compliments. *Discourse Process.* 33, 199–217. doi: 10.1207/S15326950DP3303_1

Ruch, W. (1998). "Foreword and overview. The sense of humor: a new look at an old concept," in *The Sense of Humor: Explorations of a Personality Characteristic,* ed W. Ruch (Berlin: Mouton de Gruyter), 3–14.

Ruch, W., and Heintz, S. (2016). The virtue gap in humor: exploring benevolent and corrective humor. *Transl. Issues Psychol. Sci.* 2, 35–45. doi: 10.1037/tps0000063

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. (2014). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., Köhler, G., and Van Thriel, C. (1996). Assessing the "humorous temperament": construction of the facet and standard trait forms of the State-Trait-Cheerfulness-Inventory–STCI. *Humor* 9, 303–339. doi: 10.1515/humr.1996.9.3-4.303

Schallberger, U., and Venetz, M. (1999). *Kurzversionen des MRS-Inventars von Ostendorf (1990) zur Erfassung der fünf "Grossen" Persönlichkeitsfaktoren. [Short version of the MRS Inventory by Ostendorf (1990) for the Assessement of the "big" Five Personality Factors].* Berichte aus der Abteilung Angewandte Psychologie, Nr. 30. Psychologisches Institut der Universität Zürich: Zürich. [Reports from the section of applied psychology, No. 30. Department for Psychology, University of Zurich: Zurich].

Schwoebel, J., Dews, S., Winner, E., and Srinivas, K. (2000). Obligatory processing of the literal meaning of ironic utterances: further evidence. *Metaphor Symbol* 15, 47–61. doi: 10.1080/10926488.2000.9678864

Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: distinguishing ironic utterances from nonirony. *J. Pragmat.* 32, 1777–1806. doi: 10.1016/S0378-2166(99)00116-2

Winner, E., Brownell, H., Happé, F., Blum, A., and Pincus, D. (1998). Distinguishing lies from jokes: theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain Lang.* 62, 89–106. doi: 10.1006/brln.1997.1889

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer UB and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Bruntsch and Ruch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

The following sample item of the TOVIDA given below is translated from German language. The statements for the appraisal of the situation were rated on a four-point scale as to how much the sentences apply to the situation (1 = "does not apply at all," 2 = "rather does not apply," 3 = "rather applies," 4 = "fully applies"). Statements printed in bold were used as (inversed) indicators of irony detection in the studies and averaged to build the item score. Three of the appraisal statements were designed as distractors for each item.

Ironic Praise Sample Item (IP6)

Situation:

Christian has invited three friends over for dinner. He prepares a meal trying a new recipe. Sitting at the table starting to eat, Julia asks for the saltshaker. Christian immediately apologizes that he could not salt the food to taste as he has a cold and cannot taste properly. This is when Julia says: "You are right, the food is inedible!"

Instruction: *Please indicate how much you think that each of the following statements applies to the situation.*

Statements:

1. *The saltshaker is not within Julia's reach.*
2. *Christian behaves like a typical male.*
3. *Julia is having a good time.*
4. ***Christian will feel bad because of Julia's final utterance. (−)***
5. ***Julia will not finish her plate without more salt. (−)***
6. *Christian is making up excuses.*

frontiers in Psychology

ORIGINAL RESEARCH
published: 19 December 2017
doi: 10.3389/fpsyg.2017.02215

When Sugar-Coated Words Taste Dry: The Relationship between Gender, Anxiety, and Response to Irony

Anna Milanowicz, Adam Tarnowski and Barbara Bokus*

Faculty of Psychology, University of Warsaw, Warsaw, Poland

This article approaches the question of mocking compliments and ironic praise from an interactional gender perspective. A statement such as "You're a real genius!" could easily be interpreted as a literal compliment, as playful humor or as an offensive insult. We investigate this thin line in the use of irony among adult men and women. The research introduces an interactional approach to irony, through the lens of gender stereotype bias. The main question concerns the impact of individual differences and gender effect on the perception and production of ironic comments. Irony Processing Task (IPT), developed by Milanowicz (2016), was applied in order to study the production and perception of ironic criticism and ironic praise in adult males and females. It is a rare case of a study measuring the ability to create irony because, unlike most of known irony research, it is not a multiple choice test where participants are given the response options. The IPT was also used to assess the asymmetry of affect (humor vs. malice) and impact of gender effect in the perception of ironic comments. Results are analyzed in relation to the State-Trait Anxiety Inventory (STAI) scores. The findings reveal the interactional relationship between gender and response to irony. Male responses were consistently more ironic than female's, across all experimental conditions, and female responses varied more. Both, men and women used more irony in response to male ironic criticism but female ironic praise. Anxiety proved to be a moderate predictor of irony comprehension and willingness to use irony. Data, collected in control and two gender stereotype activation conditions, also corroborates the assumption that the detection of compliments and the detection of criticism can be moderated by the attitude activation effect. The results are interpreted within the framework of linguistic intergroup bias (LIB) and natural selection strategies.

Keywords: irony, gender bias, anxiety, blame by praise, praise by blame, humor, malice

OPEN ACCESS

Edited by:
Tracey Platt,
University of Wolverhampton,
United Kingdom

Reviewed by:
Andres Mendiburo-Seguel,
Universidad Andrés Bello, Chile
Gil Greengross,
Aberystwyth University,
United Kingdom

****Correspondence:***
Anna Milanowicz
ania.milanowicz@gmail.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *31 August 2017*
Accepted: *06 December 2017*
Published: *19 December 2017*

Citation:
Milanowicz A, Tarnowski A and Bokus B (2017) When Sugar-Coated Words Taste Dry: The Relationship between Gender, Anxiety, and Response to Irony. Front. Psychol. 8:2215. doi: 10.3389/fpsyg.2017.02215

INTRODUCTION

> "L'humour est une disposition d'esprit qui fait qu'on exprime avec gravité des choses frivoles et avec légèreté des choses sérieuses."
>
> Alfred Capus

Irony is wordplay, a figure of speech that flouts the maxim of quality, requiring information provided in conversation to be truthful (Grice, 1975). However, it implies the contradiction of what is literally expressed. It is characterized by opposition and substitution between two levels of meaning. It is an Aristotelian blame-by-praise figure, criticism which sounds like a compliment, where in fact, what the speaker literally says should be taken to mean "something else," conveniently

assumed to be the exact or relative opposite of what is said. Irony can be humorous and humor can be ironic—these two concepts may overlap but are not tantamount.

Every day we feed ourselves with words and ideas, we process them and, in doing so, we often refer to stereotypes. We cook facts, we simmer with resentment, and we cool down. We find certain opinions indigestible, some humor dry, and some comments sharp. Words make us feel sick or satisfied. Most importantly, however, we do not always find the same platter equally tasty. What one person finds savory and pungent might seem quite insipid and bland to another.

We believe that this is exactly what happens with language. Some enjoy refinement and undertones while others appreciate simplicity and directness. Irony can evoke laughter, but its humorous potential might also not be recognized and taken instead as stinging and harsh. But who likes what? What are the flavors of irony? In our attempt to understand the varied research results on social functions of irony (Kreuz et al., 1991; Dews et al., 1995; Pexman and Olineck, 2002) as well as its contradictory nature itself (Grice, 1975; Giora, 1995; Gibbs and Colston, 2007) we decided to combine the so far-disclosed qualities and components and concoct a new recipe for its nature.

To this end, we designed four scenarios gathering men and women's spontaneous responses to the same ironic criticism (*blame by praise*, BbP) or ironic compliment (*praise by blame*, PbB) voiced by either ingroup (same sex) or outgroup (opposite sex) interaction partner in control and two gender stereotype activation conditions. We decided to use BbP and PbB labels (Anolli et al., 2002) due to the fact that they seem least confusing, as compared with "critical praise," "critical blame," "ironic compliments," and "ironic criticism". As Burgers et al. (2012) rightly point out,

> the terms ironic praise and ironic blame are used in two distinct ways in the irony literature. Some authors use ironic praise to refer to ironic utterances that are literally negative, such as "That's a horrible idea" (e.g., Schwoebel et al., 2000; Filipova and Astington, 2008). In contrast, other irony scholars define ironic praise in the exact opposite way, namely, by referring to ironic utterances that are literally positive, such as "That's a great idea" (e.g., Poggi et al., 2007; Poggi and D'Errico, 2010) (p. 306).

Irony as a Verbal Dimension of Social Comparisons

In general perception, irony is a funny thing (Kreuz and Glucksberg, 1989; Roberts and Kreuz, 1994; Colston and O'Brien, 2000). Ironic remarks are viewed as more playful than literal comments (Kreuz et al., 1991; Gibbs, 2000) and people who use irony are perceived as having a sense of humor (Pexman and Olineck, 2002). Other than humor, politeness is also indicated as a communication goal of irony. Leech (1983) makes reference to his *politeness principle* and proposes an *irony principle*, where irony is seen as a way of not causing offense directly and thus preventing an open conflict. However, results of research conducted by Matthews et al. (2006) indicated that humor, but not politeness, was a significant factor in a speaker's decision to use verbal irony. According to Partington (2007), the use of irony is affiliative inasmuch as it can "bind speaker and hearer when a third party is the object of criticism, it can be used in friendly teasing or it can be used in self-deprecatory humor" (p. 1,565). While Dews et al. (1995) propose the *tinge hypothesis* and tinge function of irony, namely, muting the aggression expressed in criticism and moderating the praise communicated in a complement, Brownell et al. (1990) show that, actually, ironic criticism can be rated as "meaner" than literal criticism. Ironic comments can be perceived as "mocking" (Kreuz et al., 1991) and implying the intention of being more hurtful (Pexman and Olineck, 2002).

"The study of humor, irony, and other playful forms is plagued by definitional problems" (Attardo, 2002, p. 166) and there is no common understanding among researchers as to what irony is. From Ancient Greek, irony, εἰρωνεία *(eirōneía)*, means a pretended ignorance. According to Encyclopædia Britannica (https://www.britannica.com), the term irony has its roots in the Greek comic character Eiron, a clever underdog who by his wit repeatedly triumphs over the boastful character Alazon. From being a violation of code, a figure of speech that does not mean what it says, flouting the maxim of quality (Grice, 1975), through the game of pretense (Clark and Gerrig, 1984), to the sound of an echo (Sperber and Wilson, 1981, 1984; Kumon-Nakamura et al., 1995), and indirect negation (Giora, 1995), irony still means more than its literal words.

According to Dynel (2014), there is no clear distinction in the topical literature between humorous irony and non-ironic humor, and "linguistic phenomena displaying overt untruthfulness and humor may be easily mistaken for humorous irony. "(p. 621). Due to the fact that researchers of irony face an array of interfering variables, Burgers et al. (2012) propose five requirements for an ironic utterance to be qualified as ironic: *evaluatieveness*, *incongruence* (between the literal meaning of the irony and its co- or context), *reversal of valence* (i.e., irony with a positive literal meaning, as in "Good idea, John!" when the idea was bad or irony with a negative literal meaning, as in "Bad idea, John!" when the idea was good), *target* (irony is always aimed at somebody or something), and *relevance to the communicative situation*. Instead of five, Dynel (2014) proposes a set of two "conditions serving as an acid test for irony," namely: (a) overt untruthfulness and (b) negative evaluation.

Given the multiple definitional problems and operational challenges resulting from lack of any specific measure of irony, it is not surprising that research on irony brings conflicting and contradictory results. However, differences in irony perception might also result from the simple fact that we differ in how we see the world, what we like about it and how we describe it; this may explain why there is systematic variance in irony detection performance (Bruntsch et al., 2016). Previous research by Milanowicz (2013) showed that that men would use irony with the aim to amuse others, to make fun, and to be perceived as funny, but women would rather use ironic comments to show their disapproval and smuggle in more anger and meanness. Seeing clearly the variability in the load of ironic comments, we agree with Jorgensen (1996) that in order to see how irony can be used as an effective tool for communication, we should

look into the perceptions of ironic instances. This is exactly what we are doing here. We measure production and evaluation of irony in relation to individual differences, such as gender and anxiety.

Irony can convey meanings outside the humorous frame, unlike teasing, but is not necessarily aimed at hurting others, unlike sarcasm. Essentially, according to Dynel (2014), the difference between irony and sarcasm comes down to the overt untruthfulness typical of irony, while the difference between irony and teasing is based on the negative evaluation, not always present in the latter category of humor. As mitigating, funny, critical, or mean as irony might seem, still, the ballistic repertoire used to describe its humor and "barbs" tips the scales of verbal interaction more in favor of the battlefield rather than a playground setting. In the subject matter literature, we encounter "targets" or "victims" of ironic comments, we read about "aims" of ironic remarks and "face-threatening" or "face-saving" techniques. Given that we are discussing hidden meaning, where *implicatures* must be pulled out from communication like rabbits from a hat, maybe it would be more appropriate to employ more of a wording that alludes to verbal illusion.

Also illusory for some researchers is the possibility of so-called asymmetry in irony, that is, the notion of critical (negative) irony being more frequent than praising irony (Sperber and Wilson, 1981; Clark and Gerrig, 1984; Matthews et al., 2006) While most theories approach "ironic praise" and "ironic blame" as two categories of the same genre, some researchers (Garmendia, 2010) voice their veto and deny the possibility of being ironic without criticizing —stating that the asymmetry issue is an illusion. We were quite inclined to follow that path, until our research results made us look at this asymmetry through the lens of the *linguistic intergroup bias* theory (Maass et al., 1989; Maass, 1999; Wigboldus and Douglas, 2007), which states that positive ingroup descriptions and negative outgroup descriptions are abstract and vague, while negative ingroup descriptions and positive outgroup descriptions are specific and observable. In other words, desirable behavior of ingroup members is interpreted on an abstract cognitive level, while negative behavior is interpreted on a more concrete level. This phenomenon is reversed when interpreting outgroup member behavior, which helps maintain a more positive image of one's own group.

We assume, in light of this theory, that if we accept irony as the manifestation of non-literal, indirect, and somehow abstract language, as opposed to direct and literal language, we can expect that when addressing one's own group (the ingroup, i.e., a same-sex interlocutor), irony should be used rather in a positive context (e.g., to praise), while literal language will be employed in a negative context. The opposite will be true toward anyone in the outgroup (i.e., an opposite sex interlocutor), and use of irony should then be preferred in negative contexts (e.g., to criticize). Thus, we assumed that women and men will understand irony communicated by a same-sex person differently than if communicated by an opposite-sex person, and we ventured into gender stereotype activation in the process of verbal communication.

Irony and Social Comparisons Based on Gender

Dunin (2001) attributes the common perception of gender to stereotypes based on polar opposites: aggressive and gentle; insensitive and caring; mathematician and linguist; talkative and quiet; logical and intuitive; competitive and co-operative, pink and blue. For Dunin, people are different from one another, regardless of gender, but it is precisely "the stereotypical visions of femininity and masculinity, ingrained in our minds and cultures, that differentiate people in an extremely simplified way." (Dunin, 2001)

It seems understandable that by attaching importance to these differences, which we are fed since birth and which we feed on every day, it is difficult not to react automatically to the "other" for which, in this paper, we use the term *outgroup member*. Gender differences in self-construal are to a large extent the product of social comparison processes (Guimond et al., 2007) and not any comparisons, but especially intergroup comparisons (Guimond et al., 2006). The conceptions of ingroup identity varying as a function of comparative context are already present in children (Sani et al., 2003). We self-categorize and categorize others in an attempt to balance our identity between individualism and a sense of belonging without risking either alienation or loss of identity (Turner et al., 1987; Guimond et al., 2006).

Recchia et al. (2010) showed that in family conversations at Canadian homes, mothers were likely to ask rhetorical questions and use ironic language in conflictual contexts, while fathers used hyperbole and understatement as frequently as rhetorical questions, and employed ironic language in both positive and conflictual contexts. Kałowski (2017) introduced audiovisual stimuli (recordings of women and men making ironic statements directed at the participant) and collected data in the form of recordings of utterances, and their analysis was consistent with previous research showing that males feel more positive about using irony (Jorgensen, 1996; Colston and Lee, 2004; Milanowicz, 2013). Self-stereotype activation yielded higher humor ratings of irony than non-irony used by male (but not female) actors in the stimulus videos. Thus, it is possible that "meaning well when using irony" is part of a male stereotype accessed by both genders in intergroup comparison conditions. Analogously, "not meaning well when using irony" might be an element of a female stereotype. These results also suggest that men and women might use irony for different reasons (Milanowicz, 2013; Milanowicz and Kałowski, 2016; Kałowski, 2017).

On a regular basis, we are involved in social networks and growing within the frames of our times and mores, we take it as natural to put labels onto things we see. We name these things, like them, hate them, use them, and talk about them, but do we understand them? Because of the pervasiveness of "name tags" and identification labels in social interactions, social inclusions and exclusions, of which gender stereotypes are also a manifestation, we decided to throw down the gauntlet and see how blue boys play "irony games" with pink girls.

We agree with Hyde (2005) that gender is not only a person variable but, most importantly, a social stimulus, whose

activation, like the activation of any other stereotype, impacts an individual's perception, judgments and behavior (Bodenhausen and Lichtenstein, 1987; Devine, 1989; Bargh et al., 1992; Macrae et al., 1994). We understand that gender studies are quite controversial and raise a lot of emotions. This can also be one of the reasons why, maybe quite surprisingly, given the vastness of the phenomenon, there is not much research on irony and gender. The aim of this research is not to present any misconception of gender equalities or differences. We take the category of gender as a classification subcategory, a social cluster which might help to explain the ambiguity of cognitive meanings and verbal behavior engaged in the response to one and the same stimulus: irony.

Gibbs (2000) reported that men were more likely than women to use sarcastic irony in conversation with friends. Jorgensen (1996) examined the effect of gender on the social and emotional impact of irony and reported that men were more likely than women to perceive humor in sarcastic irony and women were more likely than men to be offended or angered by sarcastic remarks. Same results were obtained in research by Milanowicz (2013). Katz et al. (2001) have investigated whether gender, as a social category, could suggest a speaker's tendency to make ironic remarks. In the light of their data, men were perceived to be more sarcastic than women. Holtgraves (2005) found gender differences in how participants rated their own tendencies to speak sarcastically, with male participants presenting higher self-reports of use of sarcasm than female participants. However, most of the research on irony might raise the question about consistency between what people claim to be (on paper or in the laboratory condition) and how they actually behave.

Lampert (1996) suggested that the primary motive for men using conversational humor is the reduction of social vulnerability: "irony can serve the self-protective function that Lampert claims is important to men and, indeed, men's ratings suggested they were more likely to use irony in most situations" (Ivanko et al., 2004, p. 266). Holtgraves (1997) showed that men rated themselves higher than women on the production factor of the Conversational Indirectness Scale (CIS)—devised to measure individual tendencies to express and interpret meanings indirectly. Irony, as the manifestation of indirect criticism, can give the impression of politeness. "Participants with higher CIS–production scores seemed more apt to recognize this politeness function, whereas female participants tended to recognize the critical (and thus impolite) function of ironic criticisms" (Ivanko et al., 2004, p. 265). In the research by Ivanko et al. (2004), females rated ironic compliments as being more sarcastic than did the male participants. The authors explain it by greater sensitivity on the part of female participants to the negative tinge of ironic compliments (Dews et al., 1995; Pexman and Olineck, 2002) The gender differences reported in the research by Ivanko et al. (2004) replicated Jorgensen's (1996) observed differences between male and female perception of politeness in ironic comments.

Colston and Lee (2004) found that irony is considered a more male-like than female-like form of communication by both men and women, reporting that "fictional speakers of unknown gender who use verbal irony to comment about relatively negative situations are thought to most likely be male" and that "males report a greater likelihood of using verbal irony in negative situations" (Colston and Lee, 2004, p. 301). They posit that men tend to use irony more often than women because their pragmatic goals in conversations more often include expressing a critical lack of approval. Alternatively, men could be more ironic because they show a greater propensity toward risk-taking (Colston and Lee, 2004) and use of irony involves a certain risk of being misunderstood. We assume that those differences might be explained not only by the willingness to take risk, but also in terms of reluctance and fear to venture into what is unknown and/or ambiguous, of which non-literal language is a representation on the symbolic level.

Irony and Anxiety

Similarly to most research on humor, most studies on irony also focus on its positive qualities.

> Ruch and Proyer (2008a,b) were the first to study gelotophobia (the fear of being laughed at) empirically as an individual differences variable that characterizes the degree to which people fear being laughed at by others. (Chłopicki et al., 2010, p. 172)

Comparatively, we believe that the application of the anxiety measure (STAI) combined with the analysis of not only funniness but also meanness in perception of verbal irony (IPT) allows for a more dimensional approach to the whole concept. It goes without saying that that *laughing at* and *laughing with* are not equipollent. Introduction of anxiety to the research on irony can be seen as a prelude to further assessment of the links between the perception of being laughed at and the motivation to laugh at others or ridicule them. Also, the Polish GEOLPH <15>, the Polish adaptation of the Inventory for Assessing Gelotophobia by Ruch and Proyer (2008b), showed that the fear of being laughed at existed widely independently from the age or sex (Chłopicki et al., 2010). Research on irony showed differences between men and women and thus we are curious to know if these differences are related to anxiety?

Both anxiety and irony relate to emotional experience and lead to emotional responses. Irony as an unexpected and ambiguous stimulus can evoke a state of "fight or flight" alertness. It is believed that, in order to arrive at a response to such a stimulus, we instinctively refer to our cognitive schemas, personal knowledge, and emotional attitude. Attitudinal responses are evaluative, and evaluation is connected with the imputation of some degree of goodness or badness to an entity (Lewin, 1935). Valence refers to intrinsic attractiveness (positive valence) and aversiveness (negative valence) of an event, situation, object, or stimulus (Lewin, 1935; Damasio, 1994), thus, affective valuation should be viewed as an integral part of meaning.

Irony being an ambiguous stimulus can be perceived not only as a harmless joke but also as threatening. Studies on interpretive and judgment biases indicate that they are already present in children with anxiety, leading them to interpret ambiguous stimuli as threatening (Taghavi et al., 2000) and exhibit avoidant responses (Chorpita et al., 1996). Another study on individual differences in children exploring associations between verbal irony comprehension and shyness (Mewhort-Buist and Nilsen,

2013) reported that shyer children ascribed a greater degree of negative attitude to speakers who made ironic criticisms. It was also demonstrated that children higher in shyness showed less appreciation of the irony muting effect.

We thus hypothesized that higher anxiety levels in adults could also lead to defensive performance in humor appreciation of irony. However, in line with the widely accepted belief that women are more emotional (Brody and Hall, 2008), more emotionally expressive than men (Kring and Gordon, 1998), and more likely to suffer from clinical anxiety than men (Remes et al., 2016), it was believed also that this effect will be moderated by gender.

Women show a greater tendency than men to interpret utterances as figurative (Holtgraves, 1991). However, a higher level of anxiety could account for the perception of ironic (ambiguous) comments as being more threatening because it seems more unknown and so more scathing than literal criticism.

Also, gender stereotypes provide a basis for socializing boys and girls about appropriate emotional behavior, where expressing fear and sadness is acceptable for girls but not for boys (Brody, 2000; Chaplin and Aldao, 2013). This emotional double-standard associated with the stereotype serves the function of preserving the social hierarchy, where women are viewed as irrational and uncontrollable and thus dangerous, legitimizing women's subordinate rank in the power hierarchy (Lutz, 1996).

We approach gender as the set of behaviors and attitudes that characterize people of a given biological sex. In this paper we write about gender (variable) because we show the existence of different patterns of verbal behavior in men and in women. These differences can also result from acceptance of different social roles pertaining to sex.

We have also decided to give importance to gender and anxiety in irony research because women are almost twice as likely as men to experience anxiety. This gender gap might result from physiological factors but also might be related to differences between men and women in how they cope with stress (Remes et al., 2016).

Therefore, it remains important to consider how inequalities among men and women in this respect might contribute to their different approaches to irony.

Due to its ambiguous nature (uncertainty as to the real meaning and interpretation), irony can possibly be a stressful stimulus for some people. Therefore, we deem it legitimate to include anxiety as an individual variable modifying human reactions to irony.

MATERIALS AND METHODS

Ethics Statement

This study was carried out in accordance with the recommendations of the Academic Ethical Review Board (Scientific Research Ethics Committee of the Faculty of Psychology, University of Warsaw). The participants provided verbal informed consent to take part in the study. Such a form of consent is customarily used in Poland in studies on adult student samples. The consent procedures were detailed in a description submitted to the institutional review board (Ethics Committee of the Faculty of Psychology, University of Warsaw), where they were granted final approval in October 2014. The participants were granted full anonymity of the data gathered for the analyses and were informed that only group results will be described.

Participants

Participants were recruited from among students (University of Warsaw and Warsaw University of Technology) and public institution employees. They participated voluntarily in the study and returning the completed questionnaires meant their consent to take part in the study. The total sample consisted of 238 subjects ($M_{age} = 23.92$; $SD = 8.120$): 127 females, age ranged from 18 to 44 ($M_{age} = 21.31$; $SD = 4.727$) and 111 males, age ranged from 18 to 60 ($M_{age} = 26.89$; $SD = 9.987$).

Measures

The State-Trait Anxiety Inventory (STAI, Polish adaptation, Spielberger et al., 1987)—was distributed among participants in order to verify if anxiety can be shown to predict perception of ironic funniness or ironic meanness. The STAI contains two 20-item scales measuring state and trait anxiety. All items are rated on a 4-point scale, ranging from "almost never" to "almost always." Both anxiety scales were used in this study. The Cronbach's alpha for the state anxiety scores was 0.928, while the Cronbach's alpha for the trait anxiety scores was 0.897.

The Irony Processing Task (IPT, Milanowicz, 2016)—a self-report questionnaire was designed to stimulate production of non-literal comments and measure not only comprehension but, most importantly, the reaction to ironic comments. The task consists of six scenarios, each depicting a short context introduction with a simple cartoon and a comment. Due to the general belief that females tend to perform better on tasks requiring decoding of non-verbal information (Hall, 1984; Collignon et al., 2010), the IPT was designed in such a way as to eliminate those cues. Any indicators of a prosodic or kinetic character, as well as facial cues, were neither considered nor present. The cartoons and ironic comments were followed by dialogue balloons for participants to write down their spontaneous replies. Some of the scenarios were followed by questions about the motivations of the speakers and emotions of the recipients. Some other actions and comments were evaluated on two five-point rating scales of "humor" and "malice" as follows: 1 (*not at all*)—5 (*very*) The IPT was developed in order to see what relationship, if any, exists between gender and response to irony. The measure aims to see how irony is understood and produced in different communicative settings and how irony is used toward different communication partners. Also, we decided to use this mode (the written word) because this is the way in which we communicate nowadays in the era of digital communication, which seems to be taking over certain aspects of face-to-face interaction.

In this paper we describe the results of two IPT experimental tasks.

The first experimental task, IPT 1, involves four context scenarios and four target statements, where the participant is asked to imagine that each ironic comment is voiced directly toward him or her. Half of these comments (one expressed by

female and one expressed by male) are ironic criticism, (BbP, i.e., positive literal meaning but negative true meaning, like "Genius!" when the idea made by the participant is very bad and what the speaker actually means is "This is so stupid"); the other half (again, one expressed by female and one expressed by male) are ironic praise (PbB, i.e., negative literal meaning but positive true meaning, like "I can see you're taking it easy" said when the participant is staying up late and what the speaker actually means is "I see you're working hard"). The hearers can choose to respond either to the dictum or the implicatum or they can engage with both meanings.

The second experimental task, IPT 2, presents two BbP criticism scenarios where research subject is no longer the direct target of the ironic comment but an observer. These stimuli are presented in two conditions: (a) stereotypically male activity—two males playing football and one misses the ball and (b) stereotypically female activity—a woman being a bad driver and taking up two spots in a parking lot. The cartoons present ironic comments expressed by one of the characters toward the friend who failed: (a) "Nice skills!"—implying good agility while, actually, his agility is poor, and (b) "Nice skills!"—implying she is a good driver, while she actually is not. The two scenarios are followed by questions about the motivations and emotions of the characters. Participants also rate ironic comments on two 5-point Likert scales for their: (a) humor—funniness, the quality that makes the comment amusing and (b) malice—desire to harm others, malicious intent and ill will, as opposed to the humorous potential of a comment.

Gender stereotype self-attribution tool—two lists of personality adjectives were used as the measure of gender stereotype activation: one list consisted of 16 positive trait adjectives and the other one consisted of 16 negative trait adjectives, where 2/3 of the adjectives related to either male or female stereotype, (e.g., independent, self-confident, brave vs. caring, sensitive, emotional, etc.) and 1/3 were considered neutral (e.g., smart, creative, arrogant, conceited).

Procedure

Participants were not told that the study specifically concerned irony. It was only explained that we were interested in knowing how people perceive and react to certain situations and comments. Participants were instructed that there were no good or bad answers. To assure study validity, male and female participants were randomly assigned to one of three experimental conditions (control, positive pretask priming, and negative pretask priming). We obtained six data sets: male control group ($n = 44$), female control group ($n = 56$), male ($n = 32$), and female ($n = 35$) groups with positive pretask priming, and male ($n = 35$) and female ($n = 36$) groups with negative pre-task priming.

Other than control, two experimental conditions were employed, where participants were conditioned by the selection of personality adjectives made available to them. The goal of the pretask conditioning was to make gender salient in order to activate gender stereotypes and induce gender stereotype-congruent inferences in the subsequent IPT. Pretask priming was based on gender self-stereotyping and the intergroup comparisons effect (Guimond et al., 2006).

In the positive pretask priming group, a social comparison paradigm was employed with a list of 16 positive trait adjectives. Participants were asked to select semantic attributes which they judged as more self-relevant when comparing themselves with outgroup members (represented by males if participants were females, or by females if participants were males). Positive trait adjectives were used with the aim of reinforcing a positive image of the self and other ingroup members at the expense of the outgroup members.

In the negative pretask priming group, a social comparison paradigm was employed with a list of 16 negative trait adjectives Participants were asked to select semantic attributes which they judged as more self-relevant when comparing themselves with out-group members. It was believed that the exposure to negative trait adjectives would provoke a negative image of the self and other ingroup members, but not outgroup members.

Conditioned self-attribution in positive and negative pre-task priming groups is illustrated by **Figure 1**.

The mention of "males" and "females" in this experimental procedure was believed to activate stereotypical knowledge of gender-relevant characteristics, which would be acknowledged as informative about the person and have an impact on the interpretation of what that person is saying. It was expected that gender category labels provided in the pretask priming would activate categorical representation and linguistic profiling.

The IPT 1 tested participants' responses to different speakers of ironic comments (ingroup or outgroup members) and to different types of irony under three different experimental conditions. The experimental design was a 2 (male vs. female) × 2 (ironic criticism vs. ironic complement) × 3 (control condition vs. positive pretask priming vs. negative pretask priming).

Where the main analyses were designed in 2 × 2 × 3 analysis of variance (ANOVA) plan, the complimentary variables were checked as unifactorial dependencies.

The data from the experiment was also used in unifactorial analyses of the relationship between (a) the level of anxiety, (b) the type of response to ironic comment (criticism and complement), (c) gender of the speaker, (d) gender of the recipient, that is, the participant (who from the role of the recipient of the ironic comment becomes herself or himself the sender of the message in either ironic or non-ironic exchange of comments).

The second experimental task, IPT 2, tested participants' perception of (a) humor and (b) malice in ironic criticism (BbP) in two different settings, involving either two ingroup members or two outgroup members. Evaluations of humor and malice were rated for each scenario on 5-point Likert scales.

In the control group, first the IPT, and then the STAI were administered to research participants. In the experimental groups, the lists of adjectives were administered first, followed by the IPT and the STAI. The order of presented tasks was kept the same for all tested individuals.

Research Objectives and Hypotheses

It was hypothesized that males and females would respond differently to ironic comments coming from ingroup (same sex) or outgroup (opposite sex) members.

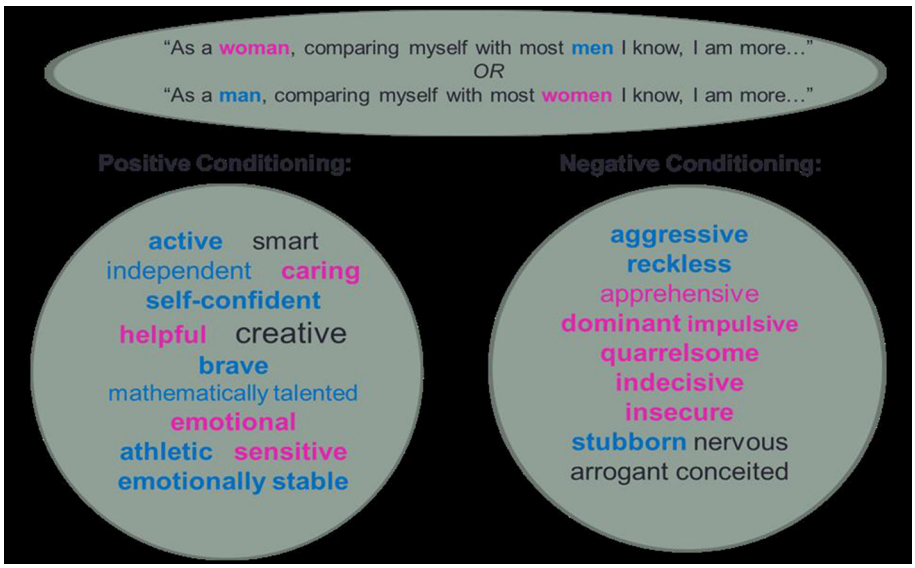


FIGURE 1 | Conditioned self-attribution.

It was also believed that application of the pretask priming on gender stereotype activation would reinforce gender differences in attributing different meanings, congruent with these stereotypes, to ironic utterances. We hypothesized that reactions to a negative stereotype (list of negative trait adjectives), that is, a stereotype threat, would negatively impact participants' test performance (lower use of irony, lower ratings of humor, and higher ratings of malice) as opposed to the stereotype boost context, where the presentation of the list of positive trait adjectives would lead participants to improved test performance (higher use of irony, higher ratings of malice, and lower ratings of humor). It was also expected that the ironic setting (male-male vs. female–female) may determine whether relevant stereotypes are activated and it might influence the perception of irony. This is in line with Wigboldus et al. (2005) suggestion that

> in an intragroup context (e.g., when females talk to females about females) a target's category membership (e.g., gender) is less likely to become salient. Consequently, stereotypic expectancies with this category are not activated, thus rendering it unlikely that linguistic biases occur. In an intergroup context, however (i.e., when either target or recipient is an outgroup member), a required category activation is more likely, and linguistic bias is expected (Beukeboom, 2014, p. 17).

We also believed that subjects with lower levels of anxiety would be more ironic in their responses to ironic comments than subjects with high levels of anxiety.

Given previous studies, it was hypothesized that males would rate the humorous potential of ironic comments higher than females. It was also hypothesized that females would rate ironic comments as more snarky and snide than males.

It was also believed that subjects with low levels of anxiety, regardless of their sex, would rate humor higher on the Likert scale.

We expected a cross-gender effect in the evaluation of humor and malice in the comments, that is, male participants giving higher rating of humor and lower rating of malice to the comment involving two outgroup members (female-female scenario) and the reverse trend in the group of female participants who would rate humor higher rather in the male-male than in the female-female scenario.

Data Analysis

Due to the exploratory nature of the study and the very distinct quality of irony in its openness to more than one interpretation, it was deemed reasonable to see what categories emerge from the collected data. Data was coded with reference to categorization by Kotthoff (2003), the classification by Clark (1996, after Hancock, 2004) and the taxonomy of irony factors and irony markers by Burgers et al. (2012). The resulting set of three categories was checked with and confirmed by two other coders. The whole multi-task IPT instrument is based on each subject's unique responses, and its reliability has been proved by high inter-rater consistency. The fourth category of "no evidence" (e.g., silence, changing the subject, no response) will not be discussed in this paper, so we are not considering it further. The two-factor analysis, based on (a) irony recognition and (b) type of response resulted in the following categories of response to irony:

1. *Ironic response* (Ir)—to what is meant: ironic intent recognized/ irony extended—acknowledgment of ironic intent and reply in terms of non-literal or equivocal language, also, but not only, with the use of ironic markers such as metaphor, hyperbole, understatement, rhetorical question, echo/repetition, question marks, emoticons, quotation marks, diminutives, laughter onomatopoeias (haha, ha ha, ahahaha, hehe, ehehehe, hohoho, ho ho);
2. *Misresponse* (Mr)—to what is said: ironic intent not recognized/ literal response—when ironic intent is missed or

misinterpreted and the reply suggests that the ironic comment is not clear or taken literally (e.g., "What do you mean?," "Really?," " Do you really think so?");

3. *Literal response* (Lr)—to what is meant: ironic intent recognized/ literal response—detection and correct interpretation of the ironic intent but irony is not extended and the response is literal and direct.

We also singled out *Laughter category* with all "hahahah," ha ha ha," emoticons, or words such as "laughter," "smile," "joke," "funny," "you must be joking," "ha ha ha really funny" (even when implying: not funny at all) that are an expression of both humor and indignation. However, due to the fact that replies containing "laughter markers" were only a few, we did not include this category in our further analysis.

Perception of humor and malice, rated on 5-point Likert scales, were counted and analyzed separately for each experimental condition (control, positive pre-task priming, and negative pre-task priming). Also, ratings of humor and malice for each participant were analyzed with regard to their level of anxiety measured with the STAI.

RESULTS

Data shows that one comment can trigger completely different verbal behaviors, ranging from acknowledgment to complete disbelief or rejection.

Reactions to Ironic Blame by Praise vs. Praise by Blame by Gender

The ANOVA model with mixed design was applied to analyze the results. The sex, priming, and anxiety (state and trait) were between-subject factors, while the experimental variables of sex of the interlocutor and irony type (BbP vs. PbB) were defined as within-subject variables. The occurence of three reaction types was the dependent variable, analyzed in three separate analyses.

Blame by Praise and Pre-task Priming Effects (Interaction of State, Priming, and Sex)

In the paradigm of ironic response to irony, the main factor was that of the interlocutor's sex, $F_{(2, 234)} = 100.28$, $p < 0.001$, $\eta^2 = 0.300$, where males are frequent targets of ironic responses. Male subjects were also more ironic, $F_{(1, 234)} = 5.36$, $p = 0.022$, $\eta^2 = 0.022$, but there was less irony in case of negative priming $F_{(2, 234)} = 3.09$, $p = 0.04$, $\eta^2 = 0.026$.

Figures 2, **3** illustrate the proportions of different types[1] of response to BbP in the three experimental conditions.

There was an interactive effect of state anxiety moderating the above dependencies, $F_{(4, 234)} = 4.444$, $p = 0.002$, $\eta^2 = 0.077$. Anxiety as a trait did not moderate the above effects. Women with median anxiety were surprisingly more ironic in the condition of positive pretask priming, while in the two other groups, they tended not to be ironic in their responses. The proportions of ironic responses and misresponses from women and men with different levels of state anxiety are presented in **Figures 4–6**.

Misresponse to irony depended on "who is speaking to whom." Significant was the sex of the interlocutor, $F_{(1, 234)} = 35.09$, $p < 0.001$, $\eta^2 = 0.13$, but more interesting was the interaction between the sex of the respondent and his/her interlocutor, $F_{(1, 234)} = 9.80$, $p = 0.002$, $\eta^2 = 0.04$. Males reacted with misresponse in an almost similar way to different interlocutors, while female subjects were significantly more likely to react with misresponse to a female interlocutor.

In a supplementary analysis, we also observed that high-anxious women mostly misresponded in the positive priming condition.

Anxiety as a trait did not moderate the above effects.

In literal responses, the main effect of the interlocutor was observed, $F_{(1, 234)} = 15.63$, $p < 0.001$, $\eta^2 = 0.063$, as well as the effect of the subject's sex, $F_{(1, 234)} = 7.10$, $p = 0.008$, $\eta^2 = 0.029$. No higher interactions were present.

No effects of state or trait anxiety were observed.

Praise by Blame and Pre-task Priming Effects (Interaction of State, Priming, and Sex)

The analysis of the expected interaction between the priming effect, subject's sex, and interlocutor's sex in ironic response to irony was not significant, $F_{(2, 234)} = 2.25$, $p > 0.1$, $\eta^2 = 0.02$. However, the main effect of the subject's sex was significant, $F_{(2, 234)} = 12.31$, $p < 0.001$ $\eta^2 = 0.05$, just as the effect of the interlocutor's sex, $F_{(2, 234)} = 7.31$, $p = 0.007$, $\eta^2 = 0.03$, but they were additive. Irony was directed more frequently at women, independently of who the speaker was. Males turned out to respond more with ironic responses (to both men and women). Priming had no direct or interactive effects in this case. Supplementary analyses did not confirm any interactions with state and trait anxiety.

Figures 7, **8** illustrate the proportions of different types[2] of response to PbB in the three experimental conditions.

Misresponse to irony was significantly more frequent in female than in male subjects. Neverthless, the expected interaction (priming × subject's sex × interlocutor's sex) was not confirmed, $F_{(2, 234)} = 1.01$, $p > 0.1$, $\eta^2 = 0.01$. No other main or interactive effects were confirmed. No interaction with state anxiety was observed, only a weak non-linear interaction of trait anxiety with the subject's sex and the interlocutor's sex was significant, $F_{(2, 208)} = 3.25$, $p = 0.044$, $\eta^2 = 0.03$. Higher levels of misresponse probability were observed in low-anxiety males and medium-anxiety females when reacting to irony coming from male.

[1]The participants' responses to ironic Blame by Praise were classified on the basis of ratings from three independent judges into three categories (*ironic, misresponse, literal*). The inter-rater consistency was measured with Kendall's coefficient; $W = 0.987$ for male responses and $W = 0.993$ for female responses. Kendalls' W was calculated from data obtained from 111 men and 127 women.

[2]The participants' responses to ironic Praise by Blame were classified on the basis of ratings from three independent judges, into three categories (*ironic, misresponse, literal*). The inter-rater consistency was measured with Kendall's coefficient; $W = 0.988$ for male responses and $W = 0.984$ for female responses. Kendalls' W was calculated from data obtained from 111 men and 127 women.

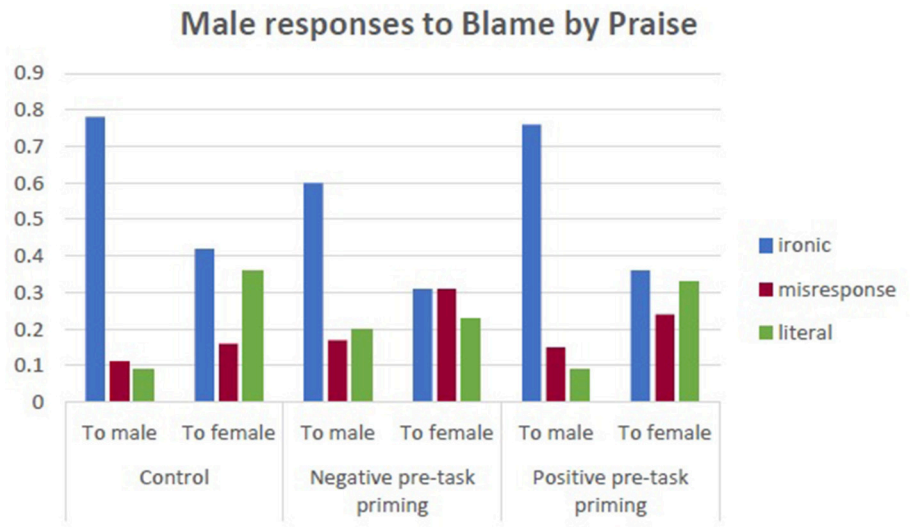


FIGURE 2 | Male responses to Blame by Praise in three experimental conditions.

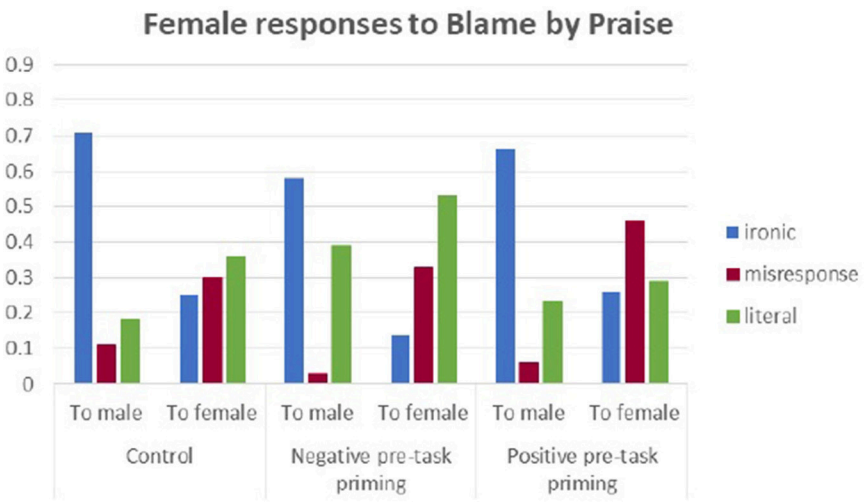


FIGURE 3 | Female responses to Blame by Praise in three experimental conditions.

Literal responses were much more probable toward male than female interlocutor, but no expected interactions of sex, interlocutor, and priming was observed, $F_{(2, 208)} = 3.25$, $p = 0.044$, $\eta^2 = 0.03$, so the effect was not moderated by sex and priming. No anxiety influence or interactions were found.

Comparative Analysis of Ironic Setting and Perception of Humor and Malice in Irony

Perceptions of humor and malice in the male-male situation vs. female-female situation were analyzed with the dependent-samples *t*-test (Supplementary Table 1), performed in six groups (for men and women in the three experimental conditions).

In the positive condition, men (see **Figure 9**) perceived irony in the male-male setting as more malicious ($t = 3.20$, $p = 0.003$) than in the female-female setting. In the control and negative conditions, women (see **Figure 10**) perceived irony in the male-male setting as more malicious than in the female-female setting ($t = 3.97$, $p = 0.000$ in the control and $t = 2.77$, $p = 0.009$ in the negative condition, respectively).

No differences in ironic humor perception between male-male and female-female settings were found (Supplementary Datasheet 1).

Anxiety and Irony Perception by Male and Female Subjects in the Three Experimental Conditions

When investigating the relationship between state/trait anxiety and irony perception, we computed the Kendall-s Tau correlations between the standardized STAI scores and

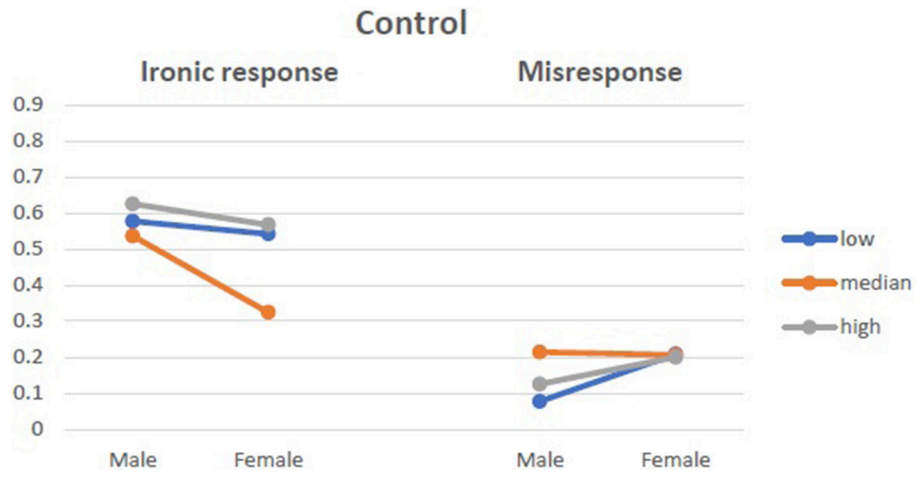


FIGURE 4 | Ironic response vs. Misresponse to Blame by Praise in relation to state anxiety in control group.

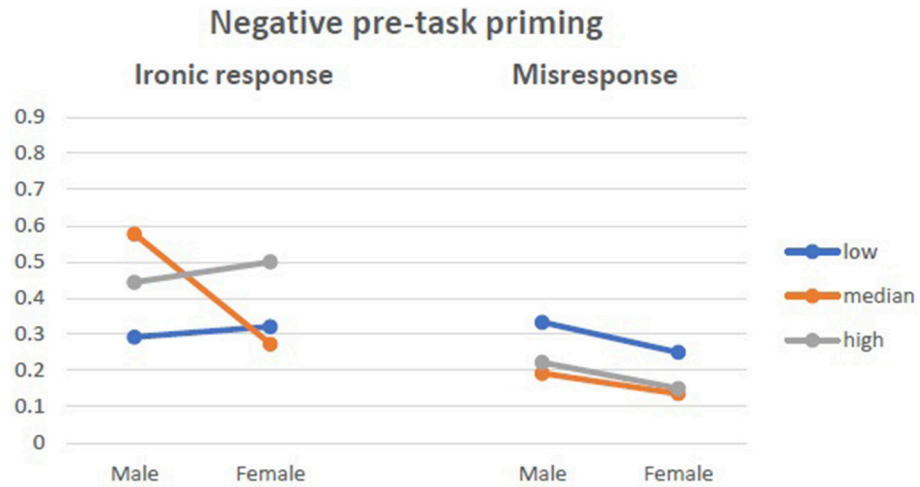


FIGURE 5 | Ironic response vs. Misresponse to Blame by Praise in relation to state anxiety in negative pre-task priming.

two scales assessing subjective perception of ironic statements (they were assessed independently in the aspects of malice and humor). The coefficients have been computed in six groups (Supplementary Table 2): for male and female subjects, according to the three experimental conditions.

In the control conditions, males with higher state anxiety perceived ironic statements made by women as more humorous ($tau = 0.251$, $p = 0.028$), and statements made by men as less humorous ($tau = -0.245$, $p = 0.030$).

In the negatively primed group, the second effect was similar: high state-anxious men perceived more humor ($tau = 0.239$, $p = 0.047$) in women's ironic statements. In the positive condition, men with higher state and trait anxiety perceived women's ironic statements as less malicious (respectively $tau = -0.252$, $p = 0.041$, and $tau = -0.238$, $p < 0.048$), and high trait-anxiety men assessed men's ironic statements as funnier and more humorous ($tau = 0.263$, $p < 0.031$).

In the control and negative conditions, women presented no significant correlation between anxiety and irony perception. In the positively primed group, women with high trait anxiety perceived men's ironic statements as slightly more malicious ($tau = 0.238$, $p = 0.047$) but they perceived women's ironic statements as definitely less malicious (the strongest correlation in the study: $tau = -0.461$, $p = 0.001$).

DISCUSSION

Gender Effect in Irony Distribution

The exploratory data analysis of reactions to irony showed a relationship between gender and response to irony, in other words, between who says what to whom, which we call the *outlook effect* (the mental attitude which refers to the gender bias in use of verbal irony in communication, where the sender modifies the content of the message by its recipient, just as in

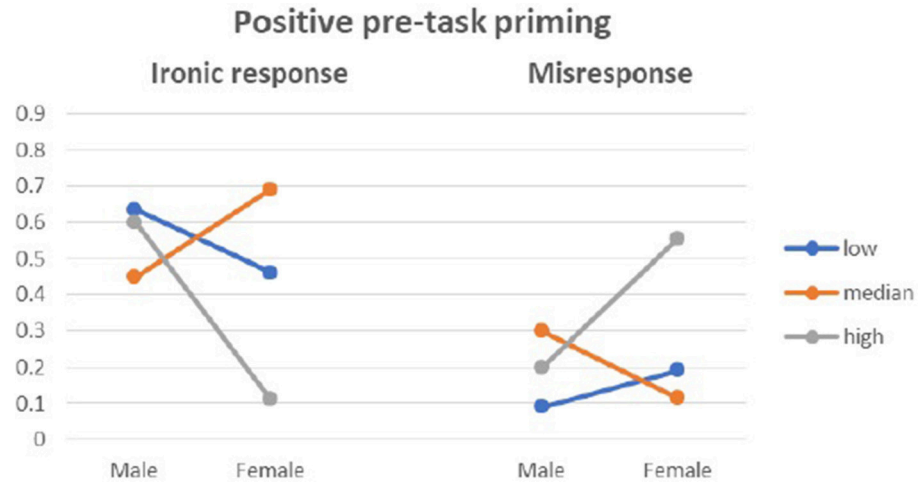


FIGURE 6 | Ironic response vs. Misresponse to Blame by Praise in relation to state anxiety in positive pre-task priming.

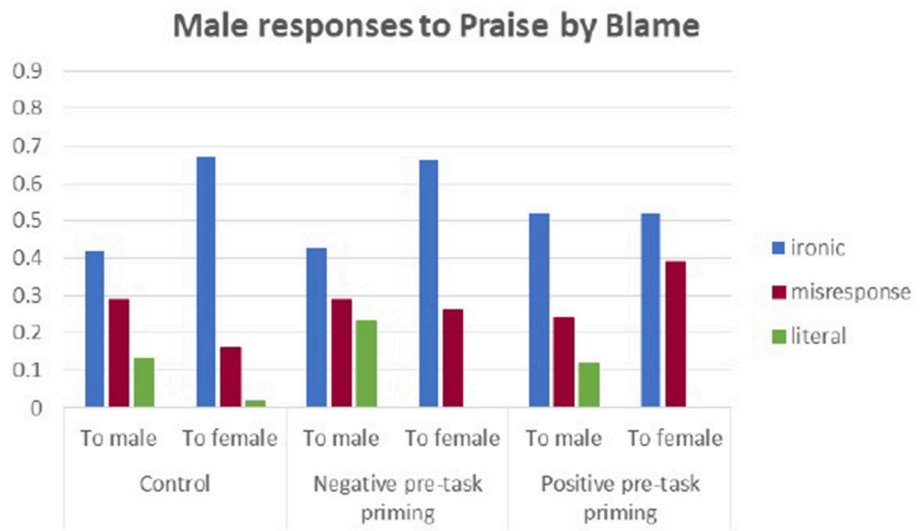


FIGURE 7 | Male responses to Praise by Blame in three experimental conditions.

personal email correspondence carrier). In five out of six male experimental groups, most participants reacted to irony with irony, in both BbP and PbB contexts, regardless of whether they spoke to another male or female. Ironic comments simply triggered mostly ironic replies in men. The results of our research are also in line with Holmes' claim that "women and men develop different patterns of language use" (Holmes, 1998, p. 462). Ironic response was significantly more frequent as a reply to males than to females. In the negative pretask priming group, ironic responses and misresponses became equally frequent categories of the reply to an ironic BbP comment from female.

While men turned out to be pretty stable in how they replied to irony, in all experimental conditions, regardless of the recipient's gender, female participants proved to be more unpredictable in their reactions. An ironic response (Ir) from a female to an ironic BbP comment coming from a male (the opposite sex, i.e., an outgroup member) showed to be the most frequent category in all the experimental conditions. This pattern of Ir > Lr> Mr was broken by the change in the sex of the interlocutor and by the shift to the PbB context. The most frequent category of reaction to BbP comment made by females was literal response (Lr, for thecontrol group and the positive pretask priming group). Misresponse (Mr, where criticism was mistaken for comfort and reassurance or when the superficial level of praise was genuinely acknowledged) was the most frequent category in the negative pretask priming female group.

In the PbB context, women replied most frequently with misresponse, regardless of who they spoke to (another female or a

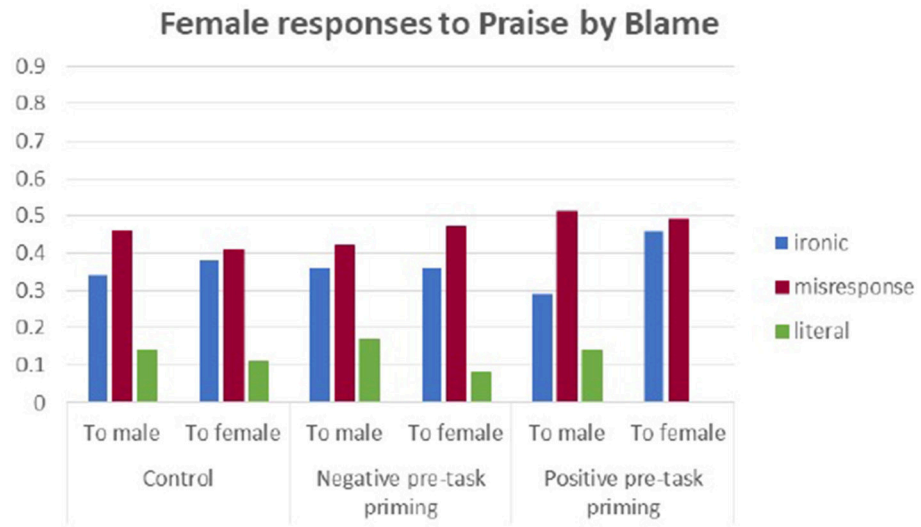


FIGURE 8 | Female responses to Praise by Blame in three experimental conditions.

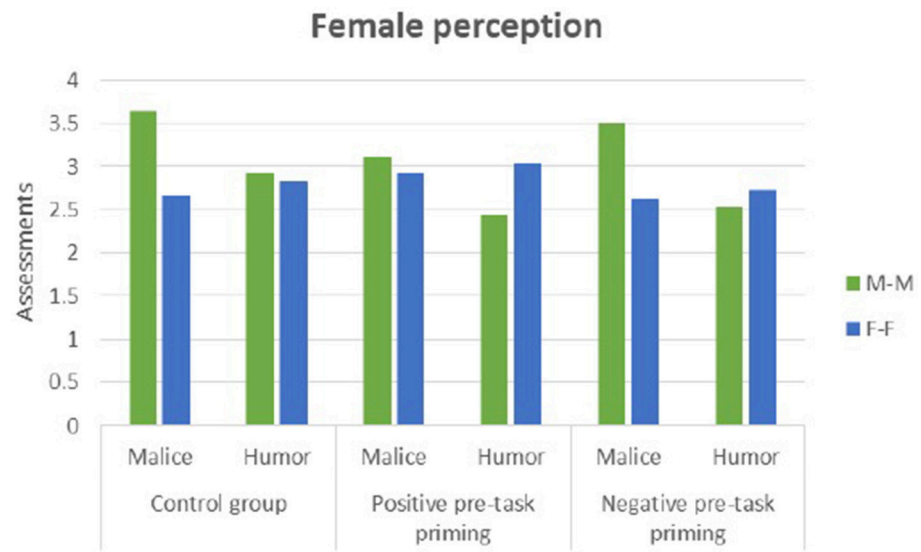


FIGURE 9 | Male perception of humor and malice in male-male (M-M) and female-female (F-F) dyads.

male) and this pattern was kept in all six experimental conditions. Ironic replies to another female were also quite frequent, while literal replies were rare and used more toward outgroup (male) members.

In the PbB context, men were more ironic in their responses to females than males. They still kept the significantly high frequency of ironic responses, but the proportion of use was reversed when compared with the BbP context, where they were more ironic toward same-sex interlocutors (i.e., another male).

We can see that two types of contextual frames constructed to convey (a) a desirable situation calling for a praise (a positive context condition) and (b) an undesirable situation endorsing criticism (a negative context condition), provoked different types of reactions.

In the BbP context, females used more irony to males but not to another female. In the PbB context, they reacted mostly with misresponses to both the ingroup and outgroup members. Additionally, literal responses were given more frequently to males (outgroup members) than females (ingroup members). However, male participants were more ironic to males in the BbP context, but to females—in the PbB context. Also, only a few literal comments were used with females and significantly more were used with males in the PbB context.

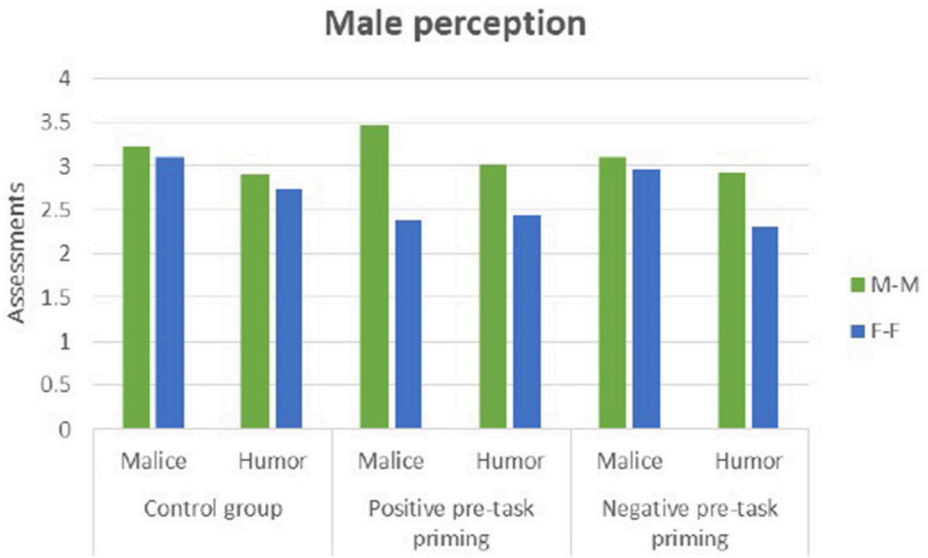


FIGURE 10 | Female perception of humor and malice in male-male (M-M) and female-female (F-F) dyads.

Women are more changeable, or rather, flexible, and more likely to adapt their behavior to circumstances than men. There are many factors that may explain some of the differences in the results, such as the greater risk aversion of women (Dwyer et al., 2002; Fletschner et al., 2010) or varied social distance between participants. However, we believe that the gender effect in irony distribution can also be attributable, at least to a certain extent, to a greater context sensitivity of women (Gilligan, 1982; Cadsby and Maynes, 1998). Gilligan (1982) claimed that women act more in terms of care orientation and cooperation toward other people while men act more in terms of abstract justice, rights, and obligations. Faced with a moral dilemma, that is, a choice to make, an individual with a care perspective will consider it in a contextualized fashion that takes into account how the individual is related to others who are involved in the dilemma. As a consequence, women may display different behavior in different contexts as a function of the contextualized features involved while men will tend to display behavior that is less context-sensitive and more rule-based. Another way of interpreting gender differences in use of irony can be in line with the reinterpretation of Cadsby and Maynes' (1998) view that women are more likely to follow conditional, as opposed to unconditional, rules—ones that depend on the specific context at hand.

The explanation of these differences can also lie in the theory of sexual selection. Females like predictability in their mates as it allows them to make good long-term decisions, and to deal with changing circumstances if they know their male is consistent (Schuett et al., 2010).

Could it be that even subtleties in our linguistic behavior reflect the true nature of our species? The study led by Schuett et al. (2010) shows that in most species, males show more consistent, predictable behaviors, particularly in relation to parental care, aggression, and risk-taking. Also, men are more inclined to engage in high-risk activities (Howland et al., 1996; Byrnes et al., 1999), of which irony can be an example on a symbolic level, like Colston and Lee (2004) suggest.

The results of Experiment 1 have also presented evidence supporting the activation of the mechanism of linguistic intergroup bias (Maass et al., 1989; Wigboldus and Douglas, 2007) in non-literal communication. The experiment shows that toward one's own group, that is, toward other women, irony was not frequently used in the context of ironic BbP, (covert criticism). Literal responses were significantly more present in responses to irony coming from an ingroup member in the negative pretask priming group. However, irony was kept in responses to ironic criticism expressed toward outgroup members, where literal comments were significantly less frequent.

Females used irony more often toward males (outgroup members) than to females (ingroup members), but only in the BbP context. However, they used more irony to ingroup members in the PbB (covert complement) than in the BbP (covert criticism) context and this trend was reversed when they used irony toward outgroup members, that is, males.

Male participants in the PbB (covert complement) context were more ironic toward females than males and used more literal responses when addressing the ingroup members, that is, other males.

These results also corroborate the results of the study by Milanowicz and Bokus (2013), which revealed that a simple change of the interactional setting, that is, of who is speaking to whom shifts the perception of the interlocutor and altogether modifies the process and the result of moral reasoning and communication.

Anxiety and Responses to Irony

The moderating effect of state anxiety on verbal reactions to ironic comments reflects the fact that irony is not a general quality of a person, but rather a state. Irony is unique depending on the context, and we do not respond to it by calling upon a database of jokes. Irony is being made again and anew, in the "here and now." Irony is more of a property of communication than of individual subjects. On the other hand, some people think of themselves as having more ironic tendencies than others. Our research draws attention to the fact that the amount of irony can be modified by conditioning. However, we do not rule out the role of other individual differences in the perception of irony and production of ironic remarks.

The perceived degree of lightheartedness or malice in ironic comments was previously linked to gender differences. Milanowicz (2013) showed that females displayed a more negative attitude to ironic comments than men, and it was hypothesized as being linked to different anxiety levels. To this end, IPT results were correlated with scores on the STAI. Anxiety proved to be a moderate predictor of irony comprehension and the willingness to use irony.

Is Irony a Funny Thing?

An unquestionable asset of this study is that it compares irony across different communicative situations within one modality: as if a real situation presented in a classic, written form. Unlike in Hancock's (2004) research paradigm, modality is kept constant and the study participants, that is, ironic speakers-to-be, can produce irony with the same set of linguistic tools. In lieu of intonation, facial expression, or gestures, they made good use of typographic cues and rhetorical figures,

We explored perception and use of irony by its direct "targets" or "recipients" (Experiment 1) as well as by its "witnesses" (Experiment 2). It is quite surprising that the self-attribution conditioning impacted male and female perception of malice and humor in a reversed way. It is also interesting that the differences are shown only in the assessment of male-male ironic setting. This might be the result of the activation of different mental representations associated with the social categories of men and women. For example, the label "woman" activates a different stereotype than the label "man." It also seems, based on these research results that irony acts as a filter, and its regulatory mechanism works different for men and different for women. It may relate to the fact that stereotypically, men have high power/status but low acceptance for expressing emotions. Women have high acceptance for expressing (negative) emotions but lower status, so irony works for men as a euphemism (for what cannot be said openly) and as loaded language for women. If ratings and evaluations of the same situation vary as to the context, mood, and perception of its evaluator, the above demonstrates that stereotypic expectancies are flexible and can be overridden under certain conditions.

Irony is omnipresent. Just take a look into private and public space to see how almost every category of contemporary reality—relationships, advertising, politics, television, or social networks—use and misuse this concept. Understanding irony (which is believed to be accessible only to humans) is maybe the closest experience we can have to mind reading. AI researchers are venturing into cracking irony, IT specialist are taking interest in linguistic profiling and HR managers are still probably confused as to whether they were funny or just mean. We are used to taking a binary approach to many things we try to understand, and we take that approach to irony as well. Something is ironic, that is, funny or something is not ironic, that is, not funny. However, we are quite certain that this binary character will soon be transgressed, as we will need an updated approach to understanding the trans-gender and trans-genetic world of the future, where the binary system of two definite states (0 or 1) will be just an illusion.

CONCLUSIONS

Irony is not a trifling matter. It is a social tool, which can make or break when it comes to having a mutual understanding and building rapport. We assumed that a higher level of anxiety in adults could account for the misperception of ambiguous comments as threatening and provoke tuning out of rather than tuning into an ironic exchange of comments. This hypothesis was not confirmed.

We showed, however, that linguistic bias is present in the use of irony, and we proposed that it might result from the gender effect and essentialist beliefs about social categories.

Taking into account individual preferences, variable across individuals while stable within individuals, can allow for more refined theoretical models of verbal irony as well as better predictions of communicative choices. More research and more data might lead to further discovery of patterns and mechanisms that impact the quality of interactions in the real-life context.

Our present study is a step forward to clarify the origins of individual differences in the perception of irony, its social functions and both, bright and dark sides of humor.

We plan in the future to explore the capacity to distinguish between the positive and negative functions of irony under the influence of gelotophobic fear, that is, the fear of laughter identified with ridicule (Ruch and Proyer, 2008a,b). We assume that gelotophobes will perceive positive irony as negative (as targeting their appearance, undermining their competence or achievements, etc.), more often than individuals with no fear of being laughed at.

LIMITATIONS

Emerging adults (i.e., adults between 18 and 25 years of age) constitute the majority of our research subjects—with mean age of 23.92 years for the total sample they form an in-between group, having completed adolescence but not yet entered adulthood. This is a stage of cognitive transition and recognizing different perspectives (Arnett, 2006).

We tried to control for the same advanced level in cognition by having a homogenous group of people with a university-level education (completed or in progress). We analyzed data collected from 238 subjects divided into six different experimental groups.

The sample size of the experimental groups may pose risk of Type II errors.

All our subjects were white, European, native Polish speakers, living in Warsaw, (the capital city—an economically developed area) Poland (almost 90% Roman Catholic), which makes us think that a more diversified ethnic background would call for cross-cultural research. For example, Irish/British citizens consider themselves highly ironic. Maybe it has something to do with the puritan Victorian society, where one just could not speak up about certain (embarrassing) things. How much has irony to do with taboo or religion and social beliefs?

And lastly, the very wording of the ironic comment (type of comment) can also provoke a given type of response on its own and act as a moderator, which is uncontrolled for.

AUTHOR CONTRIBUTIONS

The research reported in this article is part of AM's doctoral dissertation at the Faculty of Psychology of the University of Warsaw, Poland, supervised by the third author. AM: Data collection, data analysis, drafting manuscript. AT: Data analysis. BB: data collection, data analysis, revision of manuscript.

FUNDING

The publication was supported by the Faculty of Psychology, University of Warsaw.

ACKNOWLEDGMENTS

The authors thank James Beilby for his suggestions and comments on an earlier version of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02215/full#supplementary-material

REFERENCES

Anolli, L., Ciceri, R., and Infantino, M. G. (2002). From "blame by praise "to" praise by blame": analysis of vocal patterns in ironic communication. *Int. J. Psychol.* 37, 266–276. doi: 10.1080/00207590244000106

Arnett, J. (2006). *Emerging Adults in America, Coming of Age in the 21st Century*. Washington DC: American Psychological Association.

Attardo, S. (2002). "Humor and irony in interaction: from mode adoption to failure of detection," in *Say Not to Say: New Perspectives on Miscommunication*, eds L. Anolli, R. Ciceri, and G. Riva (Amsterdam: IOS Press), 159–179.

Bargh, J. A., Chaiken, S., Govender, R., and Pratto, F. (1992). The generality of the automatic attitude activation effect. *J. Pers. Soc. Psychol.* 62, 893–912. doi: 10.1037/0022-3514.62.6.893

Beukeboom, C. J. (2014). "Mechanisms of linguistic bias: How words reflect and maintain stereotypic expectancies," in *Social Cognition and Communication*, eds J. Laszlo, J. Forgas, and O. Vincze (New York, NY: Psychology Press), 313–330.

Bodenhausen, G. V., and Lichtenstein, M. (1987). Social stereotypes and information-processing strategies: the impact of task complexity. *J. Pers. Soc. Psychol.* 52, 871–880. doi: 10.1037/0022-3514.52.5.871

Brody, L. R., and Hall, J. A. (2008). "Gender and emotion in context," in *Handbook of Emotions, 3rd Edn*, eds M. Lewis, M. J. M. Haviland-Jones, and L. F. Barrett (New York, NY: The Guilford Press), 395–408.

Brody, L. R. (2000). "The socialization of gender differences in emotional expression: display rules, infant temperament, and differentiation," in *Gender and Emotion: Social Psychological Perspectives*, ed A. H. Fischer (Cambridge: Cambridge University Press), 24–47.

Brownell, H. H., Jacobs, J. R., Gardner, H., and Gianoulis, D. (1990). *Conditions for Sarcasm. [Unpublished paper from the Boston University Aphasia Research Center and Harvard Project Zero]*.

Bruntsch, R., Hofmann, J., and Ruch, W. (2016). Virgin soil in irony research: personality, humor, and the "sense of irony." *Transl. Issues Psychol. Sci.* 2, 25–34. doi: 10.1037/tps0000054

Burgers, C.h., Van Mulken, M., and Schellens, P. J. (2012). Verbal irony: differences in usage across written genres. *J. Lang. Soc. Psychol.* 31, 3. doi: 10.1177/0261927X12446596

Byrnes, J. P., Miller, D. C., and Schafer, W. D. (1999). Gender differences in risk taking: a meta-analysis. *Psychol. Bull.* 125, 367–383. doi: 10.1037/0033-2909.125.3.367

Cadsby, C. B., and Maynes, E. (1998). Gender and free riding in a threshold public goods game: experimental evidence. *J. Econ. Behav. Organ.* 34, 603–620. doi: 10.1016/S0167-2681(97)00010-3

Chaplin, T. M., and Aldao, A. (2013). Gender differences in emotion expression in children: a meta-analytic review. *Psychol. Bull.* 139, 735–765. doi: 10.1037/a0030737

Chłopicki, W., Radomska, A., Proyer, R. T., and Ruch, W. (2010). The assessment of the fear of being laughed at in Poland: translation and first evaluation of the Polish GELOPH<15>. *Polish Psychol. Bull.* 41, 172–181 doi: 10.2478/v10059-010-0022-2

Chorpita, B. F., Albano, A. M., and Barlow, D. H. (1996). Cognitive processing in children: relation to anxiety and family influences. *J. Clin. Child Adolesc. Psychol.* 25, 170–176. doi: 10.1207/s15374424jccp2502_5

Clark, H. H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

Clark, H., and Gerrig, R. (1984). On the pretense theory of irony. *J. Exp. Psychol. Gen.* 113, 121–126. doi: 10.1037/0096-3445.113.1.121

Collignon, O., Girard, S., Gosselin, F., Saint-Amour, D., Lepore, F., and Lassonde, M. (2010). Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia* 48, 220–225. doi: 10.1016/j.neuropsychologia.2009.09.007

Colston, H. L., and Lee, S. Y. (2004). Gender differences in verbal irony use. *Metaphor Symb.* 19, 289–306. doi: 10.1207/s15327868ms1904_3

Colston, H. L., and O'Brien, J. E. (2000). Contrast of kind versus contrast of magnitude: the pragmatic accomplishments of irony and hyperbole. *Discourse Process.* 30, 179–199. doi: 10.1207/S15326950DP3002_05

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Penguin Books.

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18. doi: 10.1037/0022-3514.56.1.5

Dews, S., Kaplan, J., and Winner, E. (1995). Why not say it directly? The social functions of irony. *Discourse Process.* 19, 347–367. doi: 10.1080/01638539509544922

Dunin, K. (2001). *Przereklamowana róznica*. Wysokie Obcasy. Available online at: http://www.wysokieobcasy.pl/wysokie-obcasy/1,96856,99076.html

Dwyer, P. D., Gilkeson, J. H., and List, J. A. (2002). Gender differences in revealed risk taking: evidence from mutual fund investors. *Econ. Lett.* 76, 151–158. doi: 10.1016/S0165-1765(02)00045-9

Dynel, M. (2014). Isn't it ironic? Defining the scope of humorous irony. *Humor* 27, 619–639. doi: 10.1515/humor-2014-0096

Filipova, E., and Astington, J. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Dev.* 79, 126–138. doi: 10.1111/j.1467-8624.2007.01115.x

Fletschner, D., Anderson, C. L., and Cullen, A. (2010). Are women as likely to take risks and compete? Behavioural findings from central Vietnam. *J. Dev. Stud.* 46, 1459–1479. doi: 10.1080/00220381003706510

Garmendia, J. (2010). Irony is critical. *Pragmat. Cogn.* 18, 397–421. doi: 10.1075/pc.18.2.07gar

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor Symb.* 15, 5–27. doi: 10.1080/10926488.2000.9678862

Gibbs, R. W., and Colston, H. L. (2007). *Irony in Language and thought: A Cognitive Science Reader*. New York, NY: Psychology Press.

Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.

Giora, R. (1995). On irony and negation. *Discourse Process.* 19, 239–264. doi: 10.1080/01638539509544916

Grice, H. P. (1975). "Logic and conversation," in *Syntax and Semantics Vol. 3: Speech Acts*, eds P. Cole and J. L. Morgan (New York, NY: Academic Press), 41–58.

Guimond, S., Branscombe, N. R., Brunot, S., Buunk, A. P., Chatard, A., and Yzerbyt, V. (2007). Culture, gender, and the self: variations and impact of social comparison processes. *J. Pers. Soc. Psychol.* 92, 1118–1134. doi: 10.1037/0022-3514.92.6.1118

Guimond, S., Chatard, A., Martinot, D., Crisp, R. J., and Redersdorff, S. (2006). Social comparison, self-stereotyping, and gender differences in self-construals. *J. Pers. Soc. Psychol.* 90, 221–242. doi: 10.1037/0022-3514.90.2.221

Hall, J. A. (1984). *Nonverbal Sex Differences: Communication Accuracy and Expressive Style*. Baltimore, MD: The Johns Hopkins University Press.

Hancock, J. T. (2004). Verbal irony use in face-to-face and computer-mediated conversations. *J. Lang. Soc. Psychol.* 23, 447–463. doi: 10.1177/0261927X0426958741.

Holmes, J. (1998). "Women's Talk: the question of sociolinguistci universals," in *Language and Gender: A Reader*, ed J. Coates (Oxford: Blackwell), 461–483.

Holtgraves, T. (1991). Interpreting questions and replies: effects of face-threat, question form, and gender. *Soc. Psychol. Q.* 54, 15–24. doi: 10.2307/2786785

Holtgraves, T. (1997). Yes, but positive politeness in conversation arguments. *J. Lang. Soc. Psychol.* 16, 222–239. doi: 10.1177/0261927X970162006

Holtgraves, T. (2005). Social psychology, cognitive psychology, and linguistic politeness. *J. Pol. R.* 1, 73–93. doi: 10.1515/jplr.2005.1.1.73

Howland, J., Hingson, R., Mangione, T. W., Bell, N., and Bak, S. M. P. H. (1996). Why are most drowning victims men? Sex differences in aquatic skills and behaviors. *Am. J. Public Health* 86, 93–96. doi: 10.2105/AJPH.86.1.93

Hyde, J. S. (2005). The gender similarities hypothesis. *Am. Psychol.* 60, 581–592. doi: 10.1037/0003-066X.60.6.581

Ivanko, S. L., Pexman, P. M., and Olineck, K. M. (2004). How sarcastic are you?: individual differences and verbal irony. *J. Lang. Soc. Psychol.* 23, 244–271. doi: 10.1177/0261927X04266809

Jorgensen, J. (1996). The functions of sarcastic irony in speech. *J. Pragmat.* 26, 613–634. doi: 10.1016/0378-2166(95)00067-4

Kałowski, P. (2017). *The Effect of Gender on Irony Understanding*. Master's thesis, University of Warsaw, Warsaw.

Katz, A., Piasecka, I., and Toplak, M. (2001). "Comprehending the sarcastic comments of males and females," in *Paper Presented at the 42nd Annual Meeting of the Psychonomic Society* (Orlando, FL).

Kotthoff, H. (2003). Responding to irony in different contexts: on cognition in conversation. *J. Pragmat.* 35, 1387–1411. doi: 10.1016/S0378-2166(02)00182-0

Kreuz, R., and Glucksberg, S. (1989). How to be sarcastic: the echoic reminder theory of irony. *J. Exp. Psychol. Gen.* 118, 374–386. doi: 10.1037/0096-3445.118.4.374

Kreuz, R. J., Long, D. L., and Church, M. B. (1991), On being ironic: pragmatic and mnemonic implications. *Metaphor Symb. Act.* 6, 149–162. doi: 10.1207/s15327868ms0603_1

Kring, A. M., and Gordon, A. H. (1998). Sex differences in emotion: expression, experience, and physiology. *J. Pers. Soc. Psychol.* 74, 686–703. doi: 10.1037/0022-3514.74.3.686

Kumon-Nakamura, S., Glucksberg, S., and Brown, M. (1995). How about another piece of pie: the allusional pretense theory of discourse irony. *J. Exp. Psychol. Gen.* 124, 3–21. doi: 10.1037/0096-3445.124.1.3

Lampert, M. D. (1996). "Gender differences in conversational humor," in *Social Interaction, Social Context, and Language*, eds D. I. Slobin, J. Gerhardt, A. Kyratzis, and J. Guo (Hillsdale, NJ: Lawrence Erlbaum), 579–596.

Leech, G. N. (1983). *Principles of Pragmatics*. New York, NY: Longman.

Lewin, K. (1935). *A Dynamic Theory of Personality*. New York, NY: McGraw Hill.

Lutz, C. (1996). "Engendered emotion: gender, power, and the rhetoric of emotional control in American discourse," in *The Emotions: Social, Cultural and Biological Dimensions*, eds R. Harre, and W. G. Parrot (Thousand Oaks, CA: Sage), 152–170.

Maass, A. (1999). "Linguistic intergroup bias: stereotype perpetuation through language," in *Advances in Experimental Social Psychology Vol. 31*, ed M. P. Zanna (San Diego, CA: Academic Press), 79–121.

Maass, A., Salvi, D., Acuri, L., and Semin, G. R. (1989). Language use in intergroup contexts: the linguistic intergroup bias. *J. Pers. Soc. Psychol.* 57, 981–993. doi: 10.1037/0022-3514.57.6.981

Macrae, C. N., Stangor, C., and Milne, A.B. (1994). Activating social stereotypes: a functional analysis. *J. Exp. Soc. Psychol.* 30, 370–389. doi: 10.1006/jesp.1994.1018

Matthews, J. K., Hancock, J. T., and Dunham, P. J. (2006). The roles of politeness and humor in the asymmetry of affect in verbal irony. *Discourse Process.* 41, 3–24. doi: 10.1207/s15326950dp4101_2

Mewhort-Buist, T. A., and Nilsen, E. S. (2013). What are you really saying? Associations between shyness and verbal irony comprehension. *Infant Child Dev.* 22, 180–197. doi: 10.1002/icd.1769

Milanowicz, A., and Kałowski, P. (2016). Zing Zing Bang Bang: how do you know what she really meant. Gender bias in response to irony: the role of who is speaking to whom. *Psychol. Lang. Commun.* 20, 213–234. doi: 10.1515/plc-2016-0014

Milanowicz, A. (2016). *Irony Processing Task.* Unpublished questionnaire from doctoral thesis, University of Warsaw.

Milanowicz, A., and Bokus, B. (2013). Gender and moral judgments: the role of who is speaking to whom. *J. Gender Stud.* 22, 423–443. doi: 10.1080/09589236.2012.719314

Milanowicz, A. (2013). Irony as a means of perception through communication channels. Emotions, attitude and IQ related to irony across gender. *Psychol. Lang. Commun.* 17, 115–132. doi: 10.2478/plc-2013-0008

Partington, A. (2007). Irony and reversal of evaluation. *J. Pragmat.* 39, 1547–1569. doi: 10.1016/j.pragma.2007.04.009

Pexman, P. M., and Olineck, K. M. (2002). Does sarcasm always sting? Investigating the impact of ironic insults and ironic compliments. *Discourse Process.* 33, 199–217. doi: 10.1207/S15326950DP3303_1

Poggi, I., Cavicchio, F., and Caldognetto, E., M (2007). Irony in a judicial debate: analyzing the subtleties of irony while testing the subtleties of an annotation scheme. *Lang. Res. Eval.* 41, 215–232. doi: 10.1007/s10579-007-9059-z

Poggi, I., and D'Errico, F. (2010). Dominance signals in debates. *Hum. Behav. Underst.* 6219, 163–174. doi: 10.1007/978-3-642-14715-9_16

Recchia, H. E., Howe, N., Ross, H. S., and Alexander, S. (2010). Children's understanding and production of verbal irony in family conversations. *Br. J. Dev. Psychol.* 28, 255–274. doi: 10.1348/026151008X401903

Remes, O., Brayne, C., Linde, R., and Lafortune, L. (2016). A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain Behav.* 6:e00497. doi: 10.1002/brb3.497

Roberts, R. M., and Kreuz, R. J. (1994). Why do people use figurative language? *Psychol. Sci.* 5, 159–163. doi: 10.1111/j.1467-9280.1994.tb00653.x

Ruch, W., and Proyer, R. T. (2008a). The fear of being laughed at: individual and group differences in Gelotophobia. *Humor Int. J. Humor Res.* 21, 47–67. doi: 10.1515/HUMOR.2008.002

Ruch, W., and Proyer, R. T. (2008b). Who is gelotophobic? Assessment criteria for the fear of being laughed at. *Swiss J. Psychol.* 67, 19–27. doi: 10.1024/1421-0185.67.1.19

Sani, F., Bennett, M., Mullaly, S., and MacPherson, J. (2003). On the assumption of fixity in children's stereotypes: a reappraisal. *Br. J. Dev. Psychol.* 21, 113–124. doi: 10.1348/026151003321164654

Schuett, W., Tregenza, T., and Dall, S. R. (2010). Sexual selection and animal personality. *Biol. Rev.* 85, 217–246. doi: 10.1111/j.1469-185X.2009.00101.x

Schwoebel, J., Dews, S., Winner, E., and Srinivas, K. (2000). Obligatory processing of the literal meaning of ironic utterances: further evidence. *Metaphor Symb*. 15, 47–61. doi: 10.1080/10926488.2000.9678864

Sperber, D., and Wilson, D. (1981). "Irony and the use-mention distinction," in *Radical Pragmatics*, ed P. Cole (New York, NY: Academic Press), 295–318.

Sperber, D., and Wilson, D. (1984). "Pragmatics: an overview," in *From the Lingusistics to the Social Context,* ed S. Georges (Bologna: Cooperativa Libraria Universitaria Editrice), 21–41.

Spielberger, C. D., J., Strelau, M., Tysarczyk, and Wrześniewski, K. (1987). *STAI - Inwentarz Stanu i Cechy Leku STAI*. Warsaw: Pracownia Testów Psychologicznych.

Taghavi, R., Moradi, A. R., Neshat-Doost, H. T., Yule, W., and Dalgleish, T. (2000). Interpretation of ambiguous emotional information in clinically anxious children and adolescents. *Cogn. Emot.* 14, 809–822. doi: 10.1080/02699930050156645

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., and Wetherell, M. S. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. Oxford: Blackwell.

Wigboldus, D. H. J., and Douglas, K. (2007). "Language, stereotypes, and intergroup relations," in *Social Communication*, ed K. Fiedler (New York, NY: Psychology Press), 79–106.

Wigboldus, D. H. J., Spears, R., and Semin, G. R. (2005). When do we communicate stereotypes? Influence of the social context on the linguistic expectancy bias. *Group Process. Intergr. Relat.* 8, 215–230. doi: 10.1177/1368430205053939

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Milanowicz, Tarnowski and Bokus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 February 2018
doi: 10.3389/fpsyg.2018.00092

Psychometric Comparisons of Benevolent and Corrective Humor across 22 Countries: The Virtue Gap in Humor Goes International

Sonja Heintz[1], Willibald Ruch[1], Tracey Platt[2], Dandan Pang[1], Hugo Carretero-Dios[3], Alberto Dionigi[4], Catalina Argüello Gutiérrez[3], Ingrid Brdar[5], Dorota Brzozowska[6], Hsueh-Chih Chen[7], Władysław Chłopicki[8], Matthew Collins[9], Róbert Ďurka[10], Najwa Y. El Yahfoufi[11], Angélica Quiroga-Garza[12], Robert B. Isler[13], Andrés Mendiburo-Seguel[14], TamilSelvan Ramis[15], Betül Saglam[16], Olga V. Shcherbakova[17], Kamlesh Singh[18], Ieva Stokenberga[19], Peter S. O. Wong[20] and Jorge Torres-Marín[21]*

[1] Department of Psychology, Personality and Assessment, University of Zurich, Zurich, Switzerland, [2] Faculty of Education, Health and Wellbeing, Institute of Psychology, University of Wolverhampton, Wolverhampton, United Kingdom, [3] Department of Methodology of Behavioral Sciences, Faculty of Psychology, Centro de Investigación Mente, Cerebro, y Comportamiento, University of Granada, Granada, Spain, [4] Federazione Nazionale Clown Dottori (FNC), Cesena, Italy, [5] Department of Psychology, Faculty of Humanities and Social Sciences, University of Rijeka, Rijeka, Croatia, [6] Institute of English, Faculty of Philology, University of Opole, Opole, Poland, [7] College of Education, National Taiwan Normal University, Taipei, Taiwan, [8] Department of English Studies, Jagiellonian University, Kraków, Poland, [9] School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, United Kingdom, [10] Department of Psychology, Faculty of Arts and Letters, Catholic University in Ružomberok, Ružomberok, Slovakia, [11] Department of Psychology, Faculty of Letters and Human Sciences, Lebanese University, Beirut, Lebanon, [12] Departamento Académico de Psicología, Universidad de Monterrey, San Pedro Garza García, Mexico, [13] School of Psychology, University of Waikato, Hamilton, New Zealand, [14] Facultad de Educación, Universidad Andrés Bello, Santiago, Chile, [15] Department of Psychology, HELP University, Kuala Lumpur, Malaysia, [16] Psychology Department, Üsküdar University, Istanbul, Turkey, [17] Faculty of Psychology, Saint Petersburg State University, Saint Petersburg, Russia, [18] Department of Humanities and Social Sciences, Indian Institute of Technology Delhi, New Delhi, India, [19] Department of Psychology, Faculty of Education, Psychology and Art, University of Latvia, Riga, Latvia, [20] Centre for Fundamental and Liberal Education, Universiti Malaysia Terengganu, Kuala Nerus, Malaysia, [21] Department of Experimental Psychology, Faculty of Psychology, Centro de Investigación Mente, Cerebro y Comportamiento, University of Granada, Granada, Spain*

OPEN ACCESS

Edited by:
Monika Fleischhauer, Medizinische Hochschule Brandenburg Theodor Fontane, Germany

Reviewed by:
Feng Jiang, Central University of Finance and Economics, China
Xiaodong Yue, City University of Hong Kong, Hong Kong

***Correspondence:**
Sonja Heintz
s.heintz@psychologie.uzh.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *01 October 2017*
Accepted: *22 January 2018*
Published: *09 February 2018*

Citation:
Heintz S, Ruch W, Platt T, Pang D, Carretero-Dios H, Dionigi A, Argüello Gutiérrez C, Brdar I, Brzozowska D, Chen H-C, Chłopicki W, Collins M, Ďurka R, Yahfoufi NYE, Quiroga-Garza A, Isler RB, Mendiburo-Seguel A, Ramis T, Saglam B, Shcherbakova OV, Singh K, Stokenberga I, Wong PSO and Torres-Marín J (2018) Psychometric Comparisons of Benevolent and Corrective Humor across 22 Countries: The Virtue Gap in Humor Goes International. Front. Psychol. 9:92. doi: 10.3389/fpsyg.2018.00092

Recently, two forms of virtue-related humor, benevolent and corrective, have been introduced. Benevolent humor treats human weaknesses and wrongdoings benevolently, while corrective humor aims at correcting and bettering them. Twelve marker items for benevolent and corrective humor (the BenCor) were developed, and it was demonstrated that they fill the gap between humor as temperament and virtue. The present study investigates responses to the BenCor from 25 samples in 22 countries (overall N = 7,226). The psychometric properties of the BenCor were found to be sufficient in most of the samples, including internal consistency, unidimensionality, and factorial validity. Importantly, benevolent and corrective humor were clearly established as two positively related, yet distinct dimensions of virtue-related humor. Metric measurement invariance was supported across the 25 samples, and scalar invariance was supported across six age groups (from 18 to 50+ years) and across gender. Comparisons of samples within and between four countries (Malaysia, Switzerland, Turkey, and the UK) showed that the item profiles were more similar within than between countries, though some evidence for

regional differences was also found. This study thus supported, for the first time, the suitability of the 12 marker items of benevolent and corrective humor in different countries, enabling a cumulative cross-cultural research and eventually applications of humor aiming at the good.

Keywords: humor, virtue, cross-cultural comparisons, measurement invariance, positive psychology

INTRODUCTION

Humor has been extensively studied in many areas of psychology, ranging from basic to applied research (for an overview, see Martin, 2007). In the area of individual differences in humor, different concepts of humor styles have been proposed, either as individual differences in humor behaviors (Craik et al., 1996) or in the functions of humor (Martin et al., 2003). A more recent approach emphasizes eight different comic styles that were derived from an interdisciplinary approach (Ruch et al., 2018a), namely fun, (benevolent) humor, nonsense, wit, irony, satire/corrective humor, sarcasm, and cynicism. The present investigation focuses on two comic styles, benevolent and corrective humor, which are historically, conceptually, and empirically related to virtue. The aim is to compare the 12 marker items of benevolent and corrective humor (created by Ruch, 2012) across different countries to investigate their psychometric properties across countries, age groups, and gender.

According to Ruch and Heintz (2016), benevolent and corrective humor are both morally valued and aim at doing good. Benevolent humor includes an accepting attitude toward the world and toward human weaknesses, and it treats them benevolently. It also includes being aware of one's surroundings and of everyday occurrences, which can then be reframed and commented on in a benevolent and humorous way. Corrective humor criticizes wrongdoings of both individuals and institutions, and it mocks them in order to improve them. Thus, it adds a moral goal to the criticism, which distinguishes corrective humor from pure mockery or aggressive forms of humor that lack this component. The connection of benevolent and corrective humor with morality and values can be traced back to their humanistic and philosophical roots, originating in England in the nineteenth century (for details, see Ruch and Heintz, 2016).

There are elements that benevolent and corrective humor share as well as elements where they differ. Both styles involve spotting incongruities in everyday life that are not inherently humorous, rather than processing and appreciating canned humor. Furthermore, these incongruities are processed playfully (not seriously) and they are treated humorously. Thus, in both styles the protagonist is attentive to what happens in his/her surroundings and realizes that deviations from expectations occur. This contributes to a large positive correlation between the two styles. However, in benevolent humor, the wrongdoing is not considered to be very important; for example, Nicolson (1946) suggested that humor observes human frailty indulgently, without bothering to correct it. In corrective humor, however, the difference between the real and the ideal is noticed, and funny comments are made to mock and to press someone to do the right thing. The two styles are opposite in this respect, thus reducing their overall positive correlation.

In line with these conceptualizations, the initial study (Ruch and Heintz, 2016) supported positive relationships of benevolent and corrective humor with several character strengths based on the VIA (Values in Action) classification of strengths and virtues (Peterson and Seligman, 2004). Specifically, benevolent humor uniquely related to character strengths assigned to the virtues of temperance (e.g., forgiveness), wisdom and knowledge (e.g., love of learning), transcendence (e.g., hope, humor), humanity (e.g., social intelligence), and justice (e.g., fairness). Of note, these relationships were robust when controlling for the sense of humor (as conceptualized by McGhee, 2010). By contrast, corrective humor was mostly uncorrelated with the strengths, except for positive correlations with creativity, bravery, and humor. Once mockery was controlled for, however, positive relationships emerged also with fairness and love of learning. This supports the notion that benevolent and corrective humor fill a virtue gap in humor by showing unique relationships to character strengths that serve to fulfill different virtues (such as humanity, justice, and wisdom/knowledge).

Investigating benevolent and corrective humor across several countries and languages is relevant for several reasons. First, despite the historical relevance of these two virtue-related humor styles, they have been neglected in psychological research. Establishing that the two styles can be found and distinguished across several countries would further support the relevance of the virtue gap in humor. Second, supporting the psychometric properties of the 12 marker items (or a subset thereof) would pave the way for international investigations on the nomological network of benevolent and corrective humor, as well as their predictors and virtue-relevant outcomes. Third, large-scale cross-cultural studies in the area of humor and virtues have been scarce (for exceptions, see Park et al., 2006; Proyer et al., 2009; McGrath, 2015, 2016), thus making the present study a valuable contribution to cross-cultural humor research and positive psychology more generally. Additionally, the large sample also allows comparing differences in benevolent and corrective humor across age groups and gender as two central demographic characteristics.

The present study investigates the psychometric properties of a set of 12 marker items for benevolent and corrective humor (the *BenCor*) within 25 samples from 22 countries. This includes descriptive statistics, reliability, measurement invariance, factorial validity, construct validity, profile similarity across the 12 marker items, as well as age and gender differences. Measurement invariance includes testing metric invariance (i.e., equal item loadings on the latent factor) and scalar invariance (i.e., equal item intercepts on the latent factor). Metric invariance is needed to compare the factors and slopes across the samples, and scalar invariance is needed to compare mean scores across the samples (see Chen, 2008). This allows evaluating the

suitability of the BenCor across samples from different countries, across different age groups, and across gender.

MATERIALS AND METHODS

Samples

Inclusion criteria for participants were (a) an age of at least 18 years, (b) a reasonable command of the language in which the survey was conducted, and (c) the completion of all BenCor marker items. Participants who selected the same answer option for each item (e.g., answered "strongly agree" to all items) were excluded. **Table 1** gives an overview of the resulting 25 BenCor samples in the 22 countries.

As shown in **Table 1**, sample sizes ranged from 173 (Costa Rica) to 533 (Switzerland, general community sample), with 7,226 participants overall. Gender was mostly balanced across samples ($M = 40.2\%$ males), with the percentages ranging from 29.0% males (Slovakia) to 59.7% males (Northern Ireland). The average age of the samples ranged from 20.10 years (China) to 39.15 years (Austria), with an overall mean of 28.73 years. The median age was lowest for China, Taiwan, and Northern Ireland ($Mdn = 20.00$ years), while it was highest for Austria ($Mdn = 40.00$ years). Thus, most of the samples comprised young to middle-aged adults. This is also reflected in the sample type, which were primarily students in 11 samples, primarily adults from the community in 6 samples, and both students and adults from the community in 8 samples. Finally, data collection was conducted online in 14 samples, offline in 8 samples, and both online and offline in 3 samples.

Measures

The BenCor (Ruch, 2012) assesses benevolent and corrective humor with 6 marker items each (see **Table 2**). The marker items were derived from descriptions of humor and satire (corresponding to benevolent and corrective humor, respectively) based on literary and linguistic analyses (Schmidt-Hidding, 1963). These literary concepts were transformed into psychological traits, capturing individual differences in the propensity to engage in benevolent and corrective humor (for details, see Ruch et al., 2018a). A first psychometric analysis of the 12 marker items in a German-speaking sample (Ruch and Heintz, 2016) supported (a) the two-factor structure (based on a principal component analysis), (b) the assignment of each item to the corresponding factor, (c) internal consistencies (Cronbach's alpha 0.82 for benevolent and 0.84 for corrective humor), and (d) the criterion validity of the two sets of marker items in terms of character strengths. Recent studies further supported the construct validity (self-other agreement) and the criterion validity (in terms of personality, character strengths, and well-being) of the 12 marker items (Ruch et al., 2018a,b). The BenCor employs a seven-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Additionally, demographic information was collected from the participants, such as gender and age, and also further information such as nationality, language skills, and education. In some samples, additional measures were employed that are not relevant to the present study.

Procedure

Each non-native English speaking co-author received a standardized package for the translation of the BenCor and the data collection. This included the English version of the 12 marker items (in some cases additional language versions were provided upon request), questionnaire instructions, descriptions of benevolent and corrective humor, the scoring key, the paper by Ruch and Heintz (2016), a description of the standardized translation/back-translation procedure (i.e., a translation to the local language and an independent back-translation into English), and a paper on guidelines for test translations (Van de Vijver and Hambleton, 1996). All item-translating co-authors had the opportunity to discuss their translations and the item contents with the first and second author to ensure that the items preserved their meaning in the translation. If a translation to the local language already existed, the co-authors were asked to check the applicability of the translation and to suggest adaptations if necessary. For example, the Spanish version (translated in Spain) was slightly adapted to fit to the Chilean and Costa Rican form of Spanish.

The online samples were collected by sending a link to the survey, which were hosted on different platforms (such as SurveyMonkey, Unipark, or Qualtrix). The offline samples were collected by asking participants (e.g., in libraries or classrooms) to complete the questionnaire in a paper-pencil version. These data were then manually entered into standardized data sheet (Excel or SPSS). Participants were recruited via different means, such as mailing lists, personal contacts, social media, the university campus, and thus comprise convenience samples. To analyze the data, they were either directly downloaded from online platforms or they were sent in the standardized data sheet to the first author. The 25 samples were collected in accordance with the local ethical guidelines, and participants provided either online or written informed consent in accordance with the Declaration of Helsinki.

After the data collection and initial data analyses, all co-authors completed a collaborator's form to provide details on the translated instrument, the sample description, the data collection procedure, and the interpretation of the data. For example, they reported which type of sample was investigated, the language skills and nationalities of the sample, how participants were approached, which mode of data collection was employed (i.e., online or offline), and whether any unexpected events occurred while collecting the data.

Analyses

Reliability and Validity

The internal consistencies of the samples are indicated by Cronbach's alpha. The factorial validity of the BenCor was tested in principal components analyses (PCA) with oblimin rotation and in confirmatory factor analyses (CFA). Based on the pattern matrix (factor loadings) of the PCA, Tucker's phi as an index of factor congruence was computed across the 12 items, separately for the benevolent and the corrective humor factor. According to Lorenzo-Seva and Ten Berge (2006), Tucker's phi coefficients $\geq$0.95 indicate equality and coefficients from 0.85 to 0.94 indicate a fair similarity of the factors. The CFA was computed with the *lavaan* package (Rosseel, 2012) in *R*

TABLE 1 | Overview of the 25 BenCor Samples in the 22 Countries.

Countries	Language	*N*	*% Males*	Age (*M*)	Age (*Mdn.*)	Primary sample type	Data collection
Austria	German	350	32.6	39.15	40.00	Community	Online
Chile	Spanish	227	48.0	29.29	29.00	Community & students	Online
China (Guangzhou, Beijing)	Chinese	243	38.7	20.10	20.00	Community & students	Online
Costa Rica	Spanish	173	47.4	33.37	30.00	Community	Offline
Croatia	Croatian	350	54.9	21.27	21.00	Students	Offline
Germany	German	282	36.2	34.28	29.50	Community & students	Online
India	Hindi	198	49.5	26.36	23.00	Community	Offline
Italy	Italian	193	44.0	36.77	35.00	Community	Online
Latvia	Latvian	413	33.4	33.23	30.00	Community & students	Online
Lebanon	Arabic	260	37.7	25.26	21.00	Students	Offline
Malaysia	Malay	239	42.3	24.95	24.00	Students	Online
Malaysia (Terengganu)	Malay	199	50.3	24.45	21.00	Students	Offline
Mexico	Spanish	198	49.0	20.88	21.00	Students	Offline
New Zealand	English	221	41.6	34.21	31.00	Community	Online
Poland	Polish	458	30.0	33.97	32.00	Community & students	Online
Russia	Russian	201	49.8	30.24	25.00	Community & students	Online & offline
Slovakia	Slovak	400	29.0	25.79	22.00	Students	Online & offline
Spain	Spanish	209	46.4	22.55	21.00	Students	Offline
Switzerland (students)	German	313	32.6	24.95	24.00	Students	Online
Switzerland (general)	German	533	37.9	39.09	34.00	Community	Online
Taiwan	Chinese (trad.)	440	48.4	21.00	20.00	Students	Offline
Turkey (non-graduates)	Turkish	336	33.3	25.55	22.00	Community & students	Online
Turkey (university graduates)	Turkish	320	34.4	36.73	36.00	Community & students	Online
UK (mostly England)	English	269	35.3	31.19	25.00	Students	Online & offline
UK (Northern Ireland)	English	201	59.7	23.70	20.00	Students	Online

TABLE 2 | Overview of the 12 BenCor Items Marking Benevolent (Ben) and Corrective (Cor) Humor.

No.	Humor	Items
1	Ben	I am a realistic observer of human weaknesses, and my good-natured humor treats them benevolently.
3	Ben	When my humor is aimed at human weaknesses, I include both myself and others.
5	Ben	On a large and small scale, the world is not perfect, but with a humorous outlook on the world I can amuse myself at the adversities of life.
7	Ben	I accept the imperfection of human beings and my everyday life often gives me the opportunity to smile benevolently about it.
9	Ben	Humor is suitable for arousing understanding and sympathy for imperfections and the human condition.
11	Ben	Even when facing unpleasant events I can keep my distance and discover something amusing or funny in it.
2	Cor	I have a critical attitude toward arrogant and unfair people and my mockery serves to establish equality and justice.
4	Cor	I parody people's bad habits to fight the bad and foolish behavior.
6	Cor	When fellow humans or institutions demonstrate their superiority unjustified, I use biting humor to belittle them.
8	Cor	I caricature my fellow humans' wrongdoings in a funny way to gently urge them to change.
10	Cor	I like to ridicule moral badness to induce or increase a critical attitude in other people.
12	Cor	If the circumstances are not as they actually should be, I poke fun at these moral transgressions or societal wrongdoings, hoping to improve them in the long term.

(R Development Core, 2015). The robust MLM estimator (with Satorra-Bentler corrections) was employed for all CFA analyses. The following fit indices were evaluated using the recommended cut-offs by Schermelleh-Engel et al. (2003): χ^2/df (good: $\leq$ 2, acceptable: $\leq$3), comparative fit index (CFI; good: $\geq$0.97, acceptable: $\geq$0.95), root mean square error of approximation (RMSEA; good: $\leq$0.05, acceptable: $\leq$0.08), and standardized root mean square residual (SRMR; good: $\leq$0.05, acceptable: $\leq$0.10). The one- and two-factor structure of the 12 BenCor marker items and the unidimensionality of benevolent and corrective humor (six marker items each) were investigated in CFAs. These analyses were conducted separately for each sample and across all samples.

Construct validity (discriminant validity) was assessed utilizing the average variance explained (AVE) calculation. According to Fornell and Larcker (1981), the AVE is computed by averaging the squared standardized loadings of each item on

the factor. Discriminant validity can be supported if the square root of the AVE of each factor is larger than the correlation between the factors (the Fornell-Larcker criterion). To avoid biases due to measurement error, the Fornell-Larcker criterion was evaluated in the CFAs only (separate for each sample and across the 25 samples).

Measurement Invariance

Measurement invariance was tested separately for benevolent and corrective humor using a multi-group CFA with the semTools package (semTools Contributors, 2015) in *R*. Metric invariance was tested by forcing all item loadings to be equal across groups. This model was then compared with the baseline model that allows a free estimation of the item loadings, comparing the difference in the CFI and the RMSEA. Changes of $\leq|0.01|$ in the CFI and changes of $\leq|0.015|$ in the RMSEA were used as cut-offs to indicate measurement invariance (based on the recommendations by Cheung and Rensvold, 1999; Chen, 2007). Similarly, scalar invariance was tested by forcing both the intercepts and the loadings to be equal across groups. In addition, partial measurement invariance at the item-level was investigated. A baseline model with free item loadings served as a comparison for models in which the item loadings (for metric invariance) and item intercepts (for scalar invariance) were constrained across the groups. This model was shown to be superior to a constrained-baseline model, in which each item is freed to test its differential functioning (see Stark et al., 2006). The CFI difference of $\leq|0.01|$ was used to evaluate the partial measurement invariance of single items. Metric measurement invariance was tested across the 25 samples, across gender (n = 2,906 males and n = 4,312 females), and across six age groups: 18–20 years (n = 1,624), 21–24 years (n = 1,981), 25–29 years (n = 1,081), 30–39 years (n = 1,225), 40–49 years (n = 704), and 50+ years (n = 580). Additionally, scalar invariance was tested for gender and age.

Cross-Sample Comparisons

Similarities in the 12 marker items between the 25 samples were analyzed in terms of (a) means, (b) corrected item-total correlations (CITC), (c) multidimensional scaling of item-profile similarities, and (d) profile correlations across the 12 items. For the multidimensional scaling, the item means were analyzed using the alternating least squares scaling (ALSCAL) algorithm and Euclidian distances. These analyses were conducted for all samples, with additional analyses focusing on the samples that shared a language (i.e., English, German, and Spanish) as well as samples from the same country (i.e., Malaysia, Switzerland, Turkey, and the UK).

RESULTS

Descriptive Statistics of Benevolent and Corrective Humor

Table 3 shows the descriptive statistics of the BenCor in the 25 samples.

As shown in **Table 3**, the means for benevolent humor ranged from 4.66 (Lebanon) to 5.44 (Spain), with a mean across samples of 5.16 (*slightly agree*). The means for corrective humor ranged from 3.51 (Lebanon) to 4.71 (India), with a mean of 4.18 (*neither agree nor disagree*). Additionally, every sample had numerically higher scores in benevolent than in corrective humor. The means of benevolent and corrective humor correlated positively with one another across the samples [$r_{(25)} = 0.67$, $p < 0.001$].

Regarding the variance in benevolent humor, the standard deviations ranged from 0.75 (New Zealand) to 1.17 (Costa Rica), with a mean of 0.86. For corrective humor, the variance was numerically larger and ranged from 0.93 (Croatia) to 1.46 (Costa Rica), with a mean of 1.12. Thus, both benevolent and corrective humor created sufficient variance within each sample, with a tendency for corrective humor to elicit more varied responses. Similar to the mean scores, the standard deviations of benevolent and corrective humor were strongly positively correlated [$r_{(25)} = 0.82$, $p < 0.001$].

Reliability

Next, the reliability of benevolent and corrective humor was investigated in each sample. As shown in **Table 3**, internal consistencies (Cronbach's alpha) of benevolent humor exceeded 0.60 in 21 of the 25 samples. Exceptions were India, Lebanon, Malaysia (Terengganu sample) and Turkey (graduate sample), in which internal consistencies ranged from 0.50 to 0.58. Across all samples, the median was 0.67. For corrective humor, all internal consistencies exceeded 0.60 (Mdn = 0.77). Thus, the internal consistencies were sufficient for corrective humor in all samples, and for benevolent humor in most samples.

Next, unidimensionality (or homogeneity) was tested in CFAs, separate for the six marker items of benevolent and corrective humor. **Table 4** shows the resulting fit indices for each of the two CFA models in the 25 samples.

As shown in **Table 4**, the fit indices were acceptable or good in 14 of the 25 samples for benevolent humor. In eight further samples, all fit indices indicated an acceptable fit, with the exception of the CFI. Due to the comparably large number of variables per factor (six), lower CFI values might be found even if the model is correctly specified (see Kenny and McCoach, 2003). Only in three samples (Chile, Taiwan, and the Turkey graduate sample), at least two fit indices were unacceptable. For corrective humor, 20 of the 25 samples showed acceptable or good fit indices, and two showed lower values only in the CFI (China and India). For Latvia, Lebanon, and the Turkey graduate sample, at least two fit indices were unacceptable for corrective humor. Overall, the unidimensionality of benevolent and corrective humor was supported for most samples.

Measurement Invariance across Samples, Age Groups, and Gender

Before comparing the factors, correlations, and mean scores, the measurement invariance of the BenCor was tested across samples, age, and gender. **Table 5** shows the fit indices of the baseline model (in which the item loadings were allowed to vary freely) with the metric invariance model (in which the item loadings were constrained to be equal across groups) and the scalar invariance model (in which the item loadings and

TABLE 3 | Psychometric characteristics and correlations with gender of the 25 BenCor samples in the 22 countries.

Countries	Benevolent humor					Corrective humor					r_{BenCor}	
	M	*SD*	α	φ	r_{gender}	*M*	*SD*	α	φ	r_{gender}	Scales	Factors
Austria	5.28	0.87	0.76	0.99	−0.07	4.20	1.14	0.83	0.99	−0.20***	0.40***	0.34
Chile	5.24	1.12	0.76	0.99	−0.09	4.56	1.36	0.82	0.98	−0.21**	0.37***	0.30
China (Guangzhou, Beijing)	5.11	0.82	0.74	0.93	0.11	4.19	0.94	0.73	0.95	−0.25***	0.33***	0.24
Costa Rica	5.01	1.17	0.76	0.95	0.01	4.48	1.46	0.85	0.95	−0.20**	0.61***	0.47
Croatia	5.26	0.82	0.65	0.95	0.05	4.50	0.93	0.69	0.96	−0.08	0.32***	0.24
Germany	5.04	0.87	0.74	0.98	−0.10	4.10	1.23	0.85	0.97	−0.21***	0.49***	0.39
India	5.33	0.76	0.51	0.77	0.10	4.71	1.06	0.70	0.86	−0.02	0.50***	0.25
Italy	5.38	0.79	0.66	0.94	0.04	4.50	1.13	0.80	0.95	−0.19**	0.34***	0.25
Latvia	5.36	0.89	0.77	0.92	−0.04	4.26	1.12	0.78	0.92	−0.27***	0.49***	0.34
Lebanon	4.66	0.80	0.56	0.95	0.02	3.51	1.05	0.66	0.94	−0.11	0.32***	0.26
Malaysia	5.12	0.85	0.63	0.93	−0.12	3.99	1.13	0.73	0.90	−0.32***	0.45***	0.32
Malaysia (Terengganu)	5.29	0.80	0.58	0.85	−0.08	4.31	1.05	0.69	0.86	−0.11	0.54***	0.27
Mexico	5.25	0.86	0.62	0.97	0.05	3.87	1.12	0.71	0.96	−0.21**	0.35***	0.29
New Zealand	5.40	0.75	0.62	0.99	0.04	4.26	1.08	0.79	0.99	−0.12	0.28***	0.24
Poland	5.22	0.87	0.72	0.95	0.00	4.27	1.14	0.76	0.98	−0.22***	0.34***	0.24
Russia	5.04	0.86	0.60	0.93	0.00	3.60	1.07	0.72	0.91	−0.21**	0.21**	0.15
Slovakia	5.05	0.84	0.67	0.97	−0.06	4.10	1.10	0.77	0.98	−0.21***	0.48***	0.37
Spain	5.44	0.81	0.65	0.97	−0.05	4.21	1.21	0.79	0.99	−0.19**	0.28***	0.23
Switzerland (students)	5.14	0.81	0.80	–	−0.12*	4.23	1.06	0.83	–	−0.28***	0.45***	0.43
Switzerland (general)	4.98	0.83	0.74	1.00	−0.10*	4.09	1.09	0.81	0.99	−0.27***	0.53***	0.45
Taiwan	5.07	0.85	0.72	0.97	−0.14**	4.12	1.09	0.80	0.95	−0.38***	0.37***	0.30
Turkey (non-graduates)	4.87	1.03	0.67	0.89	−0.04	3.89	1.23	0.72	0.88	−0.20***	0.54***	0.34
Turkey (graduates)	4.90	0.85	0.50	0.80	0.02	3.96	1.15	0.68	0.86	−0.20***	0.45***	0.22
UK (mostly England)	5.11	0.87	0.69	0.99	−0.11	4.19	1.11	0.78	0.94	−0.22***	0.41***	0.28
UK (Northern Ireland)	5.33	0.76	0.60	0.94	−0.05	4.41	1.07	0.75	0.92	−0.18*	0.37***	0.24

*α, Cronbach's alpha (internal consistency); φ, Tucker's phi (factor congruence to the Swiss student sample based on the pattern matrix in the principal component analysis with oblimin rotation); gender coded as 1 = male, 2 = female. *p < 0.05. **p < 0.01. ***p < 0.001.*

intercepts were constrained to be equal across groups) as well as the changes in the CFI and the RMSEA.

As shown in **Table 5**, the RMSEA changes were <|0.015| for benevolent and corrective humor in each group (i.e., the samples, age groups, and gender). The CFI changes were <|0.01| for the age groups (metric invariance) and gender (scalar invariance), but not for the samples (metric invariance) and the age groups (scalar invariance). Thus, follow-up analyses were conducted for assessing partial measurement invariance, comparing the metric invariance of each of the 12 marker items for the samples and the scalar invariance for the age groups. For the samples, metric invariance was supported for each item, as the CFI change between the baseline model and the metric invariance model was <|0.01| (range |0.001|–|0.008|). For the age groups, the CFI change was also <|0.01| for all items (range |0.000|–|0.008|) with the exception of Item 9 (|0.029|). Thus, partial metric invariance was supported across the samples, partial scalar invariance was supported across the age groups, and scalar invariance was supported for gender. This indicates (a) that benevolent and corrective humor were measured the same way across the different samples, (b) that the factors of the different samples were comparable, and (c) that the mean differences between the age groups and gender could be attributed to mean differences in benevolent and corrective humor. This allows to meaningfully compare the mean-level differences between the BenCor scores across the age groups and gender.

Factorial Validity

The factorial validity of the 12 marker items of benevolent and corrective humor was first tested in an exploratory fashion with Tucker's phi as an index of factor congruence. The 12 marker items were subjected to a PCA with oblimin rotation, in which two factors were extracted. The benevolent and corrective humor factors were then compared with the Swiss student sample, for which the BenCor was originally developed. As shown in **Table 3**, Tucker's phi indicated factor equality for 14 samples and a fair factor similarity for 8 samples. Lower values were obtained for India and the Turkey graduate sample, for which the extracted BenCor factor was not similar to the comparison sample. The median Tucker's phi value across the 25 samples was 0.95, indicating that the benevolent humor factor showed cross-cultural equality. For the corrective humor factor, 14 samples showed factor equality, and 10 samples indicated a fair factor similarity. With a median of 0.95, cross-cultural factor equality could also be supported for the corrective humor factor.

TABLE 4 | Overview of the fit indices of confirmatory factor analyses of the 6 marker items (one-factor models indicating unidimensionality/homogeneity) separate for benevolent and corrective humor across the 25 BenCor samples in the 22 countries.

Countries	Benevolent humor ($df = 9$)					Corrective humor ($df = 9$)				
	χ^2	χ^2/df	CFI	RMSEA	SRMR	χ^2	χ^2/df	CFI	RMSEA	SRMR
Austria	16.22	1.80	0.97	0.05	0.03	28.31**	3.15	0.96	0.08	0.04
Chile	24.65**	2.74	0.93	0.09	0.05	24.53**	2.73	0.96	0.09	0.04
China	17.91*	1.99	0.95	0.06	0.05	22.12**	2.46	0.93	0.08	0.05
Costa Rica	3.83	0.43	1.00	0.00	0.02	7.48	0.83	1.00	0.00	0.02
Croatia	16.83	1.87	0.95	0.05	0.04	13.23	1.47	0.98	0.04	0.03
Germany	21.04*	2.34	0.95	0.07	0.04	18.16*	2.02	0.98	0.06	0.03
India	12.48	1.39	0.93	0.04	0.05	20.98*	2.33	0.92	0.08	0.05
Italian	18.51*	2.06	0.91	0.07	0.05	12.91	1.43	0.99	0.05	0.04
Latvia	16.91	1.88	0.98	0.05	0.03	52.26***	5.81	0.92	0.11	0.05
Lebanon	25.70**	2.86	0.84	0.08	0.05	33.43***	3.71	0.87	0.10	0.06
Malaysia	21.16*	2.35	0.86	0.08	0.05	11.01	1.22	0.99	0.03	0.03
Malaysia (Terengganu)	10.61	1.18	0.97	0.03	0.04	12.47	1.39	0.97	0.04	0.04
Mexico	11.92	1.32	0.96	0.04	0.05	8.44	0.94	1.00	0.00	0.03
New Zealand	18.49*	2.05	0.86	0.07	0.06	15.01	1.67	0.98	0.06	0.04
Poland	16.57	1.84	0.98	0.04	0.03	26.47**	2.94	0.97	0.07	0.03
Russia	20.66*	2.30	0.87	0.08	0.05	17.83*	1.98	0.96	0.07	0.05
Slovakia	22.88**	2.54	0.93	0.06	0.04	26.71**	2.97	0.96	0.07	0.04
Spain	16.85	1.87	0.93	0.07	0.05	11.84	1.32	0.99	0.04	0.03
Switzerland (community)	12.50	1.39	0.99	0.03	0.02	16.16	1.80	0.99	0.04	0.02
Switzerland (students)	6.59	0.73	1.00	0.00	0.02	6.07	0.67	1.00	0.00	0.02
Taiwan	31.62***	3.51	0.93	0.08	0.05	29.35***	3.26	0.97	0.07	0.04
Turkey (non-graduates)	21.33*	2.37	0.95	0.06	0.04	15.93	1.77	0.98	0.05	0.03
Turkey (graduates)	39.89***	4.43	0.77	0.10	0.08	42.88***	4.76	0.85	0.11	0.06
UK (mostly England)	18.66*	2.07	0.95	0.06	0.05	14.45	1.61	0.99	0.05	0.03
UK (Northern Ireland)	6.74	0.75	1.00	0.00	0.03	15.07	1.67	0.97	0.06	0.04

*CFI, Comparative fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.*

Next, the factor structure was investigated in CFAs. Both one-factor and two-factor models were estimated based on the 12 marker items, and their fit indices are shown in **Table 6**.

As expected, the one-factor model indicated an unacceptable fit in all samples except for India, for which only the CFI was unacceptable. By contrast, the two-factor model showed an acceptable or good fit in all indices (except for the CFI) in 20 of the 25 samples. An unacceptable fit in at least two indices was obtained for China, Costa Rica, Latvia, and the two Turkish samples. These findings mostly support the two-factor structure of the BenCor.

Next, the intercorrelations of benevolent and corrective humor are of interest. **Table 3** shows the observed intercorrelations and the factor correlations (from the PCA with oblimin rotation), and **Table 6** shows the latent correlations in the two-factor CFA model. In line with the conceptualization of the BenCor, all correlations between benevolent and corrective humor were significant and positive (medium to large effects). The numerically lowest correlations were obtained in Russia, and the highest correlations were obtained in Costa Rica, India, and Malaysia (Terengganu sample). Median correlations were 0.40 for the observed scores, 0.28 for the PCA factors, and 0.53 for the CFA factors. Thus, both the individual samples and the median correlations suggested that benevolent and corrective humor overlap. Still, they can be distinguished from one another, with a median of 28.1% shared true-score variance. Overall, the factorial validity of the BenCor can be supported, albeit to a lesser extent for the samples from India and Turkey (mainly the graduate sample).

Factor analyses (PCA with oblimin rotation and CFA) were also conducted across the full sample of 7,226 participants. The first four eigenvalues in the PCA were 3.67, 1.52, 1.00, and 0.86. Both the scree test and Horn's parallel analysis indicated the retention of two factors, which together explained 43.3% of the variance in the 12 marker items. The loadings and factor intercorrelations are presented in **Table 7**.

As shown in **Table 7**, each item had its highest loading on the expected factor in the PCA. Main loadings ranged from 0.31 to 0.75 for the benevolent humor factor and from 0.50 to 0.77 for the corrective humor factor. A few cross-loadings were substantial. Item 3 loaded on the corrective factor almost as strongly as on the benevolent factor. By contrast, item 7 had a small negative loading on the corrective humor factor. Items 8 and 12 showed small positive loadings on the benevolent humor factor. In the

TABLE 5 | Fit indices of models assessing metric (fixed loadings) invariance of benevolent and corrective humor across samples.

Measurement invariance models	*df*	AIC	CFI	RMSEA	CFI change	RMSEA change
BENEVOLENT HUMOR						
25 samples (*N* = 7,226)						
Baseline model	225	144,126	0.95	0.06	–	–
Metric invariance	345	144,103	0.94	0.05	0.014	0.005
Age (across all samples, 6 age groups[a])						
Baseline model	54	147,378	0.95	0.05	–	–
Metric invariance	79	147,352	0.95	0.05	0.001	0.009
Scalar invariance	104	147,562	0.90	0.06	0.053	0.012
Gender (across all samples[b])						
Baseline model	18	147,890	0.96	0.05	–	–
Metric invariance	23	147,891	0.95	0.05	0.003	0.005
Scalar invariance	28	147,964	0.94	0.05	0.003	0.000
CORRECTIVE HUMOR						
25 samples (*N* = 7,226)						
Baseline model	225	156,578	0.97	0.07	–	–
Metric invariance	345	156,676	0.94	0.07	0.025	0.004
Age (across all samples, 6 age groups[a])						
Baseline model	54	159,516	0.97	0.06	–	–
Metric invariance	79	159,497	0.97	0.05	0.003	0.007
Scalar invariance	104	159,658	0.95	0.06	0.023	0.008
Gender (across all samples[b])						
Baseline model	18	159,731	0.98	0.05	–	–
Metric invariance	23	159,726	0.97	0.05	0.001	0.005
Scalar invariance	28	159,736	0.97	0.04	0.003	0.002

AIC, Akaike's information criterion; CFI, comparative fit index; RMSEA, root mean square error of approximation.
[a] *18–20 years (n = 1,624), 21–24 years (n = 1,981), 25–29 years (n = 1,081), 30–39 years (n = 1,225), 40–49 years (n = 704), 50+ years (n = 580).*
[b] *n = 2,906 males and n = 4,312 females.*

CFA, all loadings were positive and significant ($p < 0.001$). They ranged from 0.43 to 0.65 for the benevolent humor factor, and from 0.51 to 0.68 for the corrective humor factor. The fit of the two-factor CFA model was unacceptable, with $\chi^2 = 1{,}560.07$, $df = 53$, $\chi^2/df = 29.44$, CFI = 0.89, RMSEA = 0.06, and SRMR = 0.05. Still, the two-factor model clearly fitted the data better than the one-factor model ($\chi^2 = 3{,}123.43$, $df = 54$, $\chi^2/df = 57.84$, CFI = 0.78, RMSEA = 0.09, and SRMR = 0.07). According to the modification indices, the model fit of the two-factor model could be improved by freeing the loading of item 3 on corrective humor, and the loadings of items 8 and 12 on benevolent humor. The factor correlations were 0.35 for the PCA and 0.58 for the CFA, again indicating a strong overlap, yet no redundancy between the two factors. Thus, although not perfectly aligning with a simple structure, the two factors of benevolent and corrective humor could be clearly separated.

Discriminant Validity

Table 6 also shows the square root of the AVE of the benevolent and corrective humor factors for each sample. Comparing the CFA factor correlations with the square root of the AVE, the Fornell-Larcker criterion was met for benevolent humor in 13 of the 25 samples, and for corrective humor in 18 of 25 samples. The strongest deviations were found for the Indian, the Malaysian (Terengganu), and the two Turkish samples due to their large factor correlations ($rs \geq 0.65$). Conducting the same analyses across the 25 samples, the square root of the AVE of the benevolent humor factor (0.50) was smaller than the factor correlation (0.58), while the square root of the AVE of the corrective humor factor (0.59) was larger than the factor correlation. Thus, discriminant validity for the benevolent humor factor was only partially supported in terms of the Fornell-Larcker criterion, while the discriminant validity of the corrective humor factor received stronger support.

Item Comparisons across Samples

Tables 8, **9** present the means and CITCs of the benevolent and corrective humor items in the 25 samples.

As shown in **Tables 8**, **9**, the samples exhibited systematic patterns in terms of the item means and CITCs. First, the means of the benevolent humor items were rather similar across the samples, ranging from 3.69 to 4.96 for the minima and 5.23 to 6.13 for the maxima, while more variation was found for corrective humor, with the minima ranging from 2.78 to 4.31 and the maxima ranging from 3.90 to 5.47. Second, for benevolent humor, item 11 showed the lowest mean in 17 of the 25 samples,

TABLE 6 | Overview of the fit indices of confirmatory factor analyses of the 12 marker items (one-factor and two-factor models) across the 25 bencor samples in the 22 countries.

Countries	One-factor model ($df = 54$)					Two-factor model ($df = 53$)							
	χ^2	χ^2/df	CFI	RMSEA	SRMR	χ^2	χ^2/df	CFI	RMSEA	SRMR	*r*	AVE_{Ben}	AVE_{Cor}
Austria	332.77	6.16	0.68	0.12	0.10	136.19	2.57	0.91	0.07	0.06	0.47	0.61	0.67
Chile	264.63	4.90	0.67	0.13	0.12	116.83	2.20	0.90	0.07	0.07	0.45	0.62	0.67
China	228.77	4.24	0.66	0.12	0.11	154.78	2.92	0.80	0.09	0.09	0.40	0.57	0.56
Costa Rica	171.18	3.17	0.82	0.11	0.08	135.89	2.56	0.87	0.10	0.07	0.74	0.61	0.70
Croatia	174.31	3.23	0.72	0.08	0.08	104.61	1.97	0.88	0.05	0.06	0.46	0.49	0.53
Germany	204.73	3.79	0.82	0.10	0.09	103.25	1.95	0.94	0.06	0.06	0.59	0.57	0.70
India	98.80	1.83	0.83	0.07	0.07	93.09	1.76	0.85	0.06	0.07	0.81	0.42	0.53
Italian	184.92	3.42	0.71	0.11	0.10	125.67	2.37	0.84	0.08	0.08	0.40	0.52	0.64
Latvia	414.41	7.67	0.70	0.13	0.10	294.12	5.55	0.80	0.11	0.09	0.59	0.62	0.62
Lebanon	185.09	3.43	0.63	0.10	0.08	138.20	2.61	0.76	0.08	0.07	0.43	0.45	0.50
Malaysia	126.15	2.34	0.81	0.08	0.07	92.95	1.75	0.89	0.06	0.06	0.64	0.48	0.57
Malaysia (Terengganu)	118.27	2.19	0.77	0.08	0.08	115.55	2.18	0.77	0.08	0.08	0.80	0.46	0.53
Mexico	117.61	2.18	0.76	0.08	0.08	76.70	1.45	0.91	0.05	0.06	0.53	0.48	0.55
New Zealand	165.11	3.06	0.70	0.10	0.09	108.16	2.04	0.85	0.07	0.06	0.41	0.47	0.63
Poland	348.68	6.46	0.72	0.11	0.09	187.80	3.54	0.87	0.08	0.07	0.50	0.56	0.60
Russia	183.33	3.40	0.61	0.11	0.11	110.20	2.08	0.83	0.07	0.08	0.25	0.48	0.58
Slovakia	222.10	4.11	0.79	0.09	0.07	145.07	2.74	0.88	0.07	0.06	0.63	0.52	0.62
Spain	167.84	3.11	0.74	0.10	0.10	89.27	1.68	0.92	0.06	0.06	0.38	0.51	0.63
Switzerland (general)	286.50	5.31	0.82	0.09	0.07	146.47	2.76	0.93	0.06	0.05	0.65	0.57	0.66
Switzerland (students)	265.07	4.91	0.73	0.11	0.09	78.40	1.48	0.97	0.04	0.04	0.53	0.64	0.67
Taiwan	305.15	5.65	0.75	0.10	0.09	161.41	3.05	0.89	0.07	0.07	0.47	0.56	0.64
Turkey (non-graduates)	197.14	3.65	0.79	0.09	0.07	160.25	3.02	0.85	0.08	0.07	0.70	0.44	0.51
Turkey (graduates)	225.70	4.18	0.65	0.10	0.08	208.00	3.92	0.69	0.10	0.09	0.65	0.53	0.55
UK (England)	201.82	3.74	0.77	0.10	0.08	126.66	2.39	0.89	0.07	0.07	0.59	0.53	0.62
UK (Northern Ireland)	126.42	2.34	0.76	0.08	0.08	84.19	1.59	0.90	0.05	0.06	0.54	0.49	0.58

CFI, Comparative fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; r, correlation between the latent benevolent and corrective humor factors; AVE, square root of average variance explained.
All χ^2 values were significant at $p < 0.05$.

while the highest mean was found for item 5 (14 samples). For corrective humor, item 4 showed the lowest mean in 10 of the 25 samples, and the highest mean was found for item 2 (11 samples).

As also shown in **Tables 8**, **9**, none of the items exhibited negative CITCs, indicating that they were all aligned with the total score. Only four samples had CITCs below 0.20, namely India, Malaysia (Terengganu sample), and the Turkey graduate sample for benevolent humor and Russia for corrective humor. The highest values were 0.65 for benevolent humor and 0.72 for corrective humor, indicating that none of the items were redundant. Thus, the psychometric properties of the single marker items seem mostly sufficient. The lowest CITC was found for the benevolent humor item 3 (14 samples), and the highest CITC was found for item 5 (17 samples). For corrective humor, the lowest CITCs were found for items 2 and 8 (11 samples), and the highest CITCs was found for item 10 (14 samples).

Profile Similarities between the Samples

The similarities of the samples across the 12 BenCor items were investigated using multidimensional scaling. A two-dimensional solution was chosen (stress function = 0.19, variance explanation 87.4%), which is plotted in **Figure 1**.

To interpret the solution, the two resulting dimensions were correlated with benevolent and corrective humor and with the single marker items. Dimension 1 correlated strongly with both benevolent [$r_{(25)} = 0.82$, $p < 0.001$] and corrective humor [$r_{(25)} = 0.91$, $p < 0.001$]. That is, Dimension 1 was sensitive to the overall mean differences, contrasting samples with high scores in benevolent and corrective humor (e.g., Italy, India, and Chile) with samples with lower scores (e.g., Lebanon, Russia, and the two Turkish samples). As benevolent and corrective humor showed large positive correlations across the samples, it is not surprising that one dimension of mean-level differences rather than two separate dimensions emerged. Dimension 2 was not significantly correlated with either benevolent or corrective humor (all $ps \geq 0.07$), and thus correlations at the item level were investigated (for which the significance level was set to 0.01 due to the multiple comparisons). Dimension 2 showed significant correlations with the benevolent humor items 3 [$r_{(25)} = -0.55$, $p = 0.005$] and 7 [$r_{(25)} = 0.64$, $p = 0.001$] and the corrective humor items 8 [$r_{(25)} = 0.87$, $p < 0.001$] and 12 [$r_{(25)} = 0.67$,

TABLE 7 | Loadings and factor intercorrelations of a joint Principal Component Analysis (PCA with oblimin rotation) and a Confirmatory Factor Analysis (CFA with the MLM-Estimator) across the 25 samples.

	Descriptives		PCA		CFA	
	M	*SD*	Ben factor	Cor factor	Ben factor	Cor factor
BENEVOLENT HUMOR ITEMS						
Item 1	5.01	1.39	0.59	−0.04	0.43***	–
Item 3	5.10	1.56	0.31	0.30	0.44***	–
Item 5	5.50	1.34	0.69	0.06	0.65***	–
Item 7	5.31	1.32	0.75	−0.18	0.48***	–
Item 9	5.34	1.37	0.59	0.09	0.54***	–
Item 11	4.58	1.56	0.59	0.11	0.56***	–
CORRECTIVE HUMOR ITEMS						
Item 2	4.53	1.72	−0.09	0.67	–	0.51***
Item 4	3.97	1.76	−0.03	0.71	–	0.60***
Item 6	4.18	1.75	−0.07	0.73	–	0.59***
Item 8	4.19	1.63	0.23	0.50	–	0.56***
Item 10	3.96	1.73	−0.03	0.77	–	0.68***
Item 12	4.18	1.58	0.21	0.56	–	0.61***
Factor correlation				0.35		0.58***

*N = 7,226. ***p < 0.001.*

TABLE 8 | Minima and maxima of the item means and of the Corrective Item-Total Correlations (CITC) of the benevolent humor items in the 25 samples in the 22 countries.

Countries	Item means				CITC			
	Min	Item	Max	Item	Min	Item	Max	Item
Austria	4.61	11	5.74	5	0.40	1	0.62	7
Chile	4.61	1	5.72	9	0.39	1	0.63	5
China	4.85	11	5.38	5	0.41	1	0.62	5
Costa Rica	4.76	1	5.24	9	0.36	1	0.59	5
Croatia	4.69	11	5.96	5	0.35	1+9	0.45	5
Germany	4.22	11	5.40	9	0.34	3	0.54	5+7+9
India	4.38	11	5.70	5	0.10	3	0.38	1
Italian	4.41	11	5.83	5	0.27	11	0.48	1
Latvia	4.61	11	5.91	5	0.42	3	0.61	5
Lebanon	4.18	11	5.23	7	0.24	3	0.40	5
Malaysia	4.70	3	5.72	7	0.32	3+7	0.47	5
Malaysia (Terengganu)	4.37	3	6.13	7	0.18	3	0.42	1
Mexico	4.83	11	5.54	5	0.27	1	0.47	9
New Zealand	4.95	11	5.67	5	0.25	1	0.42	7
Poland	4.60	11	5.79	5	0.40	3	0.53	5
Russia	4.59	9	5.51	5	0.21	3	0.50	7
Slovakia	4.58	1	5.61	5	0.32	11	0.55	5
Spain	4.69	11	6.00	9	0.25	3	0.50	5
Switzerland (community)	4.28	11	5.24	9	0.40	3	0.56	9
Switzerland (students)	4.42	11	5.43	3	0.46	1+3	0.65	5
Taiwan	4.79	11	5.35	5	0.35	3	0.56	5
Turkey (non-graduates)	3.85	3	5.66	9	0.28	3	0.54	5
Turkey (graduates)	3.69	3	5.86	9	0.11	3	0.38	5
UK (mostly England)	4.71	11	5.46	5	0.32	1	0.50	5+9
UK (Northern Ireland)	4.96	11	5.81	5	0.27	11	0.54	5

TABLE 9 | Minima and maxima of the item means and of the Corrective Item-Total Correlations (CITC) of the corrective humor items in the 25 samples in the 22 countries.

Countries	Item means				CITC			
	Min	Item	Max	Item	Min	Item	Max	Item
Austria	3.80	4	4.90	2	0.52	8	0.66	4
Chile	4.31	12	5.00	6	0.44	8	0.63	2
China	3.52	8	4.91	2	0.39	2	0.49	4+8
Costa Rica	4.16	2	4.87	8	0.58	2	0.70	12
Croatia	4.07	6	5.17	2	0.34	8	0.55	10
Germany	3.61	4	4.81	2	0.55	2	0.72	10
India	3.91	6	5.47	12	0.35	12	0.51	4
Italian	3.75	8	5.15	6	0.50	4+6	0.64	10
Latvia	3.99	6	4.71	12	0.35	12	0.62	10
Lebanon	2.84	6	4.46	4	0.31	2	0.51	10
Malaysia	3.66	4	4.62	2	0.30	8	0.56	12
Malaysia (Terengganu)	3.09	10	5.19	8	0.33	8	0.49	4
Mexico	3.23	10	4.22	4	0.33	8	0.59	10
New Zealand	4.06	4	4.81	2	0.48	2+8	0.66	10
Poland	4.05	4	4.73	8	0.39	2	0.60	10
Russia	3.30	10	3.90	12	0.13	12	0.66	8
Slovakia	3.89	2	4.34	8	0.41	2	0.64	10
Spain	3.93	2	4.53	4	0.42	8	0.65	10
Switzerland (community)	3.73	4	4.96	2	0.43	2	0.65	4+10
Switzerland (students)	3.78	4	4.86	2	0.52	8	0.64	4
Taiwan	3.75	8	4.74	2	0.37	2	0.62	10
Turkey (non-graduates)	3.17	4	4.37	6+8	0.33	8	0.54	6
Turkey (graduates)	2.78	4	4.55	8	0.32	8	0.50	6+10
UK (mostly England)	3.90	4	4.81	2	0.38	2	0.64	10
UK (Northern Ireland)	4.04	10	5.06	2	0.36	2	0.54	8

$p < 0.001$]. Thus, this dimension distinguished samples that were comparably high in three items (7, 8, and 12) and comparably low in item 3. As shown in **Figure 1**, most samples were rather similar in this dimension, while India, Malaysia (Terengganu region), and the Turkish graduate sample had the highest scores, and Lebanon, Russia, Italy, and China had the lowest scores. This dimension might capture the extent to which item 3 had a corrective connotation and items 8 and 12 had a benevolent connotation, thus potentially decreasing the mean of item 3 and increasing the means of items 8 and 12. In fact, India, Malaysia (Terengganu region), and the Turkish graduate sample showed zero or even negative loadings of item 3 on the benevolent humor factor in the PCA, and items 8 and 12 showed large positive loadings on the benevolent and the corrective humor factor.

Focusing on the similarity of the countries that shared the same language, item-profile comparisons were conducted. **Figure 2** illustrates the item distributions of the English-, German-, and Spanish-speaking samples.

When correlating the samples across the 12 items, a median correlation of 0.97 was found for the English- and the German-speaking countries and a correlation of 0.88 was found for the Spanish-speaking countries. This similarity can also be seen in **Figure 2**, as the English- and German-speaking countries shared a similar item profile, while the Spanish countries differed more strongly from one another. This similarity was numerically higher than the correlations across the three different languages (0.94 for English and German, 0.80 for English and Spanish, and 0.76 for German and Spanish). Thus, the item mean profiles were most similar for the two Germanic languages, and less similar for Spanish (a Romance language).

Further comparisons were undertaken between the four countries that had two samples each (i.e., Malaysia, Switzerland, Turkey, and the UK). The item-profile correlations within the countries were 0.82 (Malaysia), 0.97 (Switzerland), 0.98 (Turkey), and 0.97 (the UK), indicating a strong similarity within the countries. Importantly, each of these correlations was numerically higher than the correlations between the countries, for which the medians were 0.69, 0.74, 0.66, and 0.77 (for Malaysia, Switzerland, Turkey, and the UK, respectively). This supports the notion that the item profiles of the BenCor were more similar within than between countries.

Comparisons across Age Groups and Gender

Comparisons of the six age groups were conducted with ANCOVAs, controlling for gender. The main effect of age group was significant both for benevolent humor [$F_{(5)} = 3.98$, $p = 0.001$, $\eta_p^2 = 0.002$] and corrective humor [$F_{(5)} = 5.01$,

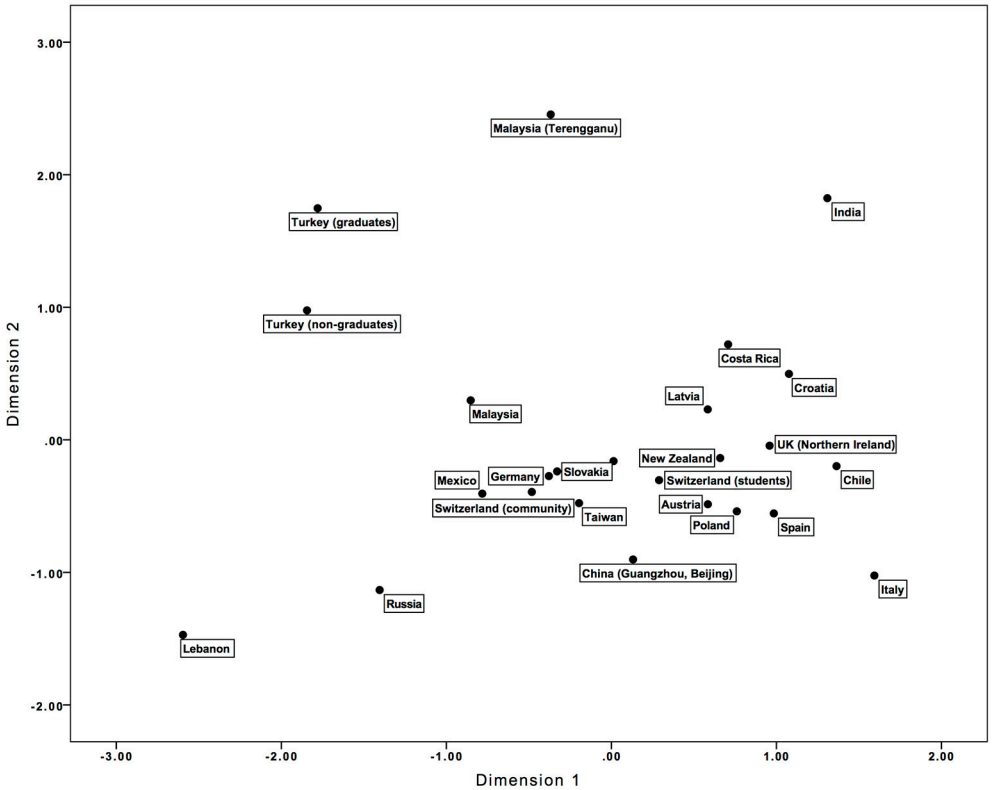


FIGURE 1 | Two-dimensional plot derived from multidimensional scaling of the 12 BenCor items.

$p < 0.001$, $\eta_p^2 = 0.003$]. Polynomial contrasts revealed a significant linear trend in benevolent humor (contrast = 0.12, $p < 0.001$), indicating a linear increase with age. For corrective humor, both the linear (contrast = −0.12, $p = 0.001$) and quadratic trends were significant (contrast = −0.15, $p < 0.001$). The means and 95% confidence intervals are shown in **Figure 3A**.

As shown in **Figure 3A**, corrective humor tended to increase until the age group of 30–39 years, and then decreased for the age groups of 40–49 and 50+ years. Taking a look at the individual items, ANCOVAs controlling for gender revealed significant main effects for all items (all ps < 0.05), except for items 2 ($p = 0.679$) and 7 ($p = 0.755$). Effect sizes were mostly negligible ($\eta_p^2 < 0.01$), with small effects obtained for items 4 ($\eta_p^2 = 0.011$) and 9 ($\eta_p^2 = 0.023$). Significant linear trends were found for the benevolent humor items 1 (contrast = 0.14, $p = 0.003$), 3 (contrast = 0.16, $p = 0.002$), 9 (contrast = 0.53, $p < 0.001$), and 11 (contrast = −0.15, $p = 0.003$). Items 1, 3, and 9 increased with age (in line with benevolent humor), while item 11 tended to decrease with age (see **Figure 3B**). For corrective humor, linear trends were significant for items 4 (contrast = −0.49, $p < 0.001$), 6 (contrast = 0.27, $p < 0.001$), 8 (contrast = −0.21, $p < 0.001$), 10 (contrast = −0.22, $p < 0.001$), and 12 (contrast = −0.11, $p = 0.039$). Additionally, significant quadratic trends were found for items 4 (contrast = −0.14, $p = 0.013$), 6 (contrast = −0.17, $p = 0.002$), 8 (contrast = −0.22, $p < 0.001$), 10 (contrast = −0.23, $p < 0.001$), and 12 (contrast = −0.12, $p = 0.015$). The negative linear and quadratic trends of Items 4, 8, 10, and 12 were in line with the age trends of corrective humor. Item 6, however, showed a positive linear trend in addition to the negative quadratic trend (see **Figure 3C**).

Regarding gender differences in benevolent and corrective humor, **Table 3** shows the correlations with gender for every sample (with males coded as 1 and females coded as 2). Most correlations with benevolent humor were small and not significant (range −0.14 to 0.11, $Mdn = -0.04$). By contrast, most correlations with corrective humor were negative and significant (range −0.02 to −0.38, $Mdn = -0.21$). When the full sample was analyzed, benevolent humor showed a negligible negative correlation with gender [$r_{(7,218)} = -0.05$, $p < 0.001$], while corrective humor showed a medium-sized negative correlation [$r_{(7,218)} = -0.22$, $p < 0.001$]. Thus, gender differences were similar across the samples, and males and females did not substantially differ in their levels of benevolent humor, while males scored higher than females in corrective humor. Comparisons were also conducted for the single items. Significant differences were found for the benevolent humor items 3 and 5, and 11 [$rs_{(7,218)} \leq -0.10$, all ps < 0.02] and for all corrective humor items [$rs_{(7,218)} = -0.11$ to −0.18, all ps < 0.001], indicating that males always scored higher than females. Thus, the benevolent humor items showed only negligible gender differences, while the corrective humor items consistently showed small gender differences.

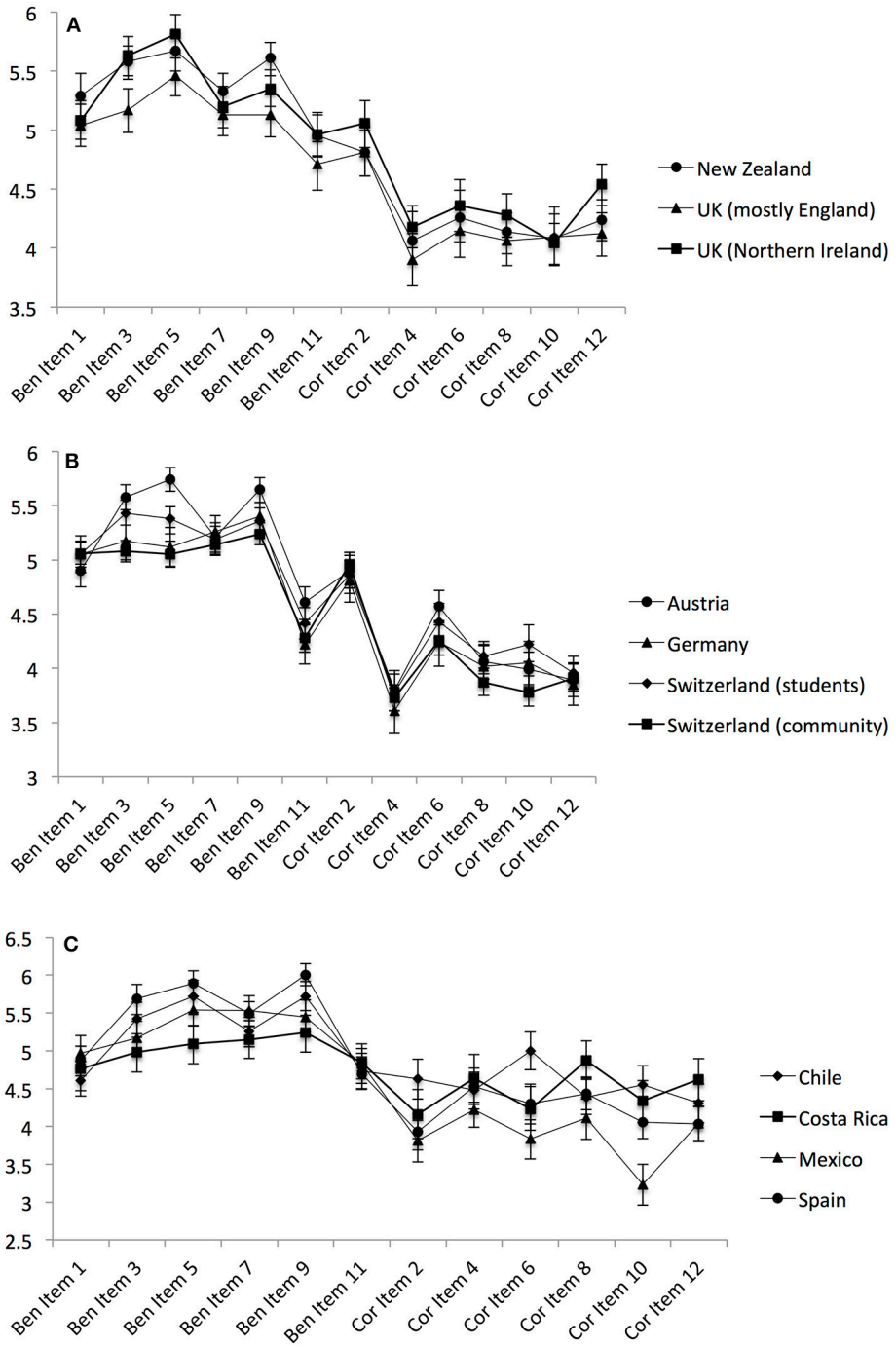


FIGURE 2 | Comparisons of the 12 BenCor items within samples sharing the same language. The upper panel depicts English-speaking samples **(A)**, the middle panel depicts German-speaking samples **(B)**, and the lower panel depicts Spanish-speaking samples **(C)**.

DISCUSSION

The aim of this study was to compare the psychometric properties of the BenCor (Ruch, 2012) across 25 samples from 22 countries. The means and standard deviations differed across the 25 samples, though they all had in common that benevolent humor was more strongly endorsed than corrective humor (around 1 scale point difference). Thus, participants across countries engaged in virtue-related humor, with the benevolent style being more prevalent than the corrective and critical style.

The reliability of both benevolent and corrective humor was supported in most of the samples. Internal consistencies

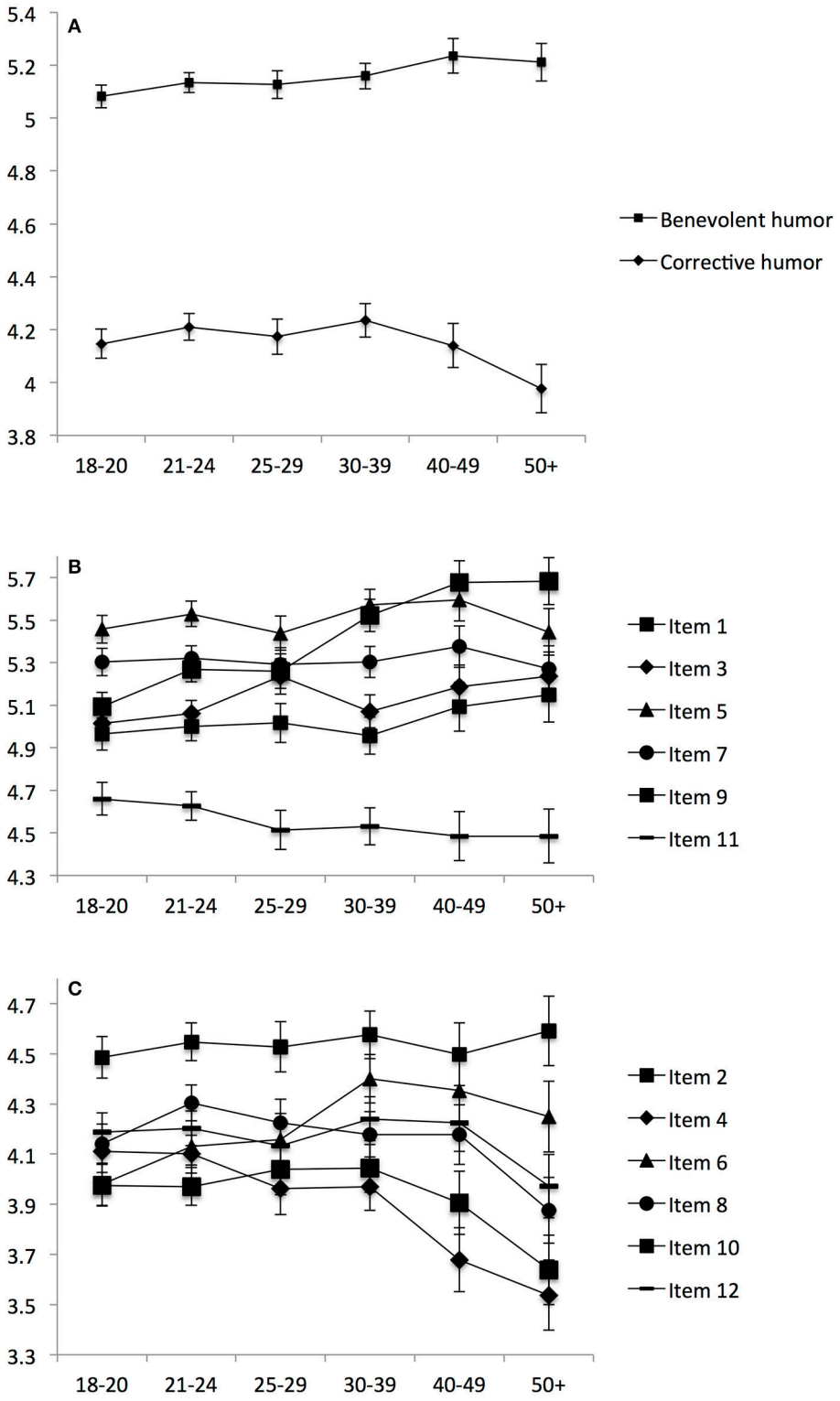


FIGURE 3 | Means with 95% confidence intervals of benevolent and corrective humor **(A)**, the benevolent humor items **(B)**, and the corrective humor items **(C)** for each of six age groups.

were acceptable, or good, in all samples for corrective humor, while benevolent humor showed somewhat lower values, which were especially low in three samples (India, the Malaysia Terengganu sample, and the Turkish graduate sample). Similarly, unidimensionality was supported in all samples, with the exception of three samples for benevolent (Chile, Taiwan, and the Turkish graduate sample) and corrective humor (Latvia, Lebanon, and the Turkish graduate sample). Thus, the reliability of the sets of marker items of benevolent and corrective humor was either fully or partially supported (except for the Turkish graduate sample). This indicates that the six marker items indeed tapped into a common underlying dimension and that their intercorrelations were positive and sufficient. Thus, despite the brevity of the questionnaire and the rather different contents covered by the marker items (see Ruch and Heintz, 2016), the BenCor seems to be able to measure benevolent and corrective humor reliably across different cultures and languages.

Next, measurement invariance was tested across samples, age groups, and gender. While metric invariance was only partially supported for benevolent and corrective humor across the 25 samples, each of the 12 marker items exhibited metric invariance, thereby allowing comparisons of the factors across the samples (Chen, 2008). For the age groups, metric invariance was supported for benevolent and corrective humor and scalar invariance was supported at the item level (with the exception of item 9). For gender, metric and scalar invariance was fully supported. Thus, both the factors and the means of these groups can be validly compared and are not biased (Chen, 2008). These findings pave the way for comparisons of benevolent and corrective in different countries, in different age groups (e.g., for investigating developmental changes), and for investigating gender differences.

The discriminant validity of the BenCor was partially confirmed using the Fornell-Larcker criterion (Fornell and Larcker, 1981). Specifically, the square root of the AVE of the latent benevolent and corrective humor factors were higher than the correlation between the two factors in 13 and 18 of the 25 samples, respectively. In other words, in more than half of the samples, the variance explanation of the latent benevolent and corrective humor factors in the 12 marker items was higher than the shared variance between the latent factors. Thus, the differences between the two styles of virtue-related humor (i.e., benevolent vs. critical treatment of human weaknesses and wrongdoings) were more pronounced than the similarities (i.e., virtuousness and aiming at the good). Still, the marker items of benevolent humor showed a comparably smaller overlap with their factor, which also fits to the finding that internal consistencies of benevolent humor were lower. Maybe the benevolent humor marker items capture more heterogeneous contents, or maybe the construct itself is more complex. The discrimination among benevolent and corrective humor could be improved by adapting some of the 12 marker items that showed cross-loadings in the PCA and high modification indices in the CFA (i.e., items 3, 8, and 12). This would help to reduce the factor correlation in the CFA. Additionally, more items could be written, which are not merely markers of benevolent and corrective humor, but which represent both constructs comprehensively.

Factorial Validity

Factorial validity for the BenCor was supported both in an exploratory and a confirmatory fashion. First, Tucker's phi indicated that the benevolent and corrective humor factors were fairly similar or equivalent to the Swiss comparison sample (except for the Indian and the Turkish graduate sample). As Tucker's phi is sensitive to differences in item loadings (see Lorenzo-Seva and Ten Berge, 2006), this is in line with the finding of metric invariance of the BenCor; in other words, all samples had similar factor loadings, and thus the meaning and conceptualization of the factors were comparable across samples. Second, CFAs within each sample showed that a two-factor structure fitted the data well in most samples, while the one-factor model did not show an acceptable fit. Also, the true-score correlation between benevolent and corrective humor was much lower than 1 (with a maximum of 64.0% shared true-score variance between the factors). Thus, despite their predictable overlap, benevolent and corrective humor constitute separate factors that capture different forms of virtue-related humor.

Regarding the suitability of the items for the two factors, the PCA across the full sample revealed cross-loadings of items 3, 7, 8, and 12. These differences also aligned well with the profile similarities across the 12 BenCor items, which revealed that the sample similarities were due to the overall mean differences in benevolent and corrective humor (Dimension 1) and due to deviations in 4 items (3, 7, 8, and 12; Dimension 2). Several explanations can be offered for these findings, drawing on both cross-cultural and culture-specific explanations.

Item 3 had similar loadings both on benevolent (0.31) and corrective humor (0.30). This could be due to the low CITCs obtained for this item in 14 of the 25 samples, indicating that this item related less strongly to the total score of benevolent humor than the other items did. It is noticeable that this is the only item that refers to the inclusion of oneself and others when making fun of human weaknesses, while the other items entail the idea of "we, as humans, are all in this together" more directly. Conversely, this item more directly incorporates making fun of human weaknesses ("aiming at"), while the other items rather refer to humor appreciation (e.g., being amused or smiling) or only indirectly entail humor production (treating benevolently). This might shift item 3 to corrective humor, as the latter directly incorporates humor production. Furthermore, PCAs within the samples revealed mismatched loadings (i.e., higher loadings on corrective than on benevolent humor) only for India, the Malaysian Terengganu sample, and for the Turkish graduate sample.

The slightly negative loading of item 7 on corrective humor could be due to it being the only benevolent humor item that explicitly includes the underlying accepting attitude. While both benevolent and corrective humor share detecting weaknesses and treating them humorously, benevolent humor treats them in an accepting manner, while in corrective humor they are not accepted, but instead corrected.

Item 8 had small positive loadings on benevolent humor, which might be due to the softener "gently urge," which bears resemblance to the benevolent and kind-hearted treatment of weaknesses in benevolent humor. Likewise, "to caricature" might imply a more playful and less critical treatment, and it might additionally be confused with drawing caricatures instead of parodying the wrongdoings physically and verbally. This item had higher loadings on benevolent than corrective humor in six samples (Croatia, India, the two Malaysian samples, and the two Turkish samples).

Finally, item 12 also had small positive loadings on benevolent humor. "Poking fun" is rather soft expression for ridiculing others and might thus have a more entertaining than critical connotation. Likewise, "hoping to improve" focuses on one's optimistic outlook, which might be similar to the humorous outlook entailed in benevolent humor. This item had higher loadings on benevolent than corrective humor in four samples (India, Latvia, Russia, and the Turkish graduate sample).

Several culture-specific differences in the understanding of the items and factors could be hypothesized, which might help to explain some of the deviations found in the factor analyses. For example, in Malaysia (Terengganu region), several informal interviews suggested that corrective humor seems to have an inherent benevolence, as close bonds exist between people and informing others about their wrongdoings in a respectful, but also humorous manner is expected and encouraged within friendships. Thus, the virtuous aspect of corrective humor might be stronger in this culture, also distinguishing this sample from the general Malaysian sample. In the Croatian, Indian, and Latvian contexts, corrective humor might not be employed at the societal level very often, perhaps because people do not feel that they can produce a change, and people might thus rather adjust than try to change the conditions with satirical remarks. Also, corrective humor might not only serve to correct transgressions, but it might also serve as a coping mechanism by venting one's feelings in making public humorous remarks about things that go wrong, independent of whether an improvement can actually be achieved or not. For the Russian context, existential freedom and implicit creative potential might be valued. Thus, there would be less need to correct rule breaking, as it would be considered a manifestation of free will, which might even arouse some sympathy. These hypotheses on cultural differences in benevolent and corrective humor should be systematically explored in future studies.

Age and Gender Differences

Going beyond cross-cultural comparisons, age and gender differences were explored. Although the differences found in these demographic variables were negligible or small, they still fitted well to the conceptualization of benevolent and corrective humor. Benevolent humor, especially item 9, showed linear increases with age. Item 9 ("Humor is suitable for arousing understanding and sympathy for imperfections and the human condition") might have had the strongest age effects for two reasons. First, it entails an attitude rather than showing humor directly. This is in line with findings that agreeableness increased with age, and extraversion and openness decreased with age (see Marsh et al., 2013). Specifically, the benevolent, serene, and accepting attitude underlying benevolent humor might increase, while making humorous remarks and enjoying humor in general might rather decrease in line with decreases in extraversion and openness (see Craik et al., 1996; Köhler and Ruch, 1996; Martin et al., 2003; Nusbaum et al., 2017). A second explanation takes into account the lack of scalar measurement invariance found for this item across age groups. Having different intercepts in the different age groups might lead to over- or underestimations of the means of specific groups, thus potentially reflecting bias instead of true mean differences (see Chen, 2008). For example, if older age groups had higher intercepts and younger age groups had lower intercepts than middle-aged adults, the means of the older groups might be overestimated and those of the younger groups underestimated.

For corrective humor, decreasing linear and quadratic trends were found. Thus, middle-aged adults engaged most often in this type of humor, followed by younger adults, with the lowest scores obtained for older adults. This developmental trajectory also fits to the increase in agreeableness and the decrease in extraversion and openness with age (Marsh et al., 2013), which would potentially explain the negative linear trend observed. The curvilinear trend was similar to the negative quadratic relationship of conscientiousness with age. Potentially, people who are more conscientious care more about what is right and wrong (i.e., they might have a stronger moral compass), which could potentially increase their levels of corrective humor. An alternative explanation could be that middle-aged adults are faced with situations in which they can employ corrective humor more often (e.g., at the workplace), and they might also believe that their humorous remarks can improve the conditions.

Regarding gender differences, men consistently scored higher in corrective humor than females, while only negligible gender differences were found for benevolent humor. This is consistent with other studies that found gender differences mostly for critical or affective forms of humor (such as sexual and aggressive humor; Martin et al., 2003; Lampert and Ervin-Tripp, 2007). By contrast, gender differences in the sense of humor and in humor as character strength (which was more strongly aligned to benevolent than to corrective humor; Ruch and Heintz, 2016) were usually small or negligible (Lampert and Ervin-Tripp, 2007; Heintz et al., 2017).

Limitations and Directions for Future Studies

The present study serves as a starting point for more extensive cross-cultural research and applications in the area of humor and particularly virtue-related forms of humor. However, several limitations can be noted. First, although the 25 samples allowed some cross-cultural comparisons, analyses at the sample level were limited due to the low statistical power. Thus, substantially increasing the number of samples is needed for additional comparisons, like correlating the samples' BenCor scores with other sample-specific indicators, such as culture dimensions (Hofstede, 2001), sample gelotophobia and character strengths scores (Proyer et al., 2009; McGrath, 2015), and broad personality traits (Schmitt et al., 2007). Additionally, employing more

samples would allow more detailed comparisons of samples from the same region vs. different regions (e.g., cities vs. rural environments, tribes of indigenous people) in the same country, from neighboring vs. adjacent countries, and from different language versions within the same country and across countries. This would help to disentangle the role of the local and national cultural norms and the influence of different languages (see Park et al., 2006; Proyer et al., 2009; McGrath, 2015) in determining similarities in the BenCor. For example, it was suggested that more collectivistic cultures, in comparison to more individualistic cultures, place higher importance on maintaining others' faces and thus rather avoid than dominate conflicts (Ting-Toomey et al., 1991). Thus, openly voicing criticism (whether humorously or not) might be less acceptable in collectivistic cultures such as China, Taiwan, and Japan, which would suggest that (a) the mean values of corrective humor would be lower, (b) corrective humor might be less seen as related to virtue, and consequently (c) the correlation between benevolent and corrective humor might be lower than in more individualistic cultures such as the United States. These hypotheses could be tested in future studies that systematically compare countries that differ in their collectivism and individualism scores.

Second, although the 12 marker items worked well in a majority of the samples, one could still think of slight adaptations that might shift them more strongly to the factor they belong to and that decrease the overlap between the two factors. For item 3, two changes are proposed, replacing "is aimed at" with "deals with" to make it less critical, and replacing "I include both myself and others" by "I refer to humans in general, including myself" (suggested rephrased item 3: "When my humor deals with human weaknesses, I refer to humans in general, including myself"). Item 8 could be simplified by replacing "caricature in a funny way" (which might be hard to understand or might be potentially misunderstood) by "making fun of", and by removing the term "gently" (suggested rephrased item 8: "I make fun of my fellow humans' wrongdoings to urge them to change"). Finally, item 12 could be made more corrective by replacing "poking fun" with "ridiculing" and by removing "hoping" ("If the circumstances are not as they actually should be, I ridicule these moral transgressions or societal wrongdoings to improve them in the long term"). The psychometric properties of these adapted marker items will be tested in future studies. If they are found to be superior to the existing marker items, these might be replaced in order to optimize the BenCor.

Third, the present study focused mainly on the psychometric properties of the BenCor and the need for separating the two concepts. Future studies can investigate their differential criterion validity in different countries. Thus far, only German-speaking countries have been investigated (Ruch and Heintz, 2016; Ruch et al., 2018a,b). For example, the BenCor could be related to different positive psychological variables such as subjective well-being (Diener et al., 2009), positive emotions (Shiota et al., 2017), and resilience (Masten et al., 2009) to establish the nomological network of benevolent and corrective humor. Replicating this nomological network in different countries would be an important task for future cross-cultural research on virtue-related humor. These studies could also include already established predictors of these outcomes (such as broad personality traits) as well as measures of the sense of humor and mockery to determine the incremental validity and unique contribution of the BenCor to the positive-psychological outcomes. Furthermore, gelotophobia (the fear of being laughed at) should be assessed as a control variable, as individuals with high scores have been shown to react less positively and more negatively to enjoyable emotions that elicit laughter (Platt et al., 2013; Ruch et al., 2015) and to have problems with intrapersonal emotion-related skills more generally (Papousek et al., 2009).

Fourth, in terms of age, the developmental trajectories of both benevolent and corrective humor deserve future studies to understand the underlying reasons for the age differences. Also, longitudinal investigations (for an overview, see Collins, 2006) would be needed to be able to distinguish among true developmental changes and cohort differences.

CONCLUSIONS

Overall, the present study supported the usefulness of the BenCor, a set of 12 marker items that assesses benevolent and corrective humor, for 22 different countries. This is especially remarkable as these historical concepts are rather complex and sophisticated, yet they could be recovered in different cultures and languages, allowing the accumulation of research findings across different cultures—at least the ones investigated so far. Thus, this study lays the foundations for closing the virtue gap in humor by providing an economic and reliable means of integrating benevolent and corrective humor in research across the world. Once the BenCor is sufficiently validated, it can fruitfully supplement existing humor applications in various areas, for example at the workplace (e.g., Robert, 2016), in clinical settings (e.g., Konradt et al., 2013), and in positive interventions (e.g., Wellenzohn et al., 2016a,b).

ETHICS STATEMENT

The studies were carried out in accordance with the recommendations of the local ethical guidelines of the committees of the following institutions: Catholic University in Ružomberok, HELP University, Indian Institute of Technology Delhi, Lebanese University, National Taiwan Normal University, Saint Petersburg State University, Universidad Andrés Bello, University of Granada, University of Latvia, Universiti Malaysia Terengganu, Universidad de Monterrey University of Rijeka University of Waikato, University of Wolverhampton, and University of Zurich. All participants provided either online or written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

WR and SH conceived the study and organized the data collection. SH conducted the data analyses and drafted the

manuscript. All authors were involved in the data collection and revisions of the manuscript.

FUNDING

AM-S thanks the Chilean Comisión Nacional de Investigación Científica y Tecnológica. His participation was funded by the Chilean Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt de Iniciación) Project no. 11160661.

ACKNOWLEDGMENTS

The authors would like to thank Jade Hooper, Mikhail Ivanov, and Veronika Sharok for their additional support in the data collection.

REFERENCES

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Model.* 14, 464–504. doi: 10.1080/10705510701301834

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *J. Pers. Soc. Psychol.* 95, 1005–1018. doi: 10.1037/a0013193

Cheung, G. W., and Rensvold, R. B. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *JOM* 25, 1–27. doi: 10.1177/014920639902500101

Collins, L. M. (2006). Analysis of longitudinal data: the integration of theoretical model, temporal design, and statistical model. *Annu. Rev. Psychol.* 57, 505–528. doi: 10.1146/annurev.psych.57.102904.190146

Craik, K. H., Lampert, M. D., and Nelson, A. J. (1996). Sense of humor and styles of everyday humorous conduct. *Humor* 9, 273–302. doi: 10.1515/humr.1996.9.3-4.273

Diener, E., Oishi, S., and Lucas, R. E. (2009). "Subjective well-being: the science of happiness and life satisfaction," in *Handbook of Positive Psychology, 2nd Edn.*, eds C. R. Snyder, and S. J. Lopez (New York, NY: Oxford University Press), 187–194.

Fornell, C., and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50. doi: 10.2307/3151312

Heintz, S., Kramm, C., and Ruch, W. (2017). A meta-analysis of gender differences in character strengths and age, nation, and measure as moderators. *J. Posit. Psychol.* doi: 10.1080/17439760.2017.1414297. [Epub ahead of print].

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations, 2nd Edn*. Thousand Oaks, CA: Sage publications.

Kenny, D. A., and McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling, *Struct. Equ. Model.* 10, 333–351. doi: 10.1207/S15328007SEM1003_1

Köhler, G., and Ruch, W. (1996). Sources of variance in current sense of humor inventories: how much substance, how much method variance? *Humor* 9, 363–397. doi: 10.1515/humr.1996.9.3-4.363

Konradt, B., Hirsch, R. D., Jonitz, M. F., and Junglas, K. (2013). Evaluation of a standardized humor group in a clinical setting: a feasibility study for older patients with depression. *Int. J. Geriatr. Psychiatry* 28, 850–857. doi: 10.1002/gps.3893

Lampert, M. D., and Ervin-Tripp, S. M. (2007). "Exploring paradigms: the study of gender and sense of humor near the end of the 20th century," in *The Sense of Humor: Explorations of a Personality Characteristic, 2nd Edn.*, ed W. Ruch (Berlin: Mouton de Gruyter), 231–270.

Lorenzo-Seva, U., and Ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology* 2, 57–64. doi: 10.1027/1614-2241.2.2.57

Marsh, H. W., Nagengast, B., and Morin, A. J. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Dev. Psychol.* 49, 1194–1218. doi: 10.1037/a0026913

Martin, R. A. (2007). *The Psychology of Humor: An Integrative Approach.* London: Academic Press.

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., and Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: development of the humor styles questionnaire. *J. Res. Pers.* 37, 48–75. doi: 10.1016/S0092-6566(02)00534-2

Masten, A. S., Cutuli, J. J., Herbers, J. E., and Reed, M. G. (2009). "Resilience in development," in *Handbook of Positive Psychology, 2nd Edn.*, eds. C. R. Snyder, and S. J. Lopez (New York, NY: Oxford University Press), 117–132.

McGhee, P. E. (2010). *Humor as Survival Training for a Stressed-Out World: The 7 Humor Habits Program*. Bloomington, IN: AuthorHouse.

McGrath, R. E. (2015). Character strengths in 75 nations: an update. *J. Posit. Psychol.* 10, 41–52. doi: 10.1080/17439760.2014.888580

McGrath, R. E. (2016). Measurement invariance in translations of the VIA inventory of strengths. *Eur. J. Psychol. Assess.* 32, 187–194. doi: 10.1027/1015-5759/a000248

Nicolson, H. (1946). *The English Sense of Humour*. London: Constable.

Nusbaum, E. C., Silvia, P. J., and Beaty, R. E. (2017). Ha ha? Assessing individual differences in humor production ability. *Psychol. Aesthet. Creat. Arts* 11, 231–241. doi: 10.1037/aca0000086

Papousek, I., Ruch, W., Freudenthaler, H. H., Kogler, E., Lang, B., and Schulter, G. (2009). Gelotophobia, emotion-related skills and responses to the affective states of others. *Pers. Individ. Dif.* 47, 58–63. doi: 10.1016/j.paid.2009.01.047

Park, N., Peterson, C., and Seligman, M. E. P. (2006). Character strengths in fifty-four nations and the fifty US states. *J. Posit. Psychol.* 1, 118–129. doi: 10.1080/17439760600619567

Peterson, C., and Seligman, M. E. P. (2004). *Character Strengths and Virtues: A Handbook and Classification*. New York, NY: Oxford University Press.

Platt, T., Hofmann, J., Ruch, W., and Proyer, R. T. (2013). Duchenne display responses towards sixteen enjoyable emotions: individual differences between no and fear of being laughed at. *Motiv. Emot.* 37, 776–786. doi: 10.1007/s11031-013-9342-9

Proyer, R. T., Ruch, W., Ali, N. S., Al-Olimat, H. S., Amemiya, T., Adal, T. A., et al. (2009). Breaking ground in cross-cultural research on the fear of being laughed at (gelotophobia): a multi-national study involving 73 countries. *Humor* 22, 253–279. doi: 10.1515/H.U.M.R.2009.012

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org/

Robert, C. (2016). *The Psychology of Humor at Work: A Psychological Perspective*. New York, NY: Routledge.

Rosseel, Y. (2012). lavaan: an r package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Ruch, W. (2012). "Towards a new structural model of the sense of humor: Preliminary findings," in *Proceedings of the AAAI Fall Symposium FS-12-02: Artificial Intelligence of Humor* (Menlo Park, CA: AAAI Press), 68–75.

Ruch, W., and Heintz, S. (2016). The virtue gap in humor: exploring benevolent and corrective humor. *Transl. Iss. Psychol. Sci.* 2, 35–45. doi: 10.1037/tps0000063

Ruch, W., Heintz, S., Platt, T., Wagner, L., and Proyer, R. T. (2018a). Broadening humor: comic styles differentially tap into temperament, character, and ability. Front. Psychol. 9:6. doi: 10.3389/fpsyg.2018.00006

Ruch, W., Hofmann, J., and Platt, T. (2015). Individual differences in gelotophobia and responses to laughter-eliciting emotions. *Pers. Individ. Dif.* 72, 117–121. doi: 10.1016/j.paid.2014.08.034

Ruch, W., Wagner, L., and Heintz, S. (2018b). Humor, the PEN model of personality, and subjective well-being: support for differential relationships with eight comic styles. *RISU* 1, 31–44. Available online at: https://www.risu.biz/wp-content/uploads/2018/01/Ruch_et_al.-RISU-11-2018-31-44-1.pdf

Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res.* 8, 23–74. Available online at: http://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130_13.pdf

Schmidt-Hidding, W. (1963). *Europäische Schlüsselwörter. Band I: Humor und Witz [European Keywords.* Vol. 1, *Humor and Wit].* Munich: Huber.

Schmitt, D. P., Allik, J., McCrae, R. R., and Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: patterns and profiles of human self-description across 56 nations. *J. Cross Cult. Psychol.* 38, 173–212. doi: 10.1177/0022022106297299

semTools Contributors (2015). *semTools: Useful Tools for Structural Equation Modeling.* R package version 0.4-9. Available online at: http://cran.r-project.org/package=semTools

Shiota, M. N., Campos, B., Oveis, C., Hertenstein, M. J., Simon-Thomas, E., and Keltner, D. (2017). Beyond happiness: building a science of discrete positive emotions. *Am. Psychol.* 72, 617–643. doi: 10.1037/a0040456

Stark, S., Chernyshenko, O. S., and Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. *J. Appl. Psychol.* 91, 1292–1306. doi: 10.1037/0021-9010.91.6.1292

Ting-Toomey, S., Gao, G., Trubisky, P., Yang, Z., Soo Kim, H., Lin, S. L., et al. (1991). Culture, face maintenance, and styles of handling interpersonal conflict: a study in five cultures. *Int. J. Confl. Manage.* 2, 275–296. doi: 10.1108/eb022702

Van de Vijver, F., and Hambleton, R. K. (1996). Translating tests. *Eur. Psychol.* 1, 89–99. doi: 10.1027/1016-9040.1.2.89

Wellenzohn, S., Proyer, R. T., and Ruch, W. (2016a). Humor-based online positive psychology interventions: a randomized placebo-controlled long-term trial. *J. Posit. Psychol.* 11, 584–594. doi: 10.1080/17439760.2015.1137624

Wellenzohn, S., Proyer, R. T., and Ruch, W. (2016b). How do positive psychology interventions work? A short-term placebo-controlled humor-based study on the role of the time focus. *Pers. Individ. Dif.* 96, 1–6. doi: 10.1016/j.paid.2016.02.056

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Heintz, Ruch, Platt, Pang, Carretero-Dios, Dionigi, Argüello Gutiérrez, Brdar, Brzozowska, Chen, Chłopicki, Collins, Ďurka, Yahfoufi, Quiroga-Garza, Isler, Mendiburo-Seguel, Ramis, Saglam, Shcherbakova, Singh, Stokenberga, Wong and Torres-Marín. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 27 February 2018
doi: 10.3389/fpsyg.2018.00199

Is an Ideal Sense of Humor Gendered? A Cross-National Study

*Sümeyra Tosun[1], Nafiseh Faghihi[2] and Jyotsna Vaid[2]**

[1] Department of Psychology, University of Pretoria, Pretoria, South Africa, [2] Department of Psychological and Brain Sciences, Texas A&M University, College Station, TX, United States

To explore lay conceptions of characteristics of an ideal sense of humor as embodied in a known individual, our study examined elicited written narratives by male and female participants from three different countries of origin: United States, Iran, and Turkey. As reported in an earlier previous study with United States-based participants (Crawford and Gressley, 1991), our study also found that the embodiment of an ideal sense of humor was predominantly a male figure. This effect was more pronounced for male than for female participants but did not differ by country. Relative mention of specific humor characteristics differed by participant gender and by country of origin. Whereas all groups mentioned creativity most often as a component of an ideal sense of humor, this attribute was mentioned significantly more often by Americans than by the other two groups; hostility/sarcasm was also mentioned significantly more often by Americans than Turkish participants who mentioned it more often than Iranian participants. Caring was mentioned significantly more often by Americans and Iranians than by Turkish participants. These findings show a shared pattern of humor characteristics by gender but group differences in the relative prominence given to specific humor characteristics. Further work is needed to corroborate the group differences observed and to pinpoint their source.

Keywords: sense of humor, ideal humor, everyday humor, gender, culture, creativity, sarcasm

OPEN ACCESS

Edited by:
Tracey Platt,
University of Wolverhampton,
United Kingdom

Reviewed by:
Jill Ann Jacobson,
Queen's University, Canada
Kuba Krys,
Institute of Psychology (PAN), Poland

****Correspondence:***
Jyotsna Vaid
jvaid@tamu.edu

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *01 October 2017*
Accepted: *05 February 2018*
Published: *27 February 2018*

Citation:
Tosun S, Faghihi N and Vaid J (2018) Is an Ideal Sense of Humor Gendered? A Cross-National Study. Front. Psychol. 9:199. doi: 10.3389/fpsyg.2018.00199

INTRODUCTION

There is an established literature on gender differences in humor perception and humor styles. Men have been noted to prefer humor that has sexual or aggressive themes whereas women appear to prefer neutral or absurd humor (Aillaud and Piolat, 2012). Whereas earlier studies showed that sexist humor (i.e., humor that upholds gender role stereotypes) is preferred over non-sexist humor (Cantor, 1976), other studies report that both men and women prefer humor that has the opposite gender as the butt (Vaid and Hull, 1998; Parekh, 1999). Furthermore, men typically rate themselves higher than women in humor initiation whereas women tend to rate themselves higher in humor appreciation, but when humor is studied in actual conversational contexts a more nuanced picture emerges (see Kramarae, 1981; Kotthoff, 1996, 2000; Schiau, 2017). Similarly, whereas some studies have found that humor produced by men is judged to be more humorous than that produced by women (Brodzinsky and Rubien, 1976), other studies have not found this effect (Hull et al., 2017), and still other work suggests a bias operating, whereby men are perceived to be the "funnier sex" regardless of how their humorous creations are actually judged (Mickes et al., 2011; Hooper et al., 2016).

Taken as a whole, the literature on gender and humor eludes easy generalization (see Martin, 2007, for a review). Methodologically, early studies have been criticized for their use of decontextualized or "canned" humor samples instead of spontaneously generated humor that arises naturally in conversation (Frecknall, 1994; Lampert and Ervin-Tripp, 1998). The literature on gender differences in humor preferences has also been critiqued for its reliance on classifications of humor type (as "hostile" or "sexual" or "sexist") based on experimenter intuitions rather than eliciting participants' own perceptions of humor type, which may not coincide with those of the experimenter (e.g., Parekh, 1999). Moreover, studies that have sought to measure the construct of a sense of humor have led to many promising instruments, such as the three Witz-Dimensionen (3WD) instrument by Ruch (1992), and to new adaptations of established instruments for use with non-English speakers (see Özdoğru, 2017). At the same time, it is recognized that for a construct as slippery and contextual as humor, it is important to consider multiple, converging measures across different groups and settings.

This recognition of the complexity of studying humor, together with a growing shift in regarding gender as performative, has led to a shift in humor scholarship in the direction of studying humor as it is enacted by men and women in a range of social contexts (e.g., Hay, 2000; Crawford, 2003), and in a range of laboratory contexts. Our own previous work has explored the relationship between cognitive, neurocognitive, and psycholinguistic aspects of humor detection and comprehension (e.g., Vaid, 2000; Vaid and Kobler, 2000; Vaid et al., 2003, 2015; Hull et al., 2005; Lopez and Vaid, 2017). Our work on humor production has sought to develop controlled ways of eliciting humor to study its cognitive and social underpinnings. For example, we developed a concept comparison task in which participants were asked to produce "catchy" ways in which the concepts were related, which invariably elicited humorous responses, e.g., MONEY and CHOCOLATE: *one swells the wallet, the other, the hips* (Hull et al., 2017). Another task involved generating rejoinders to proverbs, e.g., *Absence makes the heart grow fonder, but also makes the eyes wander* (Vaid, 2014). Other prior work in our laboratory has examined the role of culture in judgments about when humor (vs. silence) is an appropriate response to embarrassing situations encountered in daily life (Vaid et al., 2008). Finally, we have examined how individuals' perceptions of their own humor styles compare with their perceptions of humor styles of members of their gender category and/or same or different cultural group (Quiros and Vaid, 1998; Vaid, 1999, 2006).

As an extension of our interest in gender and cultural dimensions of humor, the aim of the present research was to characterize how gender and country of origin (as a proxy for culture) may shape how individuals conceptualize an ideal sense of humor. The motivation for this study was a previous study which examined the role of gender in lay conceptions of an ideal sense of humor (Crawford and Gressley, 1991) in a large sample of United States-based participants of different ages and backgrounds. Participants in this study were asked to provide a brief narrative describing the humor characteristics of a person they knew who embodied an outstanding sense of humor. Crawford and Gressley (1991) reported that a majority of the participants identified a male figure as the person who embodied an outstanding sense of humor. Indeed, of the 141 respondents (49 men, 92 women), nearly 84% of men and 67% of women selected a male figure. The researchers also classified the humor characteristics mentioned into five categories: creativity (witty, clever, quick comeback), caring (humor used to put others at ease), real life (grounding the humor in real life experiences), jokes (having a repertoire of jokes), and hostility/sarcasm (satirical, biting humor) and noted that creativity, caring and real life were mentioned most often, and that there were no discernible differences in the weighting of these characteristics as a function of either participant gender or target gender.

Over 25 years have passed since the Crawford and Gressley (1991) study. While gender continues to be a salient element structuring society, women have also become more visible in a number of domains of public life, including in the realm of comedy. It is possible that gender stereotypes may have become less entrenched in the present day. We therefore wondered if the preference for a male figure as the embodiment of an outstanding sense of humor noted previously still holds among young adults in the present age. We also wondered whether individuals from other countries would show a similar preference, given that they might be less likely to be influenced by Western gender stereotypes (including stereotypes regarding men as being the canonical humor initiator), but might have their own cultural stereotypes about humor, gender, and the relation between the two. Although there have been a few prior studies of humor stereotypes in different nationalities, the focus of our study was on how individuals from the United States compared to those from two other countries in articulating characteristics of an ideal sense of humor, as embodied in someone they knew. Our interest was to uncover patterns of commonalities as well as differences across groups and across genders.

In searching the literature, we could find only one other empirical study conducted since the study by Crawford and Gressley (1991) that used their open-ended prompt. This study, by Nevo et al. (2001), was conducted on men and women in Singapore. It, too, found that the embodiment of an outstanding sense of humor was male. Of the 18 men and 46 women in the study, 76% of respondents selected a male target (Nevo et al., 2001). The researchers further noted that the preference for a male target was more pronounced in men, but no additional analyses were reported in terms of specific humor characteristics mentioned by men and women. Thus, we felt another study was warranted.

The Present Research

Our study had two goals. The first was to investigate if the male preference first reported by Crawford and Gressley (1991) still holds. To examine this, we pooled data from United States-based college students tested from 2004 to the present. The second goal was to investigate if the pattern of a male preference as the embodiment of an ideal sense of humor is restricted to United States participants or is generalizable to other samples. In particular, we considered samples drawn

from Iran and from Turkey, as these particular groups have been understudied in the humor literature; where country-based differences have been studied, they have either tended to be within north American/European samples or have compared Western with east Asian samples (e.g., China). Turkey is considered geographically and culturally as a bridge between Asia and Europe. Thus, we aimed to compare participants raised in a Western (American), a Middle Eastern (Iranian), and a blended (Turkish) culture. We did not have *a priori* expectations of how participants across the three groups would respond on the task; our study is exploratory with regard to the cultural dimension, as our sample sizes were limited and varied in other respects (e.g., age) and we recognize that much more follow up investigation would be needed to fully understand the nature of any differential patterns uncovered.

MATERIALS AND METHODS

Participants

Participants included male and female United States born (American) and international students (born in Iran or in Turkey) recruited from a university town in the southwestern region of the United States, from a university in Istanbul, and from online responses. The American sample consisted of 279 undergraduate students (including 201 women) who ranged in age from 18 to 23 years, with a mean of 21 years. The majority self-identified as white, and the numbers of Latinx, African American, or Asian Americans were too few to permit separate subgroup analyses. The Iranian sample comprised 71 participants (47 women) who ranged in age from 17 to 54 years, with a mean age of 31.32 years, and the Turkish sample consisted of 79 undergraduate students (48 women) ranging in age from 18 to 25 years with a mean of 21.7. The American and Turkish participants completed the task as part of a class activity; the Iranian sample was recruited by placing an announcement in social media and participants completed an online version of the task. All participants received and answered the prompt in their primary language. The Iranian and Turkish data were translated into English by native Farsi- and Turkish speakers who had advanced English proficiency. Most of the data were coded by the same researcher (with gender of participants masked) to provide consistency in coding. A subset of the data were also intercoded to ensure some level of consensus (at least 80%).

Materials and Procedure

Participants were given a response sheet on which they were to write a brief narrative in response to an open-ended prompt adapted from Crawford and Gressley's (1991) study. They were instructed to think of a specific individual they knew who had an outstanding or ideal sense of humor and then to describe the characteristics of that humor, using three to five descriptors. They were then asked to describe the person who embodied that humor (we refer to this person as the *humor target*), indicating, for example, whether it was a family member, a friend, co-worker, or a comedian, and/or noting their gender, age, and ethnicity. There were no time constraints for responding. Since not all respondents stated the humor target's gender, this information sometimes had to be inferred from the stated relation to the target (e.g., brother, sister, girlfriend, particular celebrity, etc.) or from the participants' choice of pronouns in describing the person (however, this approach was helpful only for the English dataset as pronouns in Farsi and Turkish are not marked for gender).

Data Analyses

Two sets of comparisons were conducted using chi square and regression analyses. The first examined percent mention of the target gender by participant gender and country. The second examined percent mention of each of the five categories of humor descriptors identified by Crawford and Gressley (1991) in relation to participant gender and country.

The five coding categories were as follows: **Creativity:** This characteristic includes terms referring to creative aspects of humor, like *witty, quick comeback, playing with language, clever,* as well as being spontaneous or natural. An example of this characteristic from our sample is "very quick in answering with a witty comment." **Caring:** This characteristic indicates the kind of humor that makes people laugh and helps to change their mood when they are upset or in a tough situation. An example of this characteristic is "their humor helps relieve the tension." **Real Life:** This characteristic shows the ability of the humorous person to tell stories and recount real life events in a humorous way. An example of this dimension is "a great story-teller to bring out humor." **Jokes:** This characteristic refers to the use of actual jokes. An example of this dimension is "holds the crowd's attention with a simple joke." **Hostility/Sarcasm:** This category consists of attacking, insulting, and destructive humor as well as sarcasm. An indication of this characteristic is "can come up with the worst sexist insult."

RESULTS

A summary of the relative distribution of target gender of the ideal humor person is provided in **Table 1** by participant gender and country of origin. Also included in the table are the number of participants per group for whom humor target gender was not

TABLE 1 | Gender distribution of humor target per participant gender and country.

Country of origin	Participant gender	Target gender		
		Female	Male	Unspecified
American	Female	45	118	38
	Male	5	50	23
	Total	50	168	61
Turkish	Female	15	33	0
	Male	2	29	0
	Total	17	62	0
Iranian	Female	6	13	28
	Male	3	7	14
	Total	9	20	42

specified. The latter comprised 21.81% of the American sample, none of the Turkish sample, and 59.15% of the Iranian sample.

Identified Target Gender by Country of Origin

A chi square analysis was done excluding those whose target gender was unspecified to compare the relative percent mention of a male vs. female humor target, collapsed across participant gender. The analysis showed no significant effect of country of origin, $\chi^2 = 0.33$, $p = 0.85$, $N = 331$. That is, regardless of their country of origin, participants showed a consistent tendency to select a male figure as their humor ideal: 77.1% of Americans, 78.5% of Turkish, and 73.5% of Iranian participants identified a male.

Identified Target Gender by Country of Origin (American vs. Turkish vs. Iranian) and Participant Gender

A logistic regression was conducted to see if the gender of the ideal humor target person could be predicted based on the participants' gender or the participant's country of origin (American, Turkish, Iranian). Again, only participants whose responses indicated the gender of their humor ideal were included in the analysis. Dummy coding was applied for the analysis. The model was significant, $\chi^2 = 13.78$, $p = 0.003$, $df = 3$ and explained 6.2% of the variance. Gender of participant was a significant predictor of gender of humor target ($\chi^2 = 11.08$, $p = 0.001$, *odds ratio* $= 0.29$): male participants were more likely than female participants to select a male target as the embodiment of an ideal sense of humor (89.6% vs. 71.9%, respectively). Country of origin, on the other hand, was not a significant predictor ($\chi^2 = 0.32$, $p = 0.85$) (Turkish vs. Iranian: $\chi^2 = 0.13$, $p = 0.71$, *odds ratio* $= 1.19$; American vs. Iranian: $\chi^2 = 0.31$, $p = 0.57$; *odds ratio* $= 0.94$; American vs. Turkish: $\chi^2 = 0.04$, $p = 0.84$, *odds* ratio $= 0.79$). See **Figure 1** for a depiction of the percent mention of male targets per participant gender and group.

Identified Target Gender by Time Period and Participant Gender – American Sample Only

A logistic regression was conducted on the American sample to see if there was a difference related to time at testing in the percent mention of a male target by men and women. Here, Crawford and Gressley (1991) were compared with data from the American sample (which was collected over two different time periods, 2004 and 2014).

The model was significant, $\chi^2 = 14.97$, $p = 0.002$, $df = 3$ and explained 6.9% of variance. There was not a difference between the American 2014 and the 1991 data. However, the American 2004 data showed a difference than both the 1991 data, $\chi^2 = 6.33$, $p = 0.012$, $B = -1.12$ and the 2014 data, $\chi^2 = 4.88$, $p = 0.027$, $B = -1.01$. Participants from the 2004 sample (89.6%) revealed more male favored results than the 2014 sample (73%) and than the original study sample (73%).

Participants' gender was also a significant predictor, $\chi^2 = 5.29$, $p = 0.021$, $B = -0.767$. That is, the selection of a male humor ideal was significantly higher when the participant was a male than when the participant was a female. In the original study male participants' preference for a male target was 83.7% and female participants' preference for a male target was 67.4%. In our study, male preference for a male target was 90.9% while female preference for a male target was 72.4%.

Analyses of Ascribed Humor Characteristics

An additional set of analyses was conducted on the influence of participant gender on relative mention of each of five

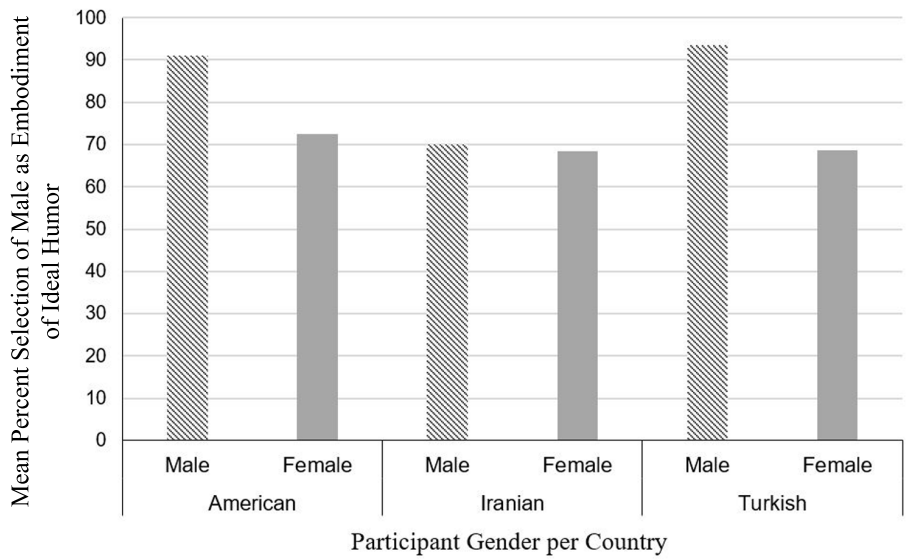


FIGURE 1 | This figure demonstrates the relative mention of man humor target per participant gender and country (*only participants who specified the target gender*).

characteristics of an ideal sense of humor. (A preliminary analysis that included target gender as an additional predictor yielded no effect of this variable and so we do not report it here.) **Table 2** provides a summary of the relative mention of each characteristic by male and female participants in each of the three groups. Note that these values represent all of the data per group, including those for whom target gender was not specified.

Humor Characteristics by Participant Gender and Country of Origin

Inspection of the relative percent mention of the five humor characteristics shows an overall predominance of mention of the creativity characteristic by men and women and across all groups. For Americans, the next most mentioned characteristic was hostility/sarcasm, followed by caring. For Iranians, by contrast, the order of mention of the five characteristics was: creativity, real life and caring, and for the Turkish sample, the order was creativity, joke and hostility/sarcasm (see **Figure 2**).

A multivariate regression analysis was conducted to jointly examine the effect of participant gender and country of origin (American, Turkish, and Iranian) with each of the five humor characteristics (creativity, caring, real life, jokes, and hostility/sarcasm) considered as separate dependent variables. The results demonstrated that overall, both participant gender ($\chi^2 = 2.67$, $p = 0.022$, $df = 5$) and country of origin ($\chi^2 = 22.15$, $p < 0.001$, $df = 5$) were significant predictors. However, no gender-specific effect was observed in any of the five humor characteristics. The results demonstrated that gender was a multivariate phenomenon, but gender did not specifically predict any of the five characteristics. Country of origin was a significant predictor for creativity ($\chi^2 = 10.30$, $p = 0.001$, $R^2 = 0.03$, *odds ratio* $= 0.905$), caring ($\chi^2 = 13.39$, $p < 0.001$, $R^2 = 0.04$, *odds ratio* $= 0.902$), and hostility ($\chi^2 = 54.85$, $p < 0.001$, $R^2 = 0.12$, *odds ratio* $= 0.816$). Creativity was mentioned significantly more by American participants (64%) than by Turkish (48%) or Iranian participants (45%). Moreover, American (37%) and Iranian (34%) participants used caring to describe their ideal humor significantly more than did Turkish participants (2.5%). Further, hostility was mentioned significantly more by American participants (41%) followed by Turkish participants (19%) and it was mentioned least by Iranians (nearly 0%).

TABLE 2 | Relative mention of each characteristic by each gender participant in each of the three groups.

Characteristic	Country	Male	Female
Creativity	Americans	56.4	66.7
	Iranians	45.8	44.7
	Turkish	45.2	50
Caring	Americans	32.1	38.3
	Iranians	12.5	29.8
	Turkish	3.2	2.1
Real life	Americans	21.8	32.3
	Iranians	12.5	27.7
	Turkish	9.7	2.1
Jokes	Americans	15.4	17.9
	Iranians	8.3	10.6
	Turkish	29	20.8
Hostility	Americans	35.9	42.8
	Iranians	0	0
	Turkish	12.9	22.9

Numbers show percent mention.

DISCUSSION

The aim of this study was to examine how men and women describe a specific person who embodies their ideal sense of humor. The study provided an opportunity to test whether the finding of a male target preference first noted by Crawford and Gressley (1991) for United States based participants and by Nevo et al. (2001) for Singaporean participants persists for Americans in the present period and is evident to the same extent among two other groups whose cultures are considered to be somewhat more traditional in terms of gender role stereotypes than American culture. Our findings show that the selection of a male as the embodiment of an ideal sense of humor was a pervasive and robust finding across the three samples we tested. Moreover, the size of this effect did not vary across the three groups. Of course, it is possible that the three samples we selected are on the gender inegalitarian end of the continuum and that had we selected a more egalitarian country we might not have found the effect. That remains for future work to test.

Our analysis of the United States samples tested at different periods of time further revealed that a male preference was actually somewhat stronger in the 2004 sample than it was for either the 1991 sample or a more recent 2014 sample. Perhaps the stronger male bias exhibited in the 2004 sample is a reflection of a public discourse in the country around that time regarding whether women can ever be as good at comedy as men. Nevertheless, it is interesting to note that, across all time periods sampled, the selection of a male target was significantly more likely when the participant was himself male. Thus, despite changes in societal consciousness about gender and humor that may have occurred (to differing degrees) over the past 25 years, there is a consistent preference for men to consider men as the embodiment of an ideal sense of humor. Moreover, this effect was found in the analysis by country of origin as well.

The finding that men are perceived as the embodiment of an ideal sense of humor may in part reflect an availability bias arising from the fact that male comedians and comedy writers still greatly outnumber female comedians and comedy writers. This difference in base rate may thus perpetuate a gender stereotype of men as the funnier sex and therefore prime people to think of men (rather than women) among their own acquaintances who exemplify an ideal sense of humor. Incidentally, among the American participants who provided information on their relationship to the gender target, a sizeable number (males and females) mentioned that the ideal humor person was their father. Further work should examine target gender demographic characteristics to provide insights into their relationship, if any, to the humor characteristics they embody.

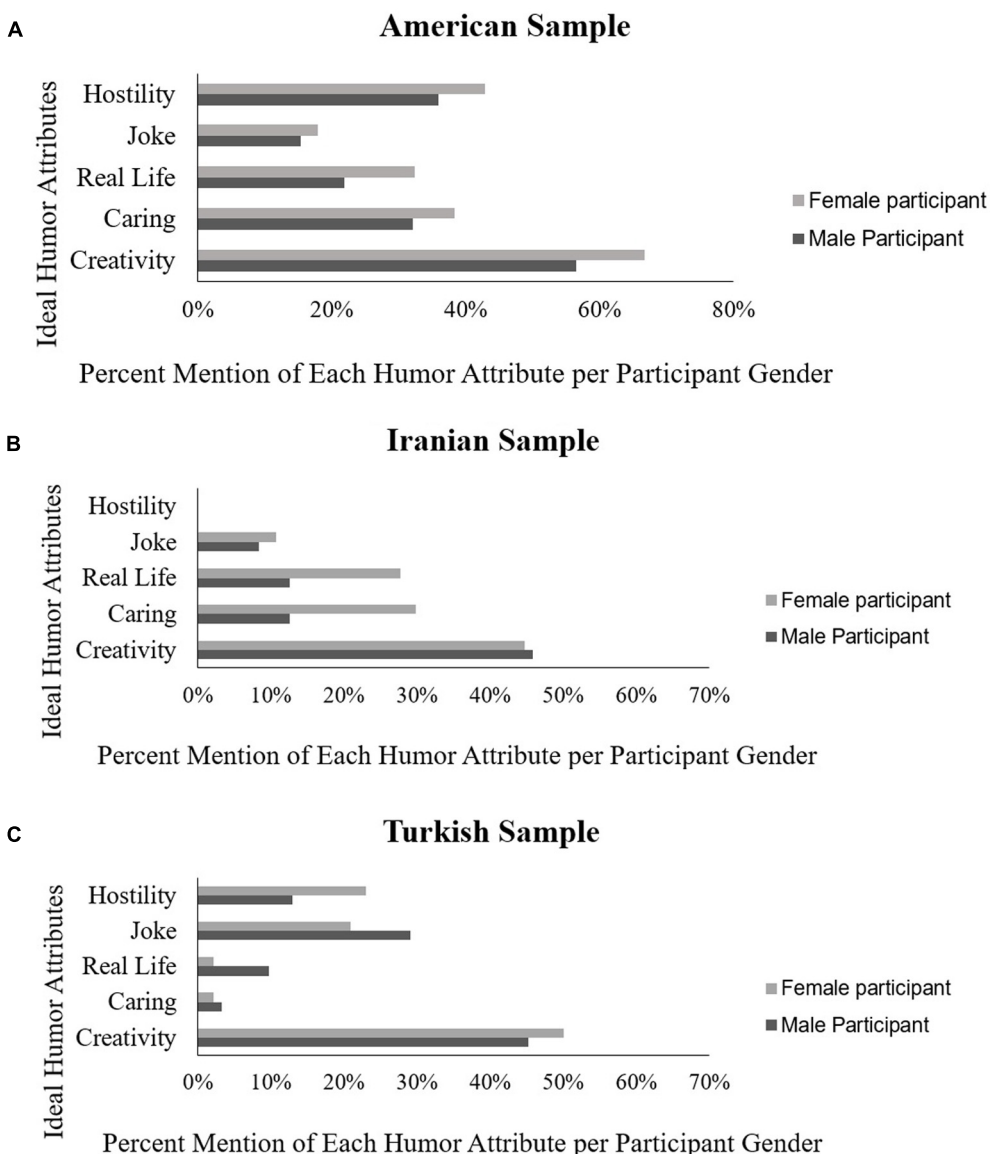


FIGURE 2 | This figure demonstrates the relative mention of each characteristic by participant gender and country. **(A)** American sample; **(B)** Iranian sample; **(C)** Turkish sample.

Is there a difference in the types of characteristics used by men and women for male vs. female humor targets? Our analysis of the five dimensions noted by Crawford and Gressley (1991) to describe an ideal humor showed that the most frequently mentioned attribute by Americans in our sample is creativity, defined here as being witty, clever, and quick in coming up with a response. Creativity was mentioned by the majority of participants of both genders. The next most frequently mentioned dimensions for the American sample were hostility/sarcasm and caring. This may seem like an odd juxtaposition at first sight but it may not be that surprising given that the characteristic of "sarcasm" was coded under "hostility" and sarcasm (in American culture) is a way of interacting with one's friends. Although the Turkish and Iranian samples also chose creativity most often as a defining characteristic of an ideal sense of humor, they differed from each other and from the American sample in other characteristics: caring was mentioned by the Iranian and the American samples to the same extent but was mentioned hardly at all by the Turkish participants. By contrast, hostility/sarcasm was mentioned hardly at all by the Iranian sample. We do not wish to over-interpret the particular group differences obtained, as we did not have *a priori* expectations. We present them here as descriptive data, in need of further exploration.

Our findings corroborate the overall pattern noted in the previous study by Crawford and Gressley (1991) on which the present research was based – namely, a preference for a male figure as the embodiment of an ideal sense of humor. However,

the pattern of mention of the five different humor characteristics of the embodiment of an ideal sense of humor does not entirely concur with the pattern noted by Crawford and Gressley (1991). As already noted, there were some clear differences across the three cultural groups in the relative frequency of mention of some of the five humor characteristics.

Moreover, we recognize that our analysis of humor characteristics may also have been influenced in a substantive way by whether the target they were thinking of was male or female. As brought up by one of the reviewers of our article, it is possible that people tend to interpret or remember a given behavior differently depending on whether it comes from a man or a woman. Because of stereotypes or violation of expectations, a joke made by a woman could be interpreted as mean whereas the same joke made by a man could be interpreted as funny. Alternatively, participants could be selectively remembering humorous statements that conform to their gendered expectations of women as being caring, and thus describe the ideal humor for a female target as more caring. Unfortunately, given the paucity of female targets in our findings, we could not analyze humor characteristics as a function of target gender, but this is an important issue for further work.

Furthermore, we confined our analysis to only five humor categories based on those used in the previous study, a number of responses mentioned by participants in our study were not easily classifiable within that coding scheme. For example, a number of participants referred to "being able to laugh at themselves/take a joke" as being a valued characteristic in the person with an ideal sense of humor. Similarly, a number of respondents emphasized that the humor of this person was "inappropriate" but "not mean-spirited." Or that it involved impersonation, funny facial expressions, etc. Based on these types of responses, it would be important in further work to do a more detailed content analysis than was possible in the present study. A need for additional investigation of this issue is underscored by our finding that a third of the Iranian sample's narratives and from 5 to 10% of the American and Turkish narratives do not refer to any of the five characteristics of humor analyzed in this paper, and instead mention other characteristics of ideal humor that were not captured by these five categories. The use of political humor in everyday discourse, humor addressing tensions between the traditional vs. the modern (e.g., Apaydin, 2005), or individual differences in the relationship between humor and psychological well-being (e.g., Martin et al., 2003) would be interesting to explore in further work. Other approaches to measuring everyday humor (e.g., Craik et al., 1998) are also worth incorporating in future investigations of the role of gender and culture in the perception of an ideal sense of humor (see also Warren and McGraw, 2016).

Moreover, in addition to looking at humor type, it would be important to consider individuals' age as another factor that will likely influence what is considered desirable in a sense of humor. As people grow older, they might prefer caring or wise humor over hostile humor. In this regard, it may be relevant to point out that the Iranian sample – which showed practically no mention of hostility/sarcasm – had a broader age range among participants than did the other two samples.

A further limitation of our study is that while the prompt was intentionally made open-ended, this made for some messiness in coding, as there were some terms, e.g, *inappropriate, crude or vulgar,* that we treated as interchangeable for the purpose of our coding, but which may not have been perceived by participants as synonyms. Similarly, a number of participants used *sarcastic* as an attribute, and in the present coding scheme it was coded as hostility, but it might not have been considered by participants to be a negative attribute. Of course, it is possible that an ideal sense of humor is perceived to include darker elements in addition to positive elements.

Another limitation of this study is that we classified participants and targets solely on the basis of their assigned gender, and thus cannot say anything about perceptions of humor by and about individuals whose gender assigned at birth does not coincide with their gender at the time of testing, or who consider themselves non-binary with regard to gender. Relatedly, since we did not administer any measures to assess participants' gender identity or gender role attitudes, our data do not allow us to say anything about how participants' attitudes toward feminism or traditional gender roles may have informed their responses. Similarly, a gendered humorous persona might not be considered appealing to some participants in this "gender fluid" era.

These limitations notwithstanding, our study, using an implicit, elicited measure of lay conceptions of an ideal sense of humor, allows us to conclude that, in highly gender inegalitarian societies, the ideal sense of humor is strongly gendered in favor of male targets, especially among men. Further, this male preference is as firmly entrenched in contemporary American culture (at least for the young adult age range sampled in the present research) now as it was nearly 25 years ago. Third, a bias for thinking of men as the embodiment of an ideal sense of humor is not restricted to those in an American cultural context but is also found among members of two other nationalities. What is important to note is that these groups, despite being considered more "traditional," nevertheless did not show a stronger gender effect than that observed in the American sample. Finally, our results indicate that across all groups the most salient dimension of an ideal sense of humor is the ability to be witty, creative, and quick; this dimension is most pronounced for the American sample, who also appeared to value hostility/sarcasm and caring. The Iranian sample in turn appeared to value real life humor and caring, whereas the Turkish group placed least value on caring but instead emphasized jokes and hostility/sarcasm.

In further work it will be important to probe deeper into the social context of humor use in everyday life to determine not just what the characteristics are of an ideal sense of humor in an abstract sense, but how those characteristics are brought to life in different kinds of interactions.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Institutional Review Board guidelines with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration

of Helsinki. The protocol was approved by the Texas A&M University Institutional Review Board Committee.

AUTHOR CONTRIBUTIONS

JV and ST contributed to the design of experimental stimuli and procedure. All authors contributed to the manuscript in terms of conception, implementation of experimental protocols, data collection and analysis, and editing of drafts.

ACKNOWLEDGMENTS

We thank Erica Dittmer, Paige Dusthimer, and Omar Garcia for assistance in data collection.

REFERENCES

Aillaud, M., and Piolat, A. (2012). Influence of gender on judgment of dark and nondark humor. *Individ. Differ. Res.* 10, 211–222.

Apaydin, G. E. (2005). Modernity as masquerade: representations of modernity and identity in Turkish humour magazines. *Identities Glob. Stud. Cult. Power* 12, 107–142. doi: 10.1080/10702890590914339

Brodzinsky, D., and Rubien, J. (1976). Humor production as a function of sex of subject, creativity, and cartoon content. *J. Consult. Clin. Psychol.* 44, 597–600. doi: 10.1037/0022-006X.44.4.597

Cantor, J. R. (1976). What is funny to whom? The role of gender. *J. Commun.* 26, 164–172. doi: 10.1111/j.1460-2466.1976.tb01920.x

Craik, K., Ware, P., and Ruch, W. (eds). (1998). *Humor and Personality in Everyday Life*. Berlin: Walter de Gruyter, 63–94. doi: 10.1515/9783110804607.63

Crawford, M. (2003). Gender and humor in social context. *J. Pragmat.* 35, 1413–1430. doi: 10.1016/S0378-2166(02)00183-2

Crawford, M., and Gressley, D. (1991). Creativity, caring, and context: women's and men's accounts of humor preferences and practices. *Psychol. Women Q.* 15, 217–231. doi: 10.1111/j.1471-6402.1991.tb00793.x

Frecknall, P. (1994). Good humor: a qualitative study of the uses of humor in everyday life. *Psychol. J. Hum. Behav.* 31, 12–21.

Hay, J. (2000). Functions of humor in the conversations of men and women. *J. Pragmat.* 32, 709–742. doi: 10.1016/S0378-2166(99)00069-7

Hooper, J., Sharpe, D., and Roberts, S. G. B. (2016). Are men funnier than women or do we just think they are? *Transl. Issues Psychol. Sci.* 2, 54–62. doi: 10.1037/tps0000064

Hull, R., Chen, H. C., Vaid, J., and Martinez, F. (2005). Great expectations: humor comprehension across hemispheres. *Brain Cogn.* 57, 281–282.

Hull, R., Tosun, S., and Vaid, J. (2017). What's so funny? Modelling incongruity in humor production. *Cogn. Emot.* 31, 484–499. doi: 10.1080/02699931.2015.1129314

Kotthoff, H. (1996). Impoliteness and conversational joking: on relational politics. *Folia Linguist.* 30, 299–324. doi: 10.1515/flin.1996.30.3-4.299

Kotthoff, H. (2000). Gender and joking: on the complexities of women's image politics in humorous narratives. *J. Pragmat.* 32, 55–80. doi: 10.1016/S0378-2166(99)00031-4

Kramarae, C. (1981). *Women and Men Speaking: Frameworks for Analysis*. New York, NY: Newbury House Publishers.

Lampert, M. D., and Ervin-Tripp, S. M. (1998). "Exploring paradigms: the study of gender and sense of humor near the end of the 20th century," in *The Sense of Humor: Explorations of a Personality Characteristic*, ed. W. Ruch (Berlin: Mouton de Gruyter), 231–270.

Lopez, B. G., and Vaid, J. (2017). "Psycholinguistic approaches to humor," in *The Routledge Handbook of Language and Humor*, ed. S. Attardo (New York, NY: Routledge), 267–281.

Martin, R. (2007). *The Psychology of Humour: An Integrative Approach*. Amsterdam: Elsevier Academic Press.

Martin, R., Puhlik-Doris, P., Larsen, G., Gray, J., and Weir, R. (2003). Individual differences in uses of humor and their relation to psychological well-being: development of the humor styles questionnaire. *J. Res. Pers.* 37, 48–75. doi: 10.1016/S0092-6566(02)00534-2

Mickes, L., Walker, D. E., Parris, J. L., Mankoff, R., and Christenfeld, N. J. (2011). Who's funny: gender stereotypes, humor production, and memory bias. *Psychon. Bull. Rev.* 19, 108–112. doi: 10.3758/s13423-011-0161-2

Nevo, O., Nevo, B., and Yin, J. L. S. (2001). Singaporean humor: a cross-cultural, cross-gender comparison. *J. Gen. Psychol.* 128, 143–156. doi: 10.1080/00221300109598904

Özdoğru, A. A. (2017). Çok boyutlu mizah duygusu ölçeği: bir güncelleme. *Avrasya Pozitif Psikoloji Kongresi'nde Sunulmuş bildiri*, İstanbul.

Parekh, M. (1999). *Humor Perception in Men and Women: Effect of Joke Type and Gender of Joke Butt*. Undergraduate thesis. Texas A&M University, College Station, TX.

Quiros, A., and Vaid, J. (1998). "Perceptions of conversational uses of humor by men and women as ascribed to the same vs. the opposite sex," in *Poster presented at Conference of the International Society for Humor Research, Holy Names College*, Oakland, CA.

Ruch, W. (1992). "Assessment of appreciation of humor: studies with the 3WD humor test," in *Advances in Personality Assessment*, Vol. 9, eds C. D. Spielberger and J. N. Butcher (Hillsdale, NJ: Erlbaum), 27–75.

Schiau, I. (2017). Women gossip and men brag: perceived gender differences in the use of humor by Romanian older women. *Rom. J. Commun. Public Relat.* 19, 69–76. doi: 10.21018/rjcpr.2017.1.231

Vaid, J. (1999). "The evolution of humor: Do those who laugh last?," in *Evolution of the Psyche*, eds D. Rosen and M. Luebbert (Westport, CT: Praeger), 123–138.

Vaid, J. (2000). New approaches to conceptual representations in bilingual memory: the case for studying humor interpretation. *Biling. Lang. Cogn.* 3, 28–30. doi: 10.1017/S1366728900290111

Vaid, J. (2006). "Joking across languages: perspectives on humor, emotion, and bilingualism," in *Bilingual Minds: Emotional Experience, Expression, and Representation*, ed. A. Pavlenko (Clevedon: Multilingual Matters), 152–182.

Vaid, J. (2014). "Absence makes the heart grow fonder. and also makes the eyes wander: optimal innovations in proverb rejoinders," in *Proceedings of International Conference on Empirical Aesthetics*, ed. A. Kozbelt (New York, NY: Hunter college).

Vaid, J., Choi, H., Chen, H. C., and Friedman, M. (2008). Perceiving and responding to embarrassing predicaments across languages: cultural influences on the mental lexicon. *Ment. Lex.* 3, 122–148. doi: 10.1075/ml.3.1.08vai

Vaid, J., Hull, M. R., Heredia, R., Gerkens, D., and Martinez, F. (2003). Getting a joke: the time course of meaning activation in verbal humor. *J. Pragmat.* 35, 1431–1449. doi: 10.1016/S0378-2166(02)00184-4

Vaid, J., and Hull, R. (1998). "Humor appreciation and the status quo," in *Poster Presented at Annual Meeting of the Psychonomics Society*, Dallas, TX.

Vaid, J., and Kobler, J. B. (2000). Laughing matters: toward a structural and neural account. *Brain Cogn.* 42, 139–141. doi: 10.1006/brcg.1999.1184

Vaid, J., Lopez, B. G., and Martinez, F. E. (2015). "Linking the figurative to the creative: Bilinguals' comprehension of metaphors, jokes, and remote associates," in *Bilingual Figurative Language Processing*, eds R. R. Heredia and A. B. Cieslicka (Cambridge: Cambridge University Press), 53–86.

Warren, C., and McGraw, A. P. (2016). Differentiating what is humorous from what is not. *J. Pers. Soc. Psychol.* 110, 407–430. doi: 10.1037/pspi0000041

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Tosun, Faghihi and Vaid. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 April 2017
doi: 10.3389/fpsyg.2017.00616

Experimentally Manipulating Items Informs on the (Limited) Construct and Criterion Validity of the Humor Styles Questionnaire

Willibald Ruch and Sonja Heintz*

Personality and Assessment, Department of Psychology, University of Zurich, Zurich, Switzerland

How strongly does humor (i.e., the construct-relevant content) in the Humor Styles Questionnaire (HSQ; Martin et al., 2003) determine the responses to this measure (i.e., construct validity)? Also, how much does humor influence the relationships of the four HSQ scales, namely affiliative, self-enhancing, aggressive, and self-defeating, with personality traits and subjective well-being (i.e., criterion validity)? The present paper answers these two questions by experimentally manipulating the 32 items of the HSQ to only (or mostly) contain humor (i.e., construct-relevant content) or to substitute the humor content with non-humorous alternatives (i.e., only assessing construct-irrelevant context). Study 1 (N = 187) showed that the HSQ affiliative scale was mainly determined by humor, self-enhancing and aggressive were determined by both humor and non-humorous context, and self-defeating was primarily determined by the context. This suggests that humor is not the primary source of the variance in three of the HQS scales, thereby limiting their construct validity. Study 2 (N = 261) showed that the relationships of the HSQ scales to the Big Five personality traits and subjective well-being (positive affect, negative affect, and life satisfaction) were consistently reduced (personality) or vanished (subjective well-being) when the non-humorous contexts in the HSQ items were controlled for. For the HSQ self-defeating scale, the pattern of relationships to personality was also altered, supporting an positive rather than a negative view of the humor in this humor style. The present findings thus call for a reevaluation of the role that humor plays in the HSQ (construct validity) and in the relationships to personality and well-being (criterion validity).

Keywords: Humor Styles Questionnaire, humor, measurement, validity, item wording, well-being, personality, scale construction

OPEN ACCESS

Edited by:
Anat Bardi,
Royal Holloway, University of London, UK

Reviewed by:
Mary Louise Cowan,
Regent's University London, UK
Feng Kong,
Shaanxi Normal University, China

****Correspondence:***
Willibald Ruch
w.ruch@psychologie.uzh.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *24 November 2016*
Accepted: *03 April 2017*
Published: *20 April 2017*

Citation:
Ruch W and Heintz S (2017) Experimentally Manipulating Items Informs on the (Limited) Construct and Criterion Validity of the Humor Styles Questionnaire. Front. Psychol. 8:616. doi: 10.3389/fpsyg.2017.00616

INTRODUCTION

Most questionnaire items contain both the construct they intend to measure (i.e., the construct-relevant content) but also additional information, which should measure the relevant content in a variety of circumstances to increase its representativeness (see Epstein, 1983). In a homogenous scale (i.e., a scale that uniformly measures a single construct), one would thus expect similar contents, as these form the core of the scale, but somewhat dissimilar contexts. For example, the construct of "liking to laugh" can be shown in different contexts, such as being with family or

friends, being told a joke, or watching a funny movie in the cinema. The tendency to laugh more than others should then generalize across the different situations. The item contexts should vary so that summing up the items over a scale strengthens the variance due to the core content and more or less averages out the different situations. This additional information might not only refer to situational contexts, but also to states, feelings, or evaluations that specify the core content in more detail. Importantly, mostly the variance contributed by the content should be the relevant one.

Besides the core content, additional elements may unintentionally produce a considerable amount of variance in a scale if it is homogenous and strongly represented or if the content is not that salient. For example, measuring "liking to laugh" with items such as "While I deliver a lecture to my class I laugh a lot," "When my colleagues make a funny remark in a faculty meeting I laugh easily," and "My assistants and I laugh a lot when we hear that our article was accepted." The answers to these items might not be determined by the tendency of liking to laugh alone, and relationships to other constructs (e.g., vocational background) would likely be biased by the additional elements in the items. Messick's (1995) mentioned this "construct-irrelevant variance" as a threat to construct validity, which occurs if "the assessment is too broad, containing excess reliable variance associated with other distinct constructs" (p. 742). Hence the amount of variance contributed by the construct-relevant content and the non-relevant context can be an indicator of the construct validity of an instrument, as the scale compositions and their relations to other constructs should be mainly driven by the construct they are intended to measure (i.e., construct-relevant variance), and less so or not all by the remainder of the item (i.e., construct-irrelevant variance).

How can the contribution of construct-relevant and construct-irrelevant variance be empirically investigated? For example, the item wording could be experimentally altered to only assess construct-relevant contents in the items, or the relevant content could be removed to yield purely construct-irrelevant items. Although not investigating construct validity, Haigler and Widiger (2001) experimentally manipulated the items of the NEO-PI-R to reverse their desirability/adaptiveness without changing the item content itself. Specifically, they changed the items from desirable/adaptive to undesirable/maladaptive, or from having a positive to a negative connotation. They simply added descriptors such as "too much" or "excessively" to the items, resulting in a reversal of desirability/adaptiveness as judged by raters. In addition, the pattern of correlations with personality disorders changed for the rephrased items in a sample of 86 adult outpatients. Most strikingly, the experimentally manipulated version of conscientiousness correlated strongly and positively with obsessive-compulsive personality, and agreeableness correlated strongly with dependent and avoidant personality disorders (while the original NEOPI-R scales showed mostly zero correlations). This study empirically supports the idea that already slight changes in item wording can change the construct that is measured (which was also found in a recent study by Blasberg et al., 2016) and its desirability/adaptiveness.

The present paper combines both Messick's (1995) ideas about construct-irrelevant variance in the contexts and the experimental manipulations of item wordings. We aim at experimentally disentangling the construct-relevant content from the remainder of the item by creating new items that only assess the core content (i.e., pure construct-relevant indicators) or by replacing the core content (i.e., pure construct-irrelevant indicators). The first study compares the similarities of the two experimentally manipulated versions with the original items and scales to yield insights into the construct validity of the original instrument. To support construct validity, relationships of the original version should be higher with the construct-relevant indicators than with the construct-irrelevant ones. Ideally, each original scale should perfectly converge with its pure construct-relevant indicators, supporting that it only assesses the construct to be measured and not other unrelated and possibly confounding elements. The second study extends the item wording manipulation to test the criterion validity of a scale. Controlling for the construct-irrelevant indicators (using the experimentally rephrased items) should reveal the "pure" correlations of the constructs under question with a set of external criteria.

This procedure is applied to the Humor Styles Questionnaire items (HSQ; Martin et al., 2003), which assesses four humor styles that represent functions of humor in everyday life, and especially those functions relevant to psychosocial well-being. The construct-relevant content hence comprises humor (including joking, laughing, and making fun of oneself and others) and functions (using humor to enhance oneself or relationships to others). The four humor styles are affiliative (amusing others, liking to laugh, and making jokes to enhance one's relationships with others), self-enhancing (amusing oneself and cheering oneself up with humor to enhance oneself), aggressive (making jokes, laughing at others, and teasing others to enhance oneself), and self-defeating (making fun of oneself and letting others laugh about oneself to enhance one's relationships with others). According to Martin et al. (2003), the affiliative humor style should be associated with better psychosocial well-being (as it should be affirming of both self and others). The self-enhancing humor style should be associated with better psychological well-being (as it entails a coping aspect). The aggressive humor style should be associated with lower social well-being (as it entails putting others down). Finally, the self-defeating humor style should be associated with lower psychological well-being (due to a negative self-evaluation and emotional avoidance underlying it).

The present investigation focuses on the humor-related contents, as the role humor plays in the HSQ is of special interest: First, the HSQ is the most widely used questionnaire in research on individual differences in humor (see Martin, 2015). Second, its interpretations usually focus on the humor-related content, for example, considering humor as a mediator in the relationship with well-being or implementing humor exercises based on findings with the HSQ. Third and foremost, inspection of its items frequently shows a salient context where it does not deem necessary (e.g., "being alone" in self-enhancing humor items; laughing at oneself "too much" in self-defeating

humor). It seems necessary to demonstrate empirically that these variations in context do average out and do not bias the overall meaning of the scale. Thus, investigating to what extent the four HSQ scales and their relationships to relevant criteria (in this case subjective well-being) are determined by humor vs. other construct-irrelevant elements is an important indicator of their construct and criterion validity.

The experimental manipulation of the 32 items of the HSQ proceeded as follows: They were rephrased to only contain their construct-relevant content (i.e., humor-related words or phrases; "*Humor-HSQ*") or the construct-relevant content was replaced ("*No-Humor-HSQ*"). To generate the No-Humor-HSQ, the items were minimally changed to replace the humor elements (substituting them by something similar but non-humorous). To generate the Humor-HSQ, everything that went beyond the humor content (be it situational conditions, thoughts or feelings during the humor behavior, or evaluations of the behavior) was stripped of. For example, the HSQ self-defeating item "I let people laugh at me or make fun at my expense more than I should" can be reduced to its humor part ("I let people laugh at me or make fun at my expense") or the humor content can be replaced ("I let people offend me or look down on me more than I should"). The former reduces the humor-related constructs to their core and the latter leaves the item intact but eliminates the reference to humor (i.e., leaves only construct-irrelevant context).

STUDY 1: COMPOSITION OF THE HUMOR STYLES QUESTIONNAIRE

Study 1 tests the construct validity of the HSQ by comparing the original HSQ with the Humor-HSQ and the No-Humor-HSQ. First, it is expected that the internal consistency of the three HSQ versions will vary in a predictable way. To the extent that the non-humorous elements produce variance, it makes the items more dissimilar, thereby increasing the internal consistencies of the Humor-HSQ scales and reducing the internal consistencies of the No-Humor-HSQ scales (in comparison to the HSQ scales). Second, the intercorrelations of the three HSQ versions should be influenced similarly. Ideally, if the non-humorous elements produce only construct-irrelevant variance that is averaged out within the four scales, then the HSQ should not correlate (or only slightly) with No-Humor-HSQ scales, and high with the Humor-HSQ (approaching unity in true-score correlations). The more construct-relevant variance is contributed by the non-humorous elements, the higher correlations can be expected between the No-Humor-HSQ and the HSQ, and the lower correlations can be expected between the Humor-HSQ and the HSQ.

Materials and Methods

Participants

Of the 289 German-speaking participants who started the survey, 201 (69.9%) completed all the items. A total of 187 participants (17.1% men) with a median age of 24 ($M = 28.81$, $SD = 10.76$) ranging from 17 to 63 years provided valid responses in this study (14 participants were excluded because they answered more than 12 items per minute, indicating inattentiveness)[1]. Participants were primarily Swiss (58.3%), German (34.2%), and from several other nations. Most participants were well-educated, with 34.2% being college or university students, 33.2% having passed tertiary education, 24.1% having A-levels, and 7.0% having an apprenticeship. A subsample of the present data was used by Ruch and Heintz (2013, study 2). None of the present results have been published before and they extend the previous study by investigating the overlap between the three HSQ versions.

Instruments

Humor Styles Questionnaire (HSQ; Martin et al., 2003; German version by Ruch and Heintz, 2016)

The HSQ consists of 32 items measuring the four humor styles. Sample items are "I don't often joke around with my friends." (affiliative, negatively keyed), "Even when I'm by myself, I'm often amused by the absurdities of life" (self-enhancing), "If someone makes a mistake, I will often tease them about it." (aggressive), and "I let people laugh at me or make fun at my expense more than I should" (self-defeating). The instrument employs a seven-point Likert scale from "totally disagree" (1) to "totally agree" (7).

Content version derived from the HSQ (Humor-HSQ)

The 32 HSQ items were rephrased to only capture the relevant humor content, resulting in four humor scales. Sample items are "I don't often joke around" (affiliative, negatively keyed), "I'm often amused by the absurdities of life." (self-enhancing), "I often tease others" (aggressive), and "I let people laugh at me or make fun at my expense." (self-defeating). The instrument employs the same Likert scale as the HSQ. The item order and keying of the original HSQ was preserved except for one self-enhancing and one aggressive item, which were positively keyed to ensure comprehensibility.

Two raters (the second author and a graduate psychology student) judged which parts of the HSQ items referred to humor vs. context. Interrater agreement (Cohen's kappa) was 0.77. Only the parts judged as containing humor were retained for the Humor-HSQ items (e.g., the item "Even when I'm by myself, I'm often amused by the absurdities of life" was rephrased into "I'm often amused by the absurdities of life"). The set of items was finalized in a discussion between the two authors.

Context version derived from the HSQ (No-Humor-HSQ)

The 32 HSQ items were rephrased to only capture the relevant context component, resulting in four humor-free context scales. Sample items are "I don't often converse with my friends" (affiliative, negatively keyed), "Even when I'm by myself, I often occupy myself with the little things in life." (self-enhancing), "If someone makes a mistake, I will often reproach them about it." (aggressive), and "I let people offend me or look down on me more than I should" (self-defeating). The instrument employs the same Likert scale, item order and keying as the HSQ. The 32

[1]Control analyses showed that none of the results were altered by excluding these participants.

items of the Humor-HSQ and the No-Humor-HSQ are listed in the **Table A1** in Appendix.

The item rephrasing process for the No-Humor-HSQ proceeded in two steps: (a) Identifying the humorous word(s) or expression(s) in each HSQ item, and (b) substituting it/them with a non-humorous, but equivalent counterpart. In step (a), the two raters judged the core humor word(s) of each item (Cohen's kappa = 0.82). In addition, every humor word that was not agreed upon (e.g., "blunder") was further analyzed using two online-thesauri (www.openthesaurus.de and www.thesaurus.com) to ensure that either the definition or one of the synonyms related to humor.

After all humorous words had been identified, they were substituted in step (b) with a non-humorous expression that was as equivalent as possible (e.g., "misapprehension" instead of "blunder," "enthralling," or "beautiful" instead of "funny"). The criterion of being humor-free was fulfilled if none of the meanings and synonyms contained a humorous word (using the two online thesauri). Equivalent meant that the word was from the same part of speech (e.g., verb, adjective, noun) and encompassed a similar level of activity (e.g., communication, action) and affect (e.g., positive, negative). In addition, nine raters (post-graduate psychologists) judged the 32 newly written No-Humor-HSQ items for their humor content (*Does the item still contain a trace/hint to humor?*), similarity (*Is/Are the replaced "humor-free" word[s] similar to the original one[s], or is there any deviation in relation to part of speech, activity, or affect?*), and overall meaningfulness (*Is the item still meaningful or are there any inconsistencies that hamper or prevent understanding the item?*). Items were iteratively improved according to each rater's judgments, and the set of items was then finalized in a discussion between the two authors to ensure that the No-Humor-HSQ items did not contain humor, that they were similar to the original, and that they were meaningful.

Procedure

The data were collected in an online survey (www.unipark.info) employing a forced-choice item format. The No-Humor-HSQ was presented first, followed by the Humor-HSQ and then the original HSQ. Further variables on personality and well-being were collected that are not relevant to the present study, yet they were used as "fillers" in between the three HSQ versions. Participants were recruited via several means, including mailing lists of the University of Zurich, social media platforms, and bulletins. They were offered a personalized feedback and/or course credit in psychology for their participation. The study was conducted in compliance with the local ethical guidelines and participants provided online informed consent.

Data Analysis

First, internal consistencies (McDonald's omega) and scale intercorrelations were computed to compare the three versions of the HSQ (original, humor, and no-humor). McDonald's omega was computed with the MBESS package (Kelley and Lai, 2012) in R (R Core Team, 2016). The differences between the (dependent) correlations were compared using the psych package (Revelle, 2015) in *R*. Correction for attenuation [according to Spearman's (1904) classical formula] was employed to reveal the true-score correlations between the scales of the three HSQ versions.

Results

Observed Scale Intercorrelations

Table 1 shows the means, standard deviations, intercorrelations, and internal consistencies of the HSQ, Humor-HSQ, and No-Humor-HSQ scales.

As shown in **Table 1**, the internal consistencies of the Humor-HSQ scales were high (≥0.80) and always numerically higher than the ones of the homologous HSQ scales. In turn, the internal consistencies of the No-humor-HSQ scales were always numerically lower than the HSQ scales, yet they still evidenced good internal consistencies (>0.70), with the exception of the aggressive scale (0.42). The correlations among the scales of the Humor- and No-Humor-HSQ with the homologous HSQ scales were all high ($rs \geq 0.61$, $ps < 0.05$), indicating that both the humor and the non-humor elements were relevant for the HSQ scales. Comparing the size of the correlations between the homologous scales of the two HSQ versions with the original HSQ, significant differences were found for the affiliative ($t = 7.03$, $p < 0.001$), self-enhancing ($t = 2.42$, $p = 0.017$), and self-defeating ($t = -2.91$, $p = 0.004$) scales. The correlations of the HSQ affiliative and self-enhancing scales were significantly larger with the Humor-HSQ than with the No-Humor-HSQ, indicating that the humor content was more relevant for these HSQ scales than the non-humorous elements. This effect was reversed for the HSQ self-defeating scale; that is, the No-Humor-HSQ, in comparison to the Humor-HSQ, correlated significantly higher with the HSQ. This indicates that the non-humorous elements were more important in the HSQ self-defeating scale than its humor core.

Numerically comparing the scale intercorrelations within each HSQ version, a few peculiarities can be noted. First, the HSQ and the No-Humor-HSQ showed small to medium intercorrelations (both positive and negative), while the Humor-HSQ scales were all positively correlated (medium to large effects). Second, the HSQ affiliative scale had large intercorrelations with all Humor-HSQ scales. Third, the Humor-HSQ self-defeating scale correlated positively with all HSQ scales (small to large effects), including the HSQ self-enhancing scale.

True-Score Scale Intercorrelations

The true-score correlations [using a double correction for attenuation with Spearman's (1904) formula] were close to one for three of the four HSQ and Humor-HSQ scales: Affiliative (0.98), self-enhancing (0.94), and aggressive (1.00), while the value was considerably lower for self-defeating (0.69). However, correlations were also close to one for three of the four HSQ and No-Humor-HSQ scales: Self-enhancing (0.94), aggressive (1.00) and self-defeating (0.95), while the true-score correlation was slightly lower for affiliative (0.85).

Item Intercorrelations

This raises the question to what extent the findings at the scale-level are also present at the level of the individual items. As each item was assessed in all three versions of the HSQ,

TABLE 1 | Means, standard deviations, intercorrelations, and internal consistencies of the Humor Styles Questionnaire (HSQ) scales and the derived Humor-HSQ and No-Humor-HSQ scales.

	M	*SD*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HSQ														
(1) AF	39.98	9.81	*0.91*											
(2) SE	34.97	8.58	0.40*	*0.87*										
(3) AG	26.93	6.90	0.25*	−0.12	*0.74*									
(4) SD	24.86	8.39	−0.05	−0.16*	0.21*	*0.88*								
HUMOR-HSQ														
(5) AF	39.14	10.47	0.91*	0.40*	0.29*	0.02	*0.94*							
(6) SE	37.69	8.26	0.61*	0.83*	0.07	−0.10	0.65*	*0.89*						
(7) AG	31.57	7.89	0.53*	0.07	0.77*	0.28*	0.61*	0.33*	*0.80*					
(8) SD	31.91	8.87	0.51*	0.23*	0.31*	0.61*	0.57*	0.43*	0.58*	*0.89*				
NO-HUMOR HSQ														
(9) AF	36.73	9.07	0.74*	0.42*	0.10	−0.25*	0.67*	0.57*	0.29*	0.30*	*0.84*			
(10) SE	35.57	7.08	0.20*	0.75*	−0.24*	−0.20*	0.18*	0.57*	−0.10	0.11	0.35*	*0.73*		
(11) AG	27.74	5.50	0.27*	−0.04	0.76*	0.01	0.29*	0.08	0.59*	0.17*	0.19*	−0.19*	*0.42*	
(12) SD	26.50	7.39	−0.23*	−0.32*	0.12	0.76*	−0.15*	−0.26*	0.11	0.30*	−0.37*	−0.33*	−0.02	*0.72*

N = 187. AF, affiliative; SE, self-enhancing; AG, aggressive; SD, self-defeating. McDonald's omegas in italics. Theoretical minimum mean of the scales = 8, maximum mean = 56.
**p < 0.05.*

comparing their correlations with one another can reveal the relative influence of humor and non-humor elements within each item. **Table 2** shows the intercorrelations of the HSQ items with the corresponding items of the Humor-HSQ and the No-Humor-HSQ.

As shown in **Table 2**, six of eight items (all except for items 9 and 29) of the HSQ affiliative scale correlated significantly higher with the homologous items of the Humor-HSQ than with the No-Humor-HSQ. For the HSQ self-enhancing scale, two items (items 6 and 22) correlated significantly higher with the Humor-HSQ than with the No-Humor-HSQ, while this effect was reversed for two other items (items 26 and 30). For the HSQ aggressive and self-defeating scales, four (items 11, 19, 27, and 31) and three items (items 20, 24, and 28), respectively, showed significantly different correlations, indicating that their relationship with the No-Humor-HSQ was significantly higher than the relationship with the Humor-HSQ.

Discussion

The aim of Study 1 was to test the construct validity of the HSQ by comparing the original HSQ with newly created Humor- and No-Humor-HSQ versions. Construct validity would be supported if the humor content turned out to be more important than the no-humor elements, evidenced by predicable patterns of internal consistencies and intercorrelations. First, the expected pattern of internal consistencies was found (Humor-HSQ scales > HSQ scales > No-Humor-HSQ). Thus, removing construct-irrelevant context made the four scales more similar, and removing the construct-relevant content made them less similar. Interestingly, the No-Humor-HSQ scales mostly had acceptable internal consistencies (McDonald's omega > 0.70, except for the aggressive scale with 0.42), indicating that participants answered the no-humor elements within each HSQ scale somewhat similarly. That is, the no-humor elements within the HSQ items did not average out at the scale-level and were thus able to contribute reliable variance to the No-Humor-HSQ scales.

Second, the pattern of intercorrelations of the affiliative and self-enhancing scales supported the primary importance of the humor core in two HSQ scales. Specifically, the intercorrelation between the HSQ and the Humor-HSQ was significantly higher than the one between the HSQ and the No-Humor-HSQ. The self-defeating scale showed the reverse effect, with the HSQ being more similar to the No-Humor- ($r^2 = 0.58$) than the Humor-HSQ ($r^2 = 0.37$). In other words, the non-humorous elements (i.e., construct-irrelevant variance) were more important than the humor core (i.e., construct-relevant variance) in the HSQ self-defeating scale.

The pattern found in the observed correlations was also corroborated in the true-score correlations. The HSQ affiliative scale was virtually identical with the Humor-HSQ scale, supporting the interpretation that it is mainly determined by humor. This was also the case for the individual items, yielding strong support for the construct validity of the HSQ affiliative scale. Along these lines, the HSQ affiliative scale correlated positively with all Humor-HSQ scales (large effects), suggesting that the humor contents of the four scales resembled the affiliative humor style, that is, amusing others, liking to laugh, and making jokes.

The true-score correlations showed that HSQ self-enhancing scale was highly similar to the Humor-HSQ scale and the No-Humor HSQ scale. Interestingly, these effects largely varied across the eight self-enhancing items. Item 6 ("Even when I'm by myself, I'm often amused by the absurdities of life.") and Item 22 ("If I am feeling sad or upset, I usually lose my sense of humor") showed higher correlations to the Humor-HSQ than the No-Humor-HSQ; that is, humor was more relevant in these two items than the context. Thus, people who are more or less frequently amused by the incongruities of life and who keep or

TABLE 2 | Intercorrelations of the Humor Styles Questionnaire (HSQ) Items with the Homologous Items of the Humor-HSQ and No-Humor-HSQ.

	AF	SE	AG	SD
ITEM 1				
HSQ with Humor-HSQ	0.73*[a]	0.67*	0.48*	0.42*
HSQ with No-Humor HSQ	0.49*[b]	0.63*	0.55*	0.41*
Humor-HSQ with No-Humor HSQ	0.46*	0.39*	0.20*	0.08
ITEM 2				
HSQ with Humor-HSQ	0.85*[a]	0.75*[a]	0.46*	0.65*
HSQ with No-Humor HSQ	0.49*[b]	0.03[b]	0.53*	0.59*
Humor-HSQ with No-Humor HSQ	0.42*	−0.07	0.36*	0.51*
ITEM 3				
HSQ with Humor-HSQ	0.48*	0.50*	0.24*[a]	0.50*
HSQ with No-Humor HSQ	0.42*	0.61*	0.53*[b]	0.59*
Humor-HSQ with No-Humor HSQ	0.39*	0.39*	0.14*	0.32*
ITEM 4				
HSQ with Humor-HSQ	0.79*[a]	0.67*	0.60*	0.30*
HSQ with No-Humor HSQ	0.64*[b]	0.68*	0.59*	0.19*
Humor-HSQ with No-Humor HSQ	0.61*	0.56*	0.32*	0.08
ITEM 5				
HSQ with Humor-HSQ	0.67*[a]	0.65*	0.43*[a]	0.43*[a]
HSQ with No-Humor HSQ	0.46*[b]	0.72*	0.73*[b]	0.67*[b]
Humor-HSQ with No-Humor HSQ	0.43*	0.51*	0.31*	0.26*
ITEM 6				
HSQ with Humor-HSQ	0.77*[a]	0.54*[a]	0.60*	0.48*[a]
HSQ with No-Humor HSQ	0.44*[b]	0.30*[b]	0.60*	0.71*[b]
Humor-HSQ with No-Humor HSQ	0.35*	0.35*	0.58*	0.29*
ITEM 7				
HSQ with Humor-HSQ	0.62*[a]	0.46*[a]	0.53*[a]	0.05[a]
HSQ with No-Humor HSQ	0.35*[b]	0.67*[b]	0.66*[b]	0.73*[b]
Humor-HSQ with No-Humor HSQ	0.33*	0.38*	0.48*	−0.15*
ITEM 8				
HSQ with Humor-HSQ	0.77*	0.27*[a]	0.14[a]	0.57*
HSQ with No-Humor HSQ	0.71*	0.71*[b]	0.35*[b]	0.61*
Humor-HSQ with No-Humor HSQ	0.54*	−0.01	0.04	0.32*

*N = 187. AF, affiliative; SE, self-enhancing; AG, aggressive; SD, self-defeating.*p < 0.05. [a,b]Correlations with different superscripts differed significantly from one another (at the 0.05 level).*

lose their sense of humor seem to do so independent of the social context or the emotional states they are in. By contrast, Item 26 ("It is my experience that thinking about some amusing aspect of a situation is often a very effective way of coping with problems") and Item 30 ("I don't need to be with other people to feel amused—I can usually find things to laugh about even when I'm by myself") showed higher correlations to the No-Humor-HSQ than the Humor-HSQ. Thus, the context in these items was more relevant than the humor. This implicates that either the context is dominant in the items (i.e., coping with problems or being by oneself), or humor is not a determining or unique factor in such situations (e.g., people cope with problems humorously, but also by non-humorous means). Thus, the construct validity of the HSQ self-enhancing scale can be mostly supported, though two of the eight items were largely determined by construct-irrelevant variance.

The true-score correlations of the HSQ aggressive scale with the homologous Humor-HSQ and No-Humor-HSQ scales were 1.00, showing that the HSQ scale was identical to both experimentally manipulated versions. Note that the latter true-score correlation exceeded 1.00 in the computation, indicating an overcorrection due to the low internal consistency of the No-Humor-HSQ aggressive scale (see Muchinsky, 1996). However, the relevance of the No-Humor-HSQ was also supported in the observed correlations and in the item-level analyses: Four of the eight HSQ aggressive items showed significantly higher correlations to the No-Humor-HSQ than to the Humor-HSQ. These effects were most pronounced for Item 11 ("When telling jokes or saying funny things, I am usually not very concerned about how other people are taking it") and Item 19 ("Sometimes I think of something that is so funny that I can't stop myself from saying it, even if it is not appropriate for the situation"). Again, this yields two possible interpretations: Either the context is dominant (not caring about others opinions or feelings, and acting impulsively and inappropriately) or humor is not a decisive factor in these items (e.g., people saying something humorous and non-humorous while not being concerned about others, or doing so impulsively and inappropriately). Thus, the construct validity of the HSQ aggressive can be partly supported, yet the strong context effects found in specific items require further scrutiny.

In contrast to the other HSQ scales, the HSQ self-defeating scale was almost identical to the homologous No-Humor-HSQ scale, but not to the Humor-HSQ scale. This effect was also found in three of the eight HSQ self-defeating items: Item 5 ("I often go overboard in putting myself down when I am making jokes or trying to be funny"), Item 6 ("When I am with friends or family, I often seem to be the one that other people make fun of or joke about") and Item 7 ("If I am having problems or feeling unhappy, I often cover it up by joking around, so that even my closest friends don't know how I really feel"). Again, either the humor in the items might not be very salient (and thus non-humorous, but similar behaviors strongly overlap with the item), or the context is dominating (e.g., going overboard, or covering up problems and negative feelings). Additionally, the self-defeating humor core was compatible with all humor styles (also with the self-enhancing one).

These findings might rather support the interpretation that the context was dominating in the HSQ self-defeating items. This suggests a potentially impactful implication: Probably the humor content can be meaningfully interpreted, yet not along the lines of the self-defeating humor style as proposed by Martin et al. (2003). This could potentially explain the contradiction between the conception of the HSQ self-defeating scale as mostly maladaptive (Martin et al., 2003), and the humor core of laughing at oneself, which is generally considered a positive trait (e.g., McGhee, 1999; see also Ruch and Heintz, 2013). The negative aspect of the HSQ self-defeating scale could be due to the primary influence of the non-humor elements of this scale, which are mostly negative connoted (like putting oneself down excessively).

Taking a closer look at the pattern of intercorrelations within one HSQ version also revealed that all HSQ-Humor scales correlated significantly and positively with one another (medium

to large effects), while this was not the case for the HSQ and No-Humor-HSQ scales. The latter two HSQ-versions also had negative scale intercorrelations. Hence, the participants rated the humor contents in the four scales quite similarly, while they differentiated the scales better once the non-humorous elements were involved. This underlines that the differentiation between the four HSQ scales might be more driven by varying their non-humorous elements *across* the scales (e.g., being with others vs. being alone, being in a sad or depressed mood vs. being cheerful) than by their humor cores.

Besides testing the construct validity, separating the construct-relevant and construct-irrelevant elements also allows for testing their contributions to correlations with other constructs and outcomes (i.e., criterion validity). For example, it was shown that correlations of the HSQ with personality traits and aspects of psychological well-being were mainly driven by the No-Humor-HSQ, while relations to other humor constructs (such as laughing at oneself) were mainly driven by the Humor-HSQ (Ruch and Heintz, 2013). This effect was most pronounced for the self-defeating scale, which is in line with the present findings. Study 2 aims at investigating the relevance of the construct-irrelevant context in the scales in relation to several criteria (personality and well-being), replicating and extending these previous findings.

STUDY 2: CRITERION VALIDITY OF THE HUMOR IN THE HUMOR STYLES QUESTIONNAIRE

In addition to construct validity, it is relevant to investigate the criterion validity of the HSQ scales. The relevant criteria of the HSQ are humor and psychosocial well-being, as the humor style concepts were derived from the literature in these two areas, and as the humor styles are defined as everyday functions of humor that are relevant to psychosocial well-being (Martin et al., 2003). Besides relating the HSQ to humor-related scales (e.g., Martin et al., 2003; Kuiper et al., 2004; Ruch and Heintz, 2016) and humor behaviors (Heintz, 2017), the HSQ is usually compared to personality traits (for a meta-analysis with the Big Five personality traits, see Mendiburo-Seguel et al., 2015) and to subjective well-being (e.g., Edwards and Martin, 2010, 2014; Jovanovic, 2011; Ruch and Heintz, 2013; Maiolino and Kuiper, 2014). These relationships have usually been associated with the humor in the HSQ scales. However, previous studies found rather low incremental validities of the HSQ scales in explaining subjective well-being (Jovanovic, 2011; Dyck and Holtzman, 2013; Ruch and Heintz, 2013; Maiolino and Kuiper, 2014; Heintz, 2017). Also the results from Study 1 cast doubt on the role of humor in the HSQ self-defeating scale, making further investigations on the criterion validity in terms of personality and well-being necessary.

Study 2 investigates the criterion validity of the HSQ scales with the Big Five personality traits, namely extraversion, agreeableness, conscientiousness, emotional stability, and culture (also labeled openness or intellect), and subjective well-being, consisting of life satisfaction as a cognitive component and positive and negative affect as affective components. In line with previous findings positive relationships are expected for the affiliative and self-enhancing scales with extraversion and openness to experience/culture. The self-enhancing scale should also positively correlate with emotional stability and with agreeableness. The aggressive scale should correlate negatively with conscientiousness and agreeableness. The self-defeating scale should correlate negatively with emotional stability and conscientiousness. In terms of subjective well-being, the affiliative and self-enhancing scales should correlate positively with life satisfaction and positive affect, and negatively with negative affect, while this pattern should be reversed for the self-defeating scale. No significant correlations are expected for the aggressive scale. The change in this relationship once the homologous No-Humor-HSQ scales are controlled for is utilized as an indicator of the criterion validity of the HSQ scales.

One previous study employed the same approach to investigate the criterion validity of the HSQ in terms of six indicators of psychological well-being (Ruch and Heintz, 2013). They found that only 3 of the 13 significant relationships remained significant once the No-Humor-HSQ was taken into account. This approach was also employed in Study 2, instead of investigating the Humor-HSQ scales directly, as the Humor-HSQ still contains some elements that are not related to humor, simply because the items needed to be meaningful by themselves (e.g., "I let others laugh at me, which keeps them in in good spirits." for self-defeating or "I usually try to think of something funny about a situation." for self-enhancing). The No-Humor-HSQ, by contrast, is parallel to the HSQ, and the only difference lies in the absence vs. presence humor-related terms and phrases. The test of criterion validity conducted in Study 2 is thus stricter, but also more precise. Based on the previous findings on the incremental validity and criterion validity of the HSQ scales and Study 1, we expected small criterion validities of the HSQ beyond the No-Humor-HSQ.

Materials and Methods

Participants

Of the 474 German-speaking participants that started the survey, 272 (57.4%) completed all the items. A total of 261 participants (30.7% men) with a median age of 24.00 (M = 27.26, SD = 10.11) ranging from 18 to 69 years provided valid responses in this study (participants were excluded if they indicated an age below 18 years [n = 9] or if they showed aberrant answer patterns like always using the same answer option or answering randomly [n = 2]). Participants were primarily Swiss (63.2%), German (26.8%), and from several other nations. Most participants were well-educated, with 50.2% being college or university students, 23.0% having passed tertiary education, 22.2% having A-levels, and 4.6% having <12 years of education. A subsample of the present data was used by Heintz (2017). None of the present results have been published before, and they extend the previous study by investigating the cross-sectional correlations among the HSQ, the No-Humor HSQ, personality, and subjective well-being.

Instruments

Humor Styles Questionnaire (HSQ; Martin et al., 2003; German version by Ruch and Heintz, 2016)
The same version of the HSQ was used as in Study 1.

Context version derived from the HSQ (No-Humor-HSQ)
The same version of the No-Humor-HSQ was used as in Study 1.

MRS-25
Inventory of minimally redundant scales (MRS; Schallberger and Venetz, 1999) . The MRS employs bipolar adjectives to assess the Big Five personality traits extraversion (e.g., talkative/quiet), agreeableness (e.g., well-tempered/short-tempered), conscientiousness (e.g., organized/disorganized), emotional stability (e.g., relaxed/oversensitive), and culture (e.g., artistic/inartistic). The 25-item version was used (five items for each trait). It employs a six-point Likert scale with mirrored labels: "very" (−3/+3), "quite" (−2/+2), and "rather" (−1/+1).

SWLS
Satisfaction with life scale (SWLS; Diener et al., 1985) . The SWLS measures life satisfaction (e.g., "I am satisfied with my life") with five items. It employs a seven-point Likert scale from "strongly disagree" (1) to "strongly agree" (7).

PANAS
Positive and negative affect schedule (PANAS; Watson et al., 1988). The PANAS measures positive affect (e.g., enthusiastic) and negative affect (e.g., nervous) with 10 items each. It employs a five-point Likert scale from "very slightly or not at all" (1) to "extremely" (5).

Procedure

The data were collected in an online survey (www.unipark.info) using the German versions of the instruments. The order of presentation was PANAS, No-Humor-HSQ, SWLS, a humor questionnaire (not relevant for the present study), MRS-25, and HSQ. All items were obligatory to answer. Participants were recruited in similar venues as those of Study 1. They were offered a personalized feedback and/or course credit in psychology for their participation. The study was conducted in compliance with the local ethical guidelines and participants provided their online informed consent.

Data Analysis

As in Study 1, internal consistencies (McDonald's omega) and scale intercorrelations were computed to compare the HSQ and the No-Humor-HSQ. Criterion validity was investigated in stepwise multiple regression, entering each No-Humor-HSQ scale in the first step and the homologous HSQ scale in the second step. Multicollinearity in the regression was low (variance inflation factors between 1.8 and 2.5).

Results

Table 3 shows the descriptive statistics, internal consistencies, and correlations with the HSQ scales and the No-Humor-HSQ scales. Replicating the findings of Study 1, the HSQ scales were always (numerically) more internally consistent than the corresponding No-Humor-HSQ scales. The No-Humor-HSQ scales again showed high internal consistencies (>0.60), this time also for the aggressive scale. Correlations between the homologous scales were high and comparable to Study 1. The true-score correlations supported the equivalence of the HSQ and No-Humor-HSQ scales for self-enhancing (0.95), aggressive (1.00), and self-defeating (1.00), but not for affiliative (0.80), again replicating the findings from Study 1. In addition, the correlations between the HSQ items and the homologous No-Humor-HSQ items (shown in **Table 4**) were highly similar to Study 1.

Relationships with Personality and Subjective Well-Being

As shown in **Table 3**, both the HSQ and the No-Humor-HSQ scales showed similar and mostly significant relationships to personality and subjective well-being: Affiliative related most strongly to extraversion, self-enhancing to emotional stability, aggressive to lower agreeableness, and self-defeating to lower emotional stability. The relationships were in general similar to the ones reported in the meta-analysis by Mendiburo-Seguel et al. (2015). Also in line with previous findings, affiliative, and self-defeating correlated positively with subjective well-being, while self-defeating was negatively related to it.

Criterion Validity beyond Context

Next, the criterion validity of the HSQ over and above its construct-irrelevant context is investigated, yielding information on the specific relationships of the humor in the HSQ (as construct-relevant content). **Table 5** provides the results of standard multiple regression analyses explaining subjective well-being with the No-Humor-HSQ in step 1 and the HSQ in step 2 (separately for each humor style).

As shown in **Table 5**, the variance that the HSQ scales explained over and above their homologous No-Humor-HSQ scales in subjective well-being was not significant. Thus, the humorous contents in the HSQ did not uniquely explain subjective well-being once the context elements were controlled for (although 10 significant correlations were originally present). The magnitude of the effects was comparable to the previous study (Ruch and Heintz, 2013). In terms of personality, seven regressions yielded significant amounts of explained variance (1.0–5.0%) for the HSQ scales (from 12 originally significant correlations). The humor in the HSQ affiliative scale was uniquely related to agreeableness and extraversion, and the humor in the HSQ self-enhancing scale was uniquely related to extraversion and openness. The humor in the HSQ aggressive scale showed a unique negative relationship to conscientiousness, and the humor in the HSQ self-defeating scale showed unique relationships to agreeableness and extraversion. Thus, while no significant criterion validities were found between the humor in the HSQ scales and subjective well-being, each HSQ scale had their unique pattern of criterion validities across the Big Five personality traits.

TABLE 3 | Descriptive statistics, internal consistencies, and correlations with the Humor Styles Questionnaire (HSQ) scales and the No-Humor-HSQ Scales.

	M	*SD*	ω	Correlations with the HSQ				Correlations with the No-Humor-HSQ			
				AF	SE	AG	SD	AF	SE	AG	SD
HSQ											
Affiliative[a]	42.62	7.99	0.88								
Self-enhancing[a]	36.87	8.11	0.84	0.42*							
Aggressive[a]	27.47	6.78	0.70	0.15*	0.04						
Self-defeating[a]	24.95	8.16	0.82	0.01	0.04	0.00					
NO-HUMOR-HSQ											
Affiliative[a]	38.89	8.29	0.83	0.68*	0.29*	0.13*	−0.12				
Self-enhancing[a]	36.06	6.73	0.70	0.23*	0.73*	−0.11	−0.06	0.28*			
Aggressive[a]	27.74	6.35	0.61	0.01	−0.07	0.75*	−0.07	0.18*	−0.14*		
Self-defeating[a]	24.72	7.28	0.71	−0.15*	−0.13*	0.03	0.77*	−0.27*	−0.21*	0.02	
PERSONALITY											
Agreeableness[b]	4.51	0.73	0.73	0.23*	0.32*	−0.44*	0.02	0.10	0.32*	−0.54*	−0.10
Conscientiousness[b]	4.29	0.94	0.87	0.00	0.01	−0.24*	−0.08	0.06	0.08	−0.19*	−0.14
Emotional stability[b]	3.76	1.01	0.86	0.33*	0.40*	0.12	−0.30*	0.48*	0.47*	0.13*	−0.48*
Extraversion[b]	4.10	1.05	0.87	0.58*	0.26*	0.03	−0.06	0.72*	0.21*	0.09	−0.22*
Culture[b]	4.46	0.87	0.87	0.29*	0.28*	−0.13*	−0.05	0.34*	0.24*	−0.10	−0.07
SUBJECTIVE WELL-BEING											
Positive affect[c]	33.42	6.12	0.85	0.25*	0.33*	0.00	−0.20*	0.43*	0.40*	0.08	−0.32*
Negative affect[c]	19.74	7.03	0.89	−0.20*	−0.35*	−0.04	0.34*	−0.32*	−0.40*	0.01	0.42*
Life satisfaction[d]	4.45	1.49	0.91	0.28*	0.34*	0.14*	−0.35*	0.48*	0.46*	0.15*	−0.49*

*N = 261. ω = McDonald's omega. *p < 0.05.*
[a] *Theoretical minimum = 8, maximum = 56.*
[b] *Theoretical minimum = 1, maximum = 6.*
[c] *Theoretical minimum = 10, maximum = 50.*
[d] *Theoretical minimum = 1, maximum = 7.*

TABLE 4 | Intercorrelations of the Humor Styles Questionnaire (HSQ) items with the Homologous items of the No-Humor-HSQ.

	AF	SE	AG	SD
Item 1	0.38*	0.50*	0.45*	0.51*
Item 2	0.44*	0.14*	0.54*	0.55*
Item 3	0.43*	0.60*	0.60*	0.65*
Item 4	0.59*	0.70*	0.44*	0.15*
Item 5	0.40*	0.68*	0.77*	0.62*
Item 6	0.43*	0.40*	0.56*	0.67*
Item 7	0.41*	0.59*	0.67*	0.78*
Item 8	0.76*	0.61*	0.35*	0.47*

*N = 261. AF, affiliative; SE, self-enhancing; AG, aggressive; SD, self-defeating.*p < 0.05.*

Discussion

Study 2 aimed at partially replicating the findings of Study 1 regarding the construct validity of the HSQ and at extending the validity analyses to the criterion validity in terms of personality and subjective well-being. The relationships between the HSQ scales and the No-Humor-HSQ scales were highly similar to Study 1, thus replicating and strengthening the previous findings on the construct validity of the HSQ.

Criterion validities varied across the two sets of criteria (personality and subjective well-being). Seven of the 12 relationships between the HSQ and the Big Five personality traits were robust beyond the No-Humor-HSQ. Thus, the humor in each of the four humor styles had a unique relevance to one or two personality traits. The humor in the HSQ affiliative scale was relevant to agreeableness and extraversion, showing that it comprised unique prosocial and social qualities. The humor in the HSQ self-enhancing scale was relevant to extraversion and culture, also supporting a unique social quality, but also a cognitive aspect. The latter might be due to recognizing incongruities in one's surroundings and being amused by them, which is a core component of humor. Openness (or culture) has thus also been implied in appreciating non-sense humor and in humor creation (e.g., Galloway and Chirico, 2008; Nusbaum, 2015). However, the relationship of the HSQ self-enhancing scale to emotional stability was not specific to humor, showing that enhancing oneself and coping with problems could be achieved non-humorously.

The humor in the HSQ aggressive scale uniquely related to lower conscientiousness (but not agreeableness). Thus, aggressive humor did not have an antisocial quality, suggesting that the label "aggressive" might not fit well to the humor content of this scale. The relationship to lower conscientiousness could probably be explained by a playful attitude underlying this humor style (e.g.,

TABLE 5 | Stepwise multiple regression analyses predicting subjective well-being with the No-Humor (NH) version of the Humor Styles Questionnaire (HSQ) in step 1 and the HSQ in step 2 (separately for each scale; only last step reported).

	Big Five personality traits										Subjective well-being					
	A		C		ES		Extraversion		Culture		PA		NA		LS	
Predictor	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	0.01		0.00		0.23*		0.52*		0.11*		0.19*		0.10*		0.23*	
NH-HSQ AF		−0.11		0.12		0.48*		0.61*		0.26*		0.43*		−0.32*		0.48*
Step 2	0.05*		0.00		0.00		0.01*		0.01		0.00		0.00		0.01	
HSQ AF		0.30*		−0.08		0.01		0.16*		0.11		−0.09		0.03		−0.10
Total R^2	0.06*		0.01		0.23*		0.53*		0.12*		0.19*		0.10*		0.24*	
Step 1	0.10*		0.01		0.22*		0.04*		0.06*		0.16*		0.16*		0.22*	
NH-HSQ SE		0.20*		0.15		0.39*		0.04		0.08		0.40*		−0.40*		0.46*
Step 2	0.01		0.01		0.01		0.03*		0.02*		0.00		0.01		0.00	
HSQ SE		0.17		−0.10		0.11		0.24*		0.22*		0.07		−0.12		−0.01
Total R^2	0.12*		0.01		0.23*		0.07*		0.08*		0.16*		0.17*		0.22*	
Step 1	0.29*		0.04*		0.02*		0.01		0.01		0.01		0.00		0.02*	
NH-HSQ AG		−0.48*		−0.02		0.08		0.15		−0.01		0.08		0.01		0.15*
Step 2	0.00		0.02*		0.00		0.00		0.01		0.01		0.00		0.00	
HSQ AG		−0.08		−0.23*		0.06		−0.08		−0.12		−0.16		−0.09		0.06
Total R^2	0.29*		0.06*		0.02		0.01		0.02		0.02		0.00		0.02*	
Step 1	0.01		0.02*		0.23*		0.05*		0.01		0.10*		0.18*		0.24*	
NH-HSQ SD		−0.28*		−0.19*		−0.59*		−0.41*		−0.08		−0.32*		0.42*		−0.49*
Step 2	0.02*		0.00		0.01		0.03*		0.00		0.01		0.00		0.00	
HSQ SD		0.24*		0.07		0.15		0.25*		0.01		0.11		0.04		0.07
Total R^2	0.03*		0.02		0.24*		0.07*		0.01		0.11*		0.18*		0.24*	

*N = 261. AF, affiliative; SE, self-enhancing; AG, aggressive; SD, self-defeating; A, agreeableness; C, conscientiousness; ES, emotional stability; PA, positive affect; NA, negative affect; LS, life satisfaction. *p < 0.05.*

teasing and making fun of others, but not in a hurtful or malicious way). This interpretation would be supported by previous studies relating the HSQ aggressive scale to lower seriousness (Martin et al., 2003) and the finding that the Humor-HSQ aggressive scale correlated positively with playfulness (Ruch and Heintz, 2013). The humor in the HSQ self-defeating scale uniquely related to agreeableness and extraversion. Thus, although the HSQ self-defeating was unrelated to both personality traits, the humor in this scale had a unique prosocial and social quality, similar to the humor in the affiliative scale. As with self-enhancing, no humor-specific effects were present for emotional stability, clearly limiting the maladaptive interpretation of self-defeating humor.

For subjective well-being, the criterion validity of the HSQ cannot be supported, as no significant amounts of variance could be explained by the HSQ scales once their homologous No-Humor-HSQ scales were controlled for. Thus, the frequently found relationships between the HSQ and subjective well-being (positively for affiliative and self-enhancing and negative for self-defeating) seem to be driven mostly or entirely by the non-humorous elements (i.e., the construct-irrelevant context) and not the humor itself (i.e., construct-relevant content). This is also in line with the usually low incremental validities of the HSQ scales in explaining subjective well-being over and above the Big Five personality traits (Jovanovic, 2011; Dyck and Holtzman, 2013; Ruch and Heintz, 2013).

Two implications can be derived from the present findings: First, humor was not a decisive factor in the relationships between the HSQ and subjective well-being. For example, it cannot be firmly concluded that affiliative and self-enhancing humor is positive and that self-defeating humor is negative. Instead, the non-humorous elements in these humor styles (e.g., liking to be with others, being able to cope with problems, or putting oneself down excessively) were the active ingredients in the relationship with subjective well-being. Second, which aspects of these non-humorous elements is most relevant in this relationship (e.g., situations, functions, states, or evaluations or combination or interaction between them) remains open for further investigation.

Does this mean that the humor in the HSQ is completely irrelevant to subjective well-being? As stated before, the present test is a rather strict one. Directly correlating the Humor-HSQ scales to six aspects of psychological well-being revealed positive correlations for affiliative and self-enhancing humor, but zero

correlations for self-defeating humor (Ruch and Heintz, 2013). In a similar vein, daily-measured humor behaviors that were similar (but not equivalent) to the Humor-HSQ scales exhibited incremental validity in explaining subjective well-being beyond personality and the HSQ (Heintz, 2017); specifically cheerful (similar to affiliative), amused (similar to self-enhancing), and self-directed (similar to self-defeating) humor behaviors. Thus, there is evidence that the humor in the HSQ can be positive in terms of psychological well-being. Most importantly, the negativity of the HSQ self-defeating scale was not supported in these less stringent analyses. This humor style can thus best be interpreted as having a negative context, yet the humor in it is either unrelated to psychological well-being or positive. This precludes drawing conclusions such as "learning how to decrease one's use of self-defeating humor" (Maiolino and Kuiper, 2014, p. 568) for enhancing one's well-being. The conclusion should rather be "putting oneself less down" (whether with humor or not) to increase one's well-being, which seems to be both a trivial and circular reasoning.

GENERAL DISCUSSION

Study 1 and 2 yielded converging evidence that the construct validity of the HSQ affiliative scale can be fully supported, while the construct validities of the HSQ self-enhancing and aggressive scales yielded mixed findings. The construct validity of the HSQ self-defeating scale could not be supported. Thus, the term "humor" in the humor styles seems appropriate for affiliative, needs be used with caution for self-enhancing and aggressive, and seems inappropriate for self-defeating. Combining these findings with the criterion validities, the humor content in the self-enhancing humor style might be rather labeled cultured or open-minded affiliative humor, and the humor in the aggressive humor style might rather be playful teasing.

The lack of criterion validity in terms of subjective well-being necessitates a reinterpretation of the role that humor plays in subjective well-being. For the affiliative and the self-enhancing humor style, the extent to which they are relevant to subjective well-being might have been overestimated in previous studies, as the primary motor of the relationships seems to lie in the non-humorous elements (e.g., the "style," or function, or contexts). While this only affects the magnitude of the relationships, the consequences are more severe for the HSQ self-defeating scale: This humor style has been implied to be negative, yet both its construct and criterion validities showed that the non-humorous elements determined this humor style more than humor did, and no negative—but rather positive—effects emerged. Importantly, this was also the case when less stringent tests were used; that is, when the humor in the self-defeating humor style were directly related to well-being (Ruch and Heintz, 2013; Heintz, 2017). Thus, the humor in the self-defeating humor style might be quite similar to the notion of an adaptive ability of laughing at yourself (McGhee, 1999) after all.

While the present study focused on one instrument of relevance for humor research, the general principle is independent of the instrument studied. Indeed, we believe that the methodology and considerations used here can be applied to psychological questionnaires in general, and in particular when the items are more complex and merge core behaviors and contextual variables. This is often the case, as traits are defined by behaviors that are consistent across time and situations. This is usually implemented by varying the context in which the behaviors occur and a strong context might generate variance itself. Also items may contain conditions for behaviors, where the conditions already have different probabilities, and hence contribute to the variance in response to the item. For example, an item "when traveling abroad, I usually prefer to stay away from problem areas" might be envisioned to be an item for prudence. However, very prudent people might disagree to the item when they just do not travel abroad at all. This made-up item demonstrates that only some of the variance is due to prudence, but the other part of the variance is actually capturing the opposite of it. Thus, the importance of item wording should not be underestimated, and it is best already considered during the process of test construction. Cognitive interviewing techniques (see e.g., Willis, 2004), for example, can detect whether items are understood in way that is intended by the creator.

Limitations and Suggestions for Future Research

First, the generalization of the results is limited to a German-speaking, young, and well-educated sample; hence, replications in other languages, cultures, and samples with a wider range in age or education are desirable. Second, the order of presentation of the different HSQ versions was not randomized, and thus any systematic influences associated with the order of presentation could have interfered with our findings. Third, the present study focused on one of two construct-relevant contents in the humor styles, namely humor (which could certainly considered to be the more prominent one given the name of the construct and their treatment in previous research). Investigating the role of "style" (as functions or uses) in the composition of the HSQ and its role in the relationships with other criteria would complement the present investigations of construct and criterion validity. Fourth, further experimental evidence is necessary to investigate the causal relationships between the HSQ, humor, personality and subjective well-being. For example, investigating which emotional states are associated and elicited by self-defeating humor experiences, or by self-defeating humor trainings, would enhance our understanding of the role that the humor entailed in the HSQ plays in criteria such as subjective well-being. Fifth, our investigations of the criterion validity of the HSQ scales focused on personality and subjective well-being. As the HSQ has been frequently studied in relation to other trait-like variables (such as character strengths; Edwards and Martin, 2014), extending the scope to further criteria would yield a more complete picture of the role that the humor in the humor styles plays.

CONCLUSION

The present studies showed that humor might not be as relevant in the humor styles as would be naturally and usually assumed. This might explain why Martin et al. (2003) found that "the

HSQ accounts for a greater proportion of variance in well-being than do several existing self-report humor scales." (p. 72 f.), which was also corroborated by Edwards and Martin (2014). If humor measures are compared to a measure that contains a large proportion of non-humorous elements that are related to well-being, the latter instrument might seem "better"—yet this does not tell us anything new about the relevance of humor in well-being. Thus, Martin et al.'s (2003) outlook that research with the HSQ "may provide better understanding of the ways in which humor may function as an adaptive resource for psychological health, as well as the ways in which it may interfere with healthy adjustment and impair relationships with others." (p. 73) seems to be hard to fulfill with the HSQ (at least in its current form). Researchers interested the relationships of humor to subjective well-being and potentially other well-being outcomes should thus be cautioned, as the HSQ scales yield rather limited information on the role that humor itself plays in these relationships (and in the case of self-defeating humor potentially misleading information). Other approaches to humor styles, such as the Humor-Behavior Q-Sort Deck (Craik et al., 1996) or comic styles (e.g., Schmidt-Hidding, 1963) might be fruitful alternatives in this regard. Future research might yield smaller, yet likely more realistic relationships, between humor and well-being.

ETHICS STATEMENT

These studies were carried out in accordance with the recommendations of the Ethical Principles of Psychologists and Code of Conduct (APA) and the Ethical Guidelines for Psychologists of the Swiss Psychological Society (SGP), as outlined by the ethics committee of the Faculty of Arts at the University of Zurich, with online informed consent from all subjects. All subjects gave online informed consent in accordance with the Declaration of Helsinki. It was not possible to obtain written informed consent, as both studies were conducted solely via the Internet, yet online informed consent was shown to be similar to written informed consent (see Varnhagen et al., 2005). The protocol was exempt from approval as stated by the guidelines of the ethics committee of the Faculty of Arts at the University of Zurich, as it passed a checklist of ethical innocuousness (which serves as ethical approval in accordance with the local guidelines).

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Blasberg, J. S., Hewitt, P. L., Flett, G. L., Sherry, S. B., and Chen, C. (2016). The importance of item wording the distinction between measuring high standards versus measuring perfectionism and why it matters. *J. Psychoeduc. Assess.* 34, 1–16. doi: 10.1177/0734282916653701

Craik, K. H., Lampert, M. D., and Nelson, A. J. (1996). Sense of humor and styles of everyday humorous conduct. *Humor* 9, 273–302. doi: 10.1515/humr.1996.9.3-4.273

Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *J. Pers. Assess.* 49, 71–75. doi: 10.1207/s15327752jpa4901_13

Dyck, K. T., and Holtzman, S. (2013). Understanding humor styles and well-being: The importance of social relationships and gender. *Pers. Individ. Dif.* 55, 53–58. doi: 10.1016/j.paid.2013.01.023

Edwards, K. R., and Martin, R. A. (2010). Humor creation ability and mental health: are funny people more psychologically healthy? *Eur. J. Psychol.* 6, 196–212. doi: 10.5964/ejop.v6i3.213

Edwards, K. R., and Martin, R. A. (2014). The conceptualization, measurement, and role of humor as a character strength in positive psychology. *Eur. J. Psychol.* 10, 505–519. doi: 10.5964/ejop.v10i3.759

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *J. Pers.* 51, 360–392. doi: 10.1111/j.1467-6494.1983.tb00338.x

Galloway, G., and Chirico, D. (2008). Personality and humor appreciation: evidence of an association between trait neuroticism and preferences for structural features of humor. *Humor* 21, 129–142. doi: 10.1515/HUMOR.2008.006

Haigler, E. D., and Widiger, T. A. (2001). Experimental manipulation of NEO-PI-R items. *J. Pers. Assess.* 77, 339–358. doi: 10.1207/S15327752JPA7702_14

Heintz, S. (2017). Putting a spotlight on daily humor behaviors: Dimensionality and relationships with personality, subjective well-being, and humor styles. *Pers. Individ. Dif.* 104, 407–412. doi: 10.1016/j.paid.2016.08.042

Jovanovic, V. (2011). Do humor styles matter in the relationship between personality and subjective well-being? *Scand. J. Psychol.* 52, 502–507. doi: 10.1111/j.1467-9450.2011.00898.x

Kelley, K., and Lai, K. (2012). *MBESS: MBESS. R Package Version 3.3.3.* Available online at: http://CRAN.R-project.org/package=MBESS

Kuiper, N. A., Grimshaw, M., Leite, C., and Kirsh, G. (2004). Humor is not always the best medicine: Specific components of sense of humor and psychological well-being. *Humor* 17, 135–168. doi: 10.1515/humr.2004.002

Maiolino, N. B., and Kuiper, N. A. (2014). Integrating humor and positive psychology approaches to psychological well-being. *Eur. J. Psychol.* 10, 557–570. doi: 10.5964/ejop.v10i3.753

Martin, R. A. (2015). On the challenges of measuring humor styles: response to Heintz and Ruch. *Humor* 28, 635–639. doi: 10.1515/humor-2015-0096

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., and Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: development of the humor styles questionnaire. *J. Res. Pers.* 37, 48–75. doi: 10.1016/S0092-6566(02)00534-2

McGhee, P. E. (1999). *The Laughter Remedy: Health, Healing and the Amuse System*. Dubuque, IA: Kendall/Hunt.

Mendiburo-Seguel, A., Páez, D., and Martínez-Sánchez, F. (2015). Humor styles and personality: a meta-analysis of the relation between humor styles and the Big Five personality traits. *Scand. J. Psychol.* 56, 335–340. doi: 10.1111/sjop.12209

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am. Psychol.* 50, 741–749. doi: 10.1037/0003-066X.50.9.741

Muchinsky, P. M. (1996). The correction for attenuation. *Educ. Psychol. Meas.* 56, 63–75. doi: 10.1177/0013164496056001004

Nusbaum, E. C. (2015). *A Meta-Analysis of Individual Differences in Humor Production and Personality*. Doctoral thesis, The University of North Carolina at Greensboro. Available online at: http://libres.uncg.edu/ir/uncg/f/Nusbaum_uncg_0154D_11766.pdf (Accessed November 23, 2016).

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna. Available online at: http://www.R-project.org/

Revelle, W. (2015). *psych: Procedures for Personality and Psychological Research*. Evanston, IL: Northwestern University. Available online at: https://CRAN.R-project.org/package=psych

Ruch, W., and Heintz, S. (2013). Humour styles, personality, and psychological well-being: what's humour got to do with it? *Eur. J. Humour Res.* 1, 1–24. doi: 10.7592/EJHR2013.1.4.ruch

Ruch, W., and Heintz, S. (2016). The german version of the humor styles questionnaire: psychometric properties and overlap with other styles of humor. *Eur. J. Psychol.* 12, 434–455. doi: 10.5964/ejop.v12i3.1116

Schallberger, U., and Venetz, M. (1999). *Kurzversionen des MRS-Inventars von Ostendorf (1990) zur Erfassung der fünf "Grossen" Persönlichkeitsfaktoren [Brief Versions of Ostendorf's (1990) MRS Inventory for the Assessment of the Big Five Personality Factors]*. Zurich: Department of Applied Psychology, University of Zurich.

Schmidt-Hidding, W. (1963). *Europäische Schlusselwörter: Band 1. Humor und Witz [European Keywords: Vol. 1. Humor and Wit]*. Munich: Huber.

Spearman, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* 15, 72–101. doi: 10.2307/1412159

Varnhagen, C. K., Gushta, M., Daniels, J., Peters, T. C., Parmar, N., Law, D., et al. (2005). How informed is online informed consent? *Ethics Behav.* 15, 37–48. doi: 10.1207/s15327019eb1501_3

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Willis, G. B. (2004). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage Publications.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Ruch and Heintz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

TABLE A1 | Overview of the 32 items of the humor- and no-humor versions of the Humor Styles Questionnaire.

Scale	Item	No-Humor-HSQ	Humor-HSQ
AF	1	I usually don't talk or converse much with other people. [recoded]	I usually don't laugh or joke around much. [recoded]
SE	2	If I am feeling depressed, I can usually put myself in a better mood with something beautiful.	I can usually cheer myself up with humor.
AG	3	If someone makes a mistake, I will often reproach them about it.	I often tease others.
SD	4	I let people offend me or look down one me more than I should.	I let people laugh at me or make fun at my expense.
AF	5	I don't have to work very hard at impressing other people—I seem to be a naturally influencing person.	I make others laugh easily—I am a humorous person.
SE	6	Even when I'm by myself, I often occupy myself with the little things in life.	I'm often amused by the absurdities of life.
AG	7	People are never offended or hurt by my manner of speaking. [recoded]	My sense of humor is never offending or hurting. [recoded]
SD	8	I will often get carried away in putting myself down if it makes my family or friends feel good.	I often put myself down and thus make others laugh.
AF	9	I rarely impress other people by telling enthralling stories about myself. [recoded]	I rarely tell funny stories about myself, which make others laugh. [recoded]
SE	10	If I am feeling upset or unhappy, I usually try to think of something beautiful about the situation to make myself feel better.	I usually try to think of something funny about a situation.
AG	11	When telling experiences or saying enthralling things, I am usually not very concerned about how other people are taking it.	I usually tell jokes or say funny things.
SD	12	I often try to make people like or accept me more by saying something appealing about my own weaknesses, misapprehensions, or faults.	I often say something funny about my own weaknesses, blunders, or faults.
AF	13	I talk or converse a lot with my friends.	I laugh and joke a lot.
SE	14	My positive outlook on life keeps me from getting overly upset or depressed about things.	I have a humorous outlook on life.
AG	15	I do not like it when people tell experiences or say things as a way of criticizing or putting someone down. [recoded]	I do not like criticizing or putting-down humor. [recoded]
SD	16	I don't often say enthralling things to put myself down. [recoded]	I rarely say funny things about myself. [recoded]
AF	17	I usually don't like to tell experiences or to impress people. [recoded]	I usually don't tell jokes or amuse people. [recoded]
SE	18	If I'm by myself and I'm feeling unhappy, I make an effort to think of something beautiful to make me feel better.	I always think of something funny to cheer myself up.
AG	19	Sometimes I think of something that is so enthralling that I can't stop myself from saying it, even if it is not appropriate for the situation.	Sometimes I think of extremely funny things.
SD	20	I often go overboard in putting myself down when I am telling experiences or trying to be communicative.	I often make jokes about myself or make fun of myself.
AF	21	I enjoy impressing people.	I make people laugh.
SE	22	If I am feeling sad or upset, I usually lose my serenity. [recoded]	I never lose my sense of humor.
AG	23	I never participate in offending others even if all my friends are doing it. [recoded]	I never laugh at others. [recoded]
SD	24	When I am with friends or family, I often seem to be the one that other people look down on or offend.	I let others often make fun of me or joke about me.
AF	25	I don't often converse with my friends. [recoded]	I don't often joke around. [recoded]
SE	26	It is my experience that thinking about some beautiful aspect of a situation is often a very effective way of coping with problems.	I often think about some amusing aspect of a situation.
AG	27	If I don't like someone, I often criticize or reproach them to put them down.	I often use humor about others or tease them.
SD	28	If I am having problems or feeling unhappy, I often cover it up by telling an experience, so that even my closest friends don't know how I really feel.	I often joke around.
AF	29	I usually can't think of enthralling things to say when I'm with other people. [recoded]	I usually can't think of witty things. [recoded]
SE	30	I don't need to be with other people to feel good—I can usually find things to occupy myself with even when I'm by myself.	I am usually amused and I can find things to laugh about.
AG	31	Even if something is really relevant to me, I will not say anything or criticize it if someone will be offended. [recoded]	I always laugh or joke about something that is really funny to me.
SD	32	Letting others offend me is my way of making my friends and family feel good.	I let others laugh at me, which keeps them in in good spirits.

AF, affiliative; SE, self-enhancing; AG, aggressive; SD, self-defeating. The order of the items and the response options are the same as in the HSQ (Martin et al., 2003).

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 August 2018
doi: 10.3389/fpsyg.2018.01470

Assessing Theory of Mind by Humor: The Humor Comprehension and Appreciation Test (ToM-HCAT)

Simge Aykan and Erhan Nalçacı*

Department of Physiology, Ankara University School of Medicine, Ankara, Turkey

Theory of Mind (ToM) may be defined as the ability to understand the mental states, such as beliefs, desires, intentions, and emotions, of others. Impairment of ToM ability leads to disorders with pathologies in social skills, such as autism spectrum disorder and schizophrenia. In addition to differences in ToM ability among patient populations, there is variation between neurotypical individuals. Unfortunately, ToM tasks are usually developed for children or patients with cognitive disorders and cannot detect variations in healthy adults. As an alternative tool, humor may be used. Humor plays a role in social communication and requires many different cognitive functions. Humor is believed to represent complex high-order cognitive processes. There are numerous types of humor; the most complex type is considered ToM humor, where an understanding of social/emotional content is necessary. Given the need for a ToM assessment test suitable for healthy adult populations, we developed a test for measuring humor comprehension and appreciation, with and without ToM content (ToM-HCAT). The present ToM-HCAT test is a performance test consisting of cartoons. The test measures perceived funniness, reaction time to perceived funniness decision, and meaning inference. Cartoons were selected after pilot studies involving 44 participants. Subscales were constituted according to expert views and confirmed by confirmatory factor analysis (N = 135). Goodness of fit values for the final 35-item test were acceptable to excellent: GFI = 0.97; AGFI = 0.97; NFI = 0.97; RFI = 0.97, and SRMR = 0.067. Both categories were internally consistent (α_1 = 0.84, α_2 = 0.94). External validity was assessed against autistic traits. One hundred and three participants completed the Autism Spectrum Quotient and were grouped by +0.5 standard deviations from the mean as high in autistic traits. The meaning-inference scores of the subscale with the ToM cartoons were significantly lower (p = 0.034) for the high autistic traits group, providing evidence of external validity. In conclusion, we developed and validated a test for assessment of ToM by humor comprehension and appreciation. We believe that the present test will be useful for the detection of variations in ToM ability in the healthy adult population.

Keywords: autistic traits, cartoons, humor, reliability, theory of mind, validity

OPEN ACCESS

Edited by:
Willibald Ruch,
Universität Zürich, Switzerland

Reviewed by:
Ursula Beermann,
Universität Innsbruck, Austria
Raymond A. Mar,
York University, Canada

****Correspondence:***
Simge Aykan
saykan@ankara.edu.tr

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *04 April 2018*
Accepted: *26 July 2018*
Published: *13 August 2018*

Citation:
Aykan S and Nalçacı E (2018) Assessing Theory of Mind by Humor: The Humor Comprehension and Appreciation Test (ToM-HCAT). Front. Psychol. 9:1470. doi: 10.3389/fpsyg.2018.01470

INTRODUCTION

As highly social beings, humans encounter a variety of interactions during their daily lives. Being successful in this environment requires insight into the social and emotional context and understanding of others' intentions and aims, which is enabled by empathy. Empathy can be described as "*Any process that emerges from the fact that observers understand others' states by activating personal, neural and mental representations of that state, including the capacity to be affected by and share the emotional state of another; assess the reasons for the other's state; and identify with the other, adopting his or her perspective*" (de Waal and Preston, 2017, p. 498). In other words, empathy is characterized by the sharing of emotions and consideration of the perspectives of other people. Empathy may be divided into two categories: affective and cognitive (Singer, 2006; Zaki et al., 2012). Affective empathy involves the ability to match others' emotions, while cognitive empathy refers to the ability to imagine how others feel. A type of cognitive empathy is theory of mind (ToM), which may be defined as the ability to understand the mental states of others, such as their beliefs, desires, intentions, and emotions (Wellman and Estes, 1986). In brief, ToM refers to the ability to understand one's own, and others', minds (Baron-Cohen, 2000). The importance of ToM may be illustrated by disorders in which ToM is impaired, such as autism spectrum disorders (ASDs) and schizophrenia (Baron-Cohen, 2000; Bora and Pantelis, 2013; Chung et al., 2014). Regarding autism, one of two main areas of impairment is that of social skills/communication: it is known that ASD patients exhibit the ability to share the emotions of others, but cannot mentalize (Smith, 2006), which is an indicator of impaired cognitive empathy. As underlying causes of disorders, ToM impairments have been studied extensively (Baron-Cohen, 2000; Losh et al., 2012; Chung et al., 2014; Sommer et al., 2018). Regarding schizophrenia, ToM impairments have been shown in unaffected relatives, ultra high-risk individuals, and first-episode patients as evidence of the trait-based nature of the disease (Bora and Pantelis, 2013; Lavoie et al., 2013).

In addition to differences in patient populations, variation in ToM abilities is observed among neurotypical individuals (Baron-Cohen et al., 2001a). However, ToM assessments are usually developed for children or cognitively disabled people (for a review, see Turner and Felisberti, 2017). As a result, the task used may not be sufficiently difficult for healthy adults with strong cognitive and social skills. These tasks have ceiling effects (near 100% accuracy) for healthy control participants (Corcoran et al., 1995; Gallagher et al., 2000; Brüne, 2003; Marjoram et al., 2005), which makes the detection of variations impossible. This limits the investigation of ToM in healthy populations, which is unfortunate as such investigations may shed light on its underlying mechanisms. Although the investigation of a cognitive mechanism in disabled populations provides useful insights, the results may be confounded by the presence of comorbid conditions. In addition, the disabled cognitive mechanism might be compensated by other processes, which will again lead to misinterpretation. Thus, investigation of the mechanism underlying a cognitive process should be accompanied by research in the healthy population. ToM variations are known to exist in healthy individuals (Baron-Cohen et al., 2001a), examples of whom include healthy first-degree relatives of schizophrenia patients (Janssen et al., 2003; Anselmetti et al., 2009; Bora and Pantelis, 2013) and relatives of individuals with ASD (Baron-Cohen and Hammer, 1997; Losh and Piven, 2007; Gokcen et al., 2009). Further, self-reported and neuroimaging data indicate variance in social cognition in the normal population (Hooker et al., 2010; Wagner et al., 2011; Regenbogen et al., 2015); however, behavioral data are lacking.

Current tasks used to measure ToM vary from social vignettes (e.g., false belief tasks, social animation tasks) to narrative fictional stories and films (e.g., strange stories tasks). There are limited number of ToM ability tests sensitive to variation in healthy population: the Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001a), the Faux Pas Test (Stone et al., 1998; Gregory et al., 2002), the Yoni Test (Shamay-Tsoory and Aharon-Peretz, 2007), the DANVA (Nowicki and Duke, 2001), among others. As ToM is a highly complex process with cognitive and affective components that can be implicit and explicit, using a diversity of approaches for assessment is necessary. In addition to current tests, the use of humor represents a potential alternative method for ToM assessment in the healthy population. Humor might be described as anything that people say or do that is perceived as funny and makes others laugh (Martin, 2007). Humor is a way of communicating ideas, strengthening relations, improving group harmony, and expressing aggressiveness in a positive manner. Humor is the most flexible tool for social interaction. Therefore, it is important to express and understand humor to communicate more effectively. Humor is a stimulus encountered often in our daily lives, and the evaluation of humorous material may be considered similar to real-life situations, which makes it an appropriate tool for measuring ToM ability. The simplest form of humor is the pun, which uses visual or semantic resemblance, and the most complex form is ToM humor, which requires ToM abilities (Vrticka et al., 2013).

Humor processing consists of two stages: comprehension (the first stage) and appreciation (the second; Suls, 1972; Wyer and Collins, 1992; Vrticka et al., 2013). The most accepted theory of these is 'incongruity detection and resolution,' which states that humor requires the introduction of the incongruity as a violation of expectations, followed by a resolution associated with enjoyment (Shultz, 1972; Martin, 2007). Humor comprehension requires understanding of the context and detection of incongruity (Ruch, 1992; Coulson et al., 2006; Uekermann et al., 2007). Necessary cognitive processes for incongruity detection may vary from recognition of simple visual resemblance to mentalizing, which requires ToM ability. The second stage, humor appreciation, requires both integration of newly formed meaning in an amusing way and a positive emotional response (Wyer and Collins, 1992; Coulson et al., 2006; Uekermann et al., 2007). Therefore, humor appreciation represents a complex, high order process that involves cognitive, behavioral, physiological, emotional, and social components (Martin, 2007).

In addition to the previously mentioned ToM disability in ASD, another relevant trait is humor impairment. Asperger's

Syndrome (AS; one of the subtypes of ASD) was first defined by Hans Asperger in 1944. Individuals with AS are known to exhibit differences in terms of their perception of humor. This observation is supported by the fact that they have problems in understanding irony or sarcasm (Happé, 1995). Since Asperger's work, researchers have verified humor-related deficits in ASD (Baron-Cohen, 1997; Emerich et al., 2003; Samson and Hegenloh, 2010). Samson and Hegenloh (2010) showed that humor appreciation in ASD individuals depends on the stimulus material. Appreciation was low for ToM cartoons, whereas no difference was observed for visual puns (Samson and Hegenloh, 2010). This result shows that humor appreciation is not reduced when ToM is not necessary.

Humor comprehension and appreciation differences in ASD would be expected to extend to the healthy population with autistic traits. Autistic traits are subthreshold deficits similar to those present in ASD, such as social interaction and communication deficits, as well as restrictive/repetitive behaviors (Constantino and Todd, 2003). The main difference between individuals with ASD and healthy people with autistic traits is the severity of the symptoms. A theory for ASD is that social adaptation and communication skills exhibit a normal distribution among the population, and individuals at the negative end cannot adapt to the social requirements of the population and, thus, constitute the ASD group (Constantino and Todd, 2003; Robinson et al., 2011; Lundström et al., 2012). Accordingly, it is widely known that ASD occurs as a spectrum in the diagnosed population; moreover, this spectrum is also observed among the general population (Baron-Cohen, 1995; Lundström et al., 2012; Ruzich et al., 2015). There is a genetic and biological overlap in the etiology of ASD and autistic traits (Bralten et al., 2018). Therefore, individuals at the end of this spectrum, with a high level of deficits, constitute the ASD group. Consistent with this view, studies examining autistic traits in the healthy population have been increasing in recent years. In addition, studies demonstrating differences in humor styles and appreciation (Eriksson, 2013; Rawlings, 2013) among healthy people with autistic traits similar to those in the ASD population have been reported.

To date, several instruments have been developed to assess the various dimensions of humor. These can be divided into two main groups: questionnaires and performance tests. As questionnaires are not relevant to the aim of this study, they are not discussed here (for a list of questionnaires, refer Ruch, 2007). Approximately 18 performance tests, with different measurement aims covering humor comprehension, appreciation, reasoning, and motivation, are have been constructed (for a list, see Ruch, 2007). None of these tests were developed directly for ToM assessment. However, studies assessing ToM using cartoons and jokes in patient populations and neurotypical individuals, which are not structured and have not been validated, have been reported (Happé et al., 1999; Gallagher et al., 2000; Samson et al., 2008). The stimuli are mostly unstructured cartoons or jokes that are used only in one study, reducing the possibility of replication.

Cartoons may be classified as one particular type of humorous material, i.e., static visual stimuli, and can be described as jokes in pictorial form (Nilsen and Nilsen, 2000). Cartoons may either consist of both text and pictures, or only pictures. The advantages of cartoons are that they do not depend merely on linguistic abilities, but also enable the depiction of characters' emotions via their facial expressions or body postures. In contrast, in verbal humor, characters' emotions must be described explicitly (Hempelmann and Samson, 2008). Henceforth, we will discuss studies using cartoons for ToM assessment, as our test consisted only of cartoons.

Cartoons with ToM content, which have been used in more than one study, were developed by Gallagher et al. (2000) and Marjoram et al. (2006). The stimuli consisted of cartoons in three categories; ToM, non-ToM, and jumbled pictures. The cartoons were grouped into categories by researchers and applied to 20 people before actual use. Meaning inference was assessed by open-ended questions and scored by a researcher as correct or incorrect. Another unstructured cartoon set was developed by Happé et al. (1999) and Snowden et al. (2003). Similarly to the above-mentioned study, the cartoons were divided into two categories (physical state and ToM) by researchers. Meaning inference for cartoons was assessed by open-ended questions and scored by researchers (Happé et al., 1999). One final example are the cartoons used by Samson et al. (2008). In this study, selected cartoons were pre-examined in several ways. Cartoons were categorized by five people in three categories as puns, those involving ToM, or as semantic, and cartoons with 90% total agreement were put into the related category. Twenty-one participants rated cartoons for funniness, complexity, and originality, with categories balanced regarding these parameters (Samson et al., 2008).

The research to date indicates that cartoon-based ToM assessment may be very useful; however, a structured, reliable, and validated test is currently not available. In addition, humor is a useful tool assessing ToM in healthy adults without a ceiling effect (Adolphs, 2003). Moreover, humor and ToM problems seem to co-occur, as seen in schizophrenia and ASD populations (Bozikas et al., 2007; Samson and Hegenloh, 2010). Hence, a test that measures both humor and ToM would be useful. Measuring both in the same test will provide an opportunity to understand whether these processes are disabled independently or in relation to each other. Finally, there is no structured humor test currently validated for use with a Turkish population.

Based on these demands, in the present study, we aimed to develop a humor test that measures humor comprehension and appreciation using cartoons with and without ToM content. More specifically, we aimed to create a task that: (i) was sensitive to differences in ToM ability in the healthy adult population, without a ceiling effect; (ii) was able to measure humor comprehension and appreciation ability with and without ToM ability; (iii) has adequate psychometric properties, being both reliable and valid; (iv) was objectively scored; and (v) was easy and quick to apply. Cartoons were presented, and time taken to decide whether the cartoon was funny or not (i.e., reaction time), scoring of funniness level (i.e., funniness score), answers for meaning of cartoons (i.e., meaning-inference score) were collected. The test was validated in relation to autistic traits.

METHODS AND RESULTS

Participants

A total of 147 (79 females and 68 males, mean age = 22.56 years, *SD* = 4.41 years), undergraduate or graduate students from different faculties participated. As humor appreciation and comprehension change with aging (Greengross, 2013) we only included younger adults in order to constitute a more homogenous sample. The inclusion criterion was an age ranging between 18 and 35 years, and exclusion criteria were uncorrected visual impairment, a diagnosed neuropsychiatric disorder, and taking neuropsychiatric medication. The study was approved by the Ethical Committee of Ankara University School of Medicine.

Test Development

The study was conducted in a series of four steps; for simplicity, the methods and results for each step are presented together. The first step of test development consisted of the selection of cartoons and piloting. The second step comprised experts grouping the cartoons. In subsequent steps three and four, reliability and validity were analyzed. In the test, three parameters for cartoons were assessed: reaction time (time taken to decide whether the cartoon was funny or not), funniness score (scoring of funniness level), and meaning-inference score (correct answers for meaning of cartoons). Confirmatory factor analysis was performed using AMOS 21.0 (Arbuckle, 2012), and all other analyses were performed using SPSS version 20.0 software (IBM Corp., 2011).

Step 1: Cartoon Selection

As either preference or dislike for sexual cartoons is known to correlate with personality characteristics (Ruch and Hehl, 1998) and was detected in all the factor analytic studies independent of the structural content (Eysenck, 1942; Herzog and Larwin, 1988; Ruch and Hehl, 1998), cartoons with high sexual content were excluded. Cartoons with low sexual, political, and violence content were collected from printed media or the internet. Colored cartoons were converted to black and white to exclude the facilitating effect of color on object recognition (Rossion and Pourtois, 2004), as this might cause a difference between colored and non-colored cartoons' reaction times. Written cartoons with more than 70 characters were excluded as reading speed might have interfered with reaction times.

Participants were given instructions comprising a two-cartoon demo test, before the actual test session. Funniness scores and reaction times were collected using a computer. Funniness evaluation part of the test was presented in a dimly lit, soundproof room using a laptop with a 15.6″, 1366 × 768 pixel resolution screen. MATLAB R2013a (MathWorks) with Psychtoolbox 3.0 (Kleiner et al., 2007) was used for presenting the cartoons and to record the reaction times and funniness scores. Participants were instructed as "A cartoon will appear on the screen and click to mouse when you decide if the cartoon is funny or not. Then a second screen will appear with numbers from one to seven, you should rate the funniness using a scale from one not funny to seven extremely funny." Cartoons were projected on a gray background in randomized order. The time duration between cartoon presentation and mouse click was recorded as reaction time in seconds. The next screen display consisted of numbers from 1 to 7 and the words "Evaluate funniness level." After the funniness level had been chosen, a new cartoon appeared on the screen.

The second part of the test is the meaning-inference test. It is a paper-based test and cartoons are presented in a booklet with one cartoon per page and with the question "Which one of the following represents the meaning of the cartoon most?" followed by four choices (see **Figure 1** for an example).

Step 1.1: Pilot study 1

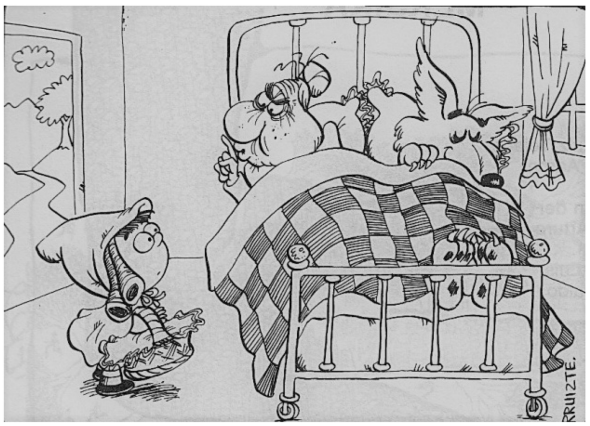
A pilot sample of 12 individuals participated in this step (six females and six males, mean age = 23.75 years, *SD* = 3.33 years). Sixty cartoons were shown to the participants. The mean funniness-score was 3.07 ± 0.84 (range [1.50; 4.92]). Cartoons of funniness scores less than 2.5 (n = 22) were eliminated as they were considered to be unfunny for the target population. Mean reaction-time was 7.59 ± 2.30 s (range [3.61; 13.81]). Cartoons with a reaction time of greater than 12 s (n = 3) were eliminated due to their longer processing time, which may have indicated that they were more complex than the other cartoons. New cartoons were added to replace those removed. The next step then commenced with 60 cartoons.

Step 1.2: Pilot study 2

Thirty-two people participated in the second pilot study (16 females and 16 males, mean age = 26.63, *SD* = 5.11). Each participant scored the cartoons for funniness. In addition, each participant evaluated 20 cartoons for familiarity and meaning (10 evaluations per cartoon). They then answered two questions: the first question was "Have you seen this cartoon before?" and the second was "Write down the meaning of the cartoon in a single sentence. If you do not think it makes sense at all, you may write down 'meaningless'."

The maximum familiarity was 4/10, and cartoons with a familiarity of 2/10 or more (n = 20) were discarded. The study then continued with the remaining 40 cartoons. The mean funniness score and reaction time of each cartoon were calculated. Reaction times with z-scores over ± 3.0 were assigned as outliers as participants might have paused during the test or might have been distracted, and those values were excluded from the analysis. The mean reaction time was 7.03 s (N = 130–135, *SD* = 1.74, range [3.72; 10.73]). The mean funniness score was 3.57 (N = 135, *SD* = 0.46, range [2.65; 4.60]).

For the first phase of meaning-inference test development, answers to the above mentioned question were collected. In the second phase, four options were created by researchers based on these answers. Four options were designed as follows: one option was the main meaning, two options were secondary meanings, and one option was "meaningless." The place of the main meaning was randomized in the first three choices and the fourth choice was always "meaningless." The main meaning option was considered the correct answer. Every



1. Which one of the following represents the meaning of the cartoon most?

a) Grandma does not want the beautifully sleeping wolf to wake up

b) Grandma doesn't want the red riding hood to tell anyone that she was in bed with the wolf

c) Grandma wants the red riding hood to go away to protect herself from the wolf

d) Meaningless

FIGURE 1 | Sample question from meaning-inference test. Option b represents the main meaning. Options a and c are secondary meanings. Option d is "meaningless." Cartoon is reprinted with permission from the Aydın Doğan Foundation. Copyright© 1986, Aydın Doğan Foundation.

correct answer was scored as 1 point. The total number of points was referred to as the "meaning-inference score." The meaning-inference test was presented to participants in two orders in opposite directions to eliminate the possible confounding effect of losing concentration toward the end. The test was applied to a pilot group of 10 participants. For three cartoons, the targeted choice was chosen by fewer than 50% of the participants and, thus, the options have been rearranged.

Step 2: Grouping of Cartoons Depending on ToM Content: The Experts' View

To group cartoons into the two categories as ToM/Non-ToM, nine experts with at least a doctoral degree (social psychology n = 4, clinical psychology n = 2, developmental psychology n = 1, physiology n = 2) answered inquired to answer either yes or no to the question: "Do you think that social relations, values, feelings, and thoughts of people need to be understood in order to understand this cartoon?" The number of experts was chosen as an odd number, as this will always result in predominance of either the "yes" or "no" answer. Cartoons were assigned to the ToM category if the majority decision was "yes." Cartoons with a majority of "no" were assigned to the Non-ToM (N-ToM) category. A Mann–Whitney U test indicated that the amount of "yes" votes that were associated with the ToM subscale (Mdn = 7) was significantly higher than the amount of "yes" votes that were associated with the N-ToM subscale (Mdn = 3), U = 0, $p < 0.001$. The ToM group consisted of 27 cartoons, whereas the N-ToM group consisted of 13 cartoons.

Step 3: Reliability

A group of 103 people participated in this part of the study (57 females and 46 males, mean age = 19.68, SD = 1.85). The funniness scores of 32 participants, who took part in pilot study 2 were included in the analysis. Therefore, data obtained from 135 participants (73 females and 62 males, mean age = 21.33, SD = 4.20 years) were used for reliability analysis.

The reliability of the subscales was assessed by three methods. First, Cronbach's alpha coefficients were calculated as a measure of internal consistency (Cronbach and Meehl, 1955). For Cronbach's alpha, values over 0.70 are accepted as good (Streiner and Norman, 1995; Kline, 2000). The Cronbach's alpha coefficient was 0.84 for the N-ToM group and 0.94 for the ToM group, indicating good internal consistency. Both subscales showed good reliability.

Second, split-half reliability was used as another measure of internal reliability. A basic assumption of split-half reliability is that the two halves of the test should yield similar true scores and error variances (Brown, 1910; Spearman, 1910). In each subgroup, cartoons were divided into two groups (even- and odd-numbered). Spearman–Brown coefficients (r_{sb}) were calculated. The coefficient for the N-ToM group was r_{sb} = 0.83, and for the ToM group it was r_{sb} = 0.95, indicating good consistency.

The third method involved the calculation of item-total correlations. Descriptive statistics, corrected item-total correlations of all items can be found in **Table 1**. The correlation coefficient is expected to be positive, above 0.30 (Nunnally and Bernstein, 1994). As shown in **Table 1**, item-total correlations

TABLE 1 | Psychometric characteristics of the scale.

Item	Subscale	*M*	*SD*	CITC	SRC	SMC
1	N-ToM	3.26	1.89	0.54	0.58	0.34
2	N-ToM	2.91	1.80	0.58	0.64	0.42
3	ToM	2.96	1.72	0.62	0.67	0.44
4	ToM	3.36	1.70	0.62	0.66	0.43
5	ToM	2.72	1.53	0.56	0.60	0.36
6	N-ToM	3.38	1.83	0.59	0.62	0.39
7	ToM	3.99	1.95	0.57	0.59	0.34
8	ToM	3.43	1.95	0.51	0.54	0.29
9	ToM	3.75	1.87	0.68	0.68	0.46
10	ToM	3.69	1.66	0.61	0.61	0.38
11	N-ToM	3.12	1.84	0.48	0.48	0.23
12	ToM	3.34	1.77	0.69	0.71	0.50
13	N-ToM	4.12	1.81	0.45	0.47	0.22
14	ToM	3.50	1.89	0.64	0.66	0.43
15	ToM	3.57	1.82	0.64	0.62	0.39
16	ToM	3.60	1.90	0.62	0.62	0.39
17	ToM	3.86	1.80	0.62	0.62	0.39
18	ToM	3.69	1.94	0.60	0.62	0.39
19	ToM	4.04	1.82	0.70	0.71	0.50
20	N-ToM	3.40	1.88	0.55	0.60	0.36
21	ToM	4.58	1.85	0.52	0.57	0.32
22	ToM	3.56	1.76	0.51	0.54	0.29
23	N-ToM	2.66	1.57	0.48	0.50	0.25
24	ToM	3.81	1.68	0.61	0.62	0.39
25	ToM	3.80	1.90	0.58	0.60	0.35
26	ToM	3.72	1.91	0.59	0.61	0.37
27	N-ToM	3.60	2.00	0.36	0.48	0.23
28	ToM	3.59	1.76	0.48	0.50	0.25
29	ToM	4.57	1.74	0.63	0.64	0.41
30	N-ToM	2.93	1.69	0.58	0.58	0.33
31	N-ToM	4.00	1.77	0.50	0.52	0.28
32	N-ToM	3.28	1.88	0.46	0.51	0.26
33	ToM	4.17	1.97	0.62	0.62	0.39
34	N-ToM	4.04	1.90	0.54	0.60	0.37
35	ToM	3.95	1.89	0.49	0.52	0.27
36	ToM	3.83	1.87	0.64	0.67	0.45
37	ToM	3.42	1.98	0.62	0.64	0.41
38	ToM	2.78	1.68	0.57	0.55	0.31
39	ToM	3.10	1.75	0.57	0.59	0.35
40	N-ToM	3.94	1.81	0.62	0.69	0.48

N = 134; M, mean; SD, standard deviation; CITC, corrected item-total correlation; SRC, standardized regression coefficient; SMC, squared multiple correlation; N-ToM, non-theory of mind; ToM, theory of mind.

were above 0.30 and positive, demonstrating the consistency of each item.

Step 4: Validity

Step 4.1: Construct validity

To evaluate the construct validity of the resulting model, confirmatory factor analysis was performed on the funniness scores. Data from 135 participants (the sample that was regarded within the reliability analysis) were used for confirmatory factor analysis.

Bartlett's Sphericity Test and Keiser–Meyer–Olkin (KMO) were calculated as measures of the suitability of data for structure detection. For data to be considered suitable, the Bartlett's test should be significant and the KMO value should be over 0.80 (Bartlett, 1954; Kaiser and Rice, 1974). The data were suitable for factoring as the Bartlett's test was significant ($p < 0.001$) and the KMO value was 0.90.

According to the Mahalanobis distance measure, one participant was detected as a multivariate outlier and, thus, was excluded from the sample (Mahalanobis, 1936). Fit indices were estimated using the unweighted least-squares (ULS) method as

TABLE 2 | Descriptives of test scores and reaction times for main study group.

Score	Subscale											
	N-ToM						ToM					
	Min	Max	*M*	*SD*	95% CI		Min	Max	*M*	*SD*	95% CI	
					LL	UL					LL	UL
Funniness	10	63	33.25	11.75	30.96	35.55	30	151	91.15	30.57	85.17	97.12
Reaction time[a]	2.11	12.67	7.08	2.11	6.66	7.49	2.49	12.92	7.21	1.74	6.87	7.55
Meaning-inference	2	10	6.67	2.00	6.28	7.06	5	24	17.68	3.53	16.99	18.37

N = 135; N-ToM, non-theory of mind; ToM, theory of mind; Min, minimum; Max, maximum; M, mean; SD, standard deviation; CI, confidence interval; LL, lower limit; UL, upper limit; a, reaction time is in seconds.

TABLE 3 | Descriptives and comparison of ToM-HCAT scores and reaction times for cartoons in subscales.

Score	Subscale											
	N-ToM (*n* = 10)					ToM (*n* = 25)						
	Min	Max	*M*	*SD*	Median	Min	Max	*M*	*SD*	Median	*U*	*p*
Funniness	2.66	4.04	3.36	0.49	3.32	2.72	4.58	3.65	0.46	3.69	85.5	0.151
Reaction time[a]	4.37	9.81	6.96	1.97	6.85	3.72	10.13	7.08	1.68	7.49	120.0	0.872
Meaning-inference[b]	47	89	67.3	12.58	69.50	47	90	71.32	13.31	72.00	100.0	0.377

N-ToM, non-theory of mind; ToM, theory of mind; Min, minimum; Max, maximum; M, mean; SD, standard deviation; CI, confidence interval; LL, lower limit; UL, upper limit; a, reaction time is in seconds; b, correct answer percentage of cartoons.

kurtosis (163.23) and critical ratio (16.30) values suggested a non-normal distribution and data were ordinal in structure (de los Ángeles Morata-Ramírez and Holgado-Tello, 2013).

The assessment of model fit was based on several indices. The goodness-of-fit (GFI), adjusted goodness-of-fit (AGFI), standardized root mean square residual (SRMR), normed-fit index (NFI), and Bollen's relative fit index (RFI) were used. The absolute fit indices (GFI and AGFI) calculate the proportion of variance that is accounted for by the model covariance. The SRMR shows the difference between the residuals of the sample covariance matrix and the hypothesized covariance model. The NFI shows the fit of the estimated model with the hypothesized model, and RFI considers inconsistency between the two models (Hooper et al., 2008). For GFI, AGFI, NFI, and RFI scores, >0.95 suggests a good fit whereas scores that are >0.80 suggest an acceptable fit (Bentler and Bonett, 1980; Bollen, 1989; Jöreskog and Sörbom, 1996). A SRMR <0.05 suggests a good data-model fit, while <0.08 suggests an acceptable fit (Hu and Bentler, 1999). Fit indices for the initial model are as follows: CFI = 0.97, AGFI = 0.97, NFI = 0.96, RFI = 0.96, SRMR = 0.070. The GFI, AGFI, NFI, and RFI suggested a good fit, and the SRMR suggested an acceptable fit.

Standardized regression coefficients of the variables were expected to be over 0.40, and all items were above that value. Squared multiple correlations should be over 0.30 but can be tolerated toward 0.10 if the other values are acceptable. Standardized regression coefficients and squared multiple correlations for the items are presented in **Table 1**. All regression values were above the expected value of 0.40. Squared multiple correlation values of 10 items were near but below 0.30. Items with low correlations were excluded from the model one by one, and fit indices were calculated. Items with lower indices were left out of the model. Three cartoons from the N-ToM group and two cartoons from the ToM group were excluded. The final model fit indices are as follows: CFI = 0.97, AGFI = 0.97, NFI = 0.97, RFI = 0.97, SRMR = 0.067. All the indices suggested an acceptable to good fit. The final test consisted of 35 cartoons: 10 cartoons from the N-ToM and 25 cartoons from the ToM.

Descriptives of test scores and reaction times for participants in this study group can be found in **Table 2**.

Descriptive statistics for the ToM-HCAT scores and reaction times, as well as comparisons of the subgroups, are given in **Table 3**. **Table 3** shows that no difference was found for funniness score, reaction time and meaning-inference score between cartoons in ToM and N-ToM subscales.

Step 4.2: External validity

In the main study group, participants (n = 103) completed the Turkish version of the Autism Spectrum Quotient (AQ; Kose et al., 2010) for calculation of autistic trait scores in addition to the humor test. The maximum score for the AQ is 50 points; higher scores indicate higher levels of autistic traits (Baron-Cohen et al., 2001b). As the ASD group constitutes the higher end of the distribution for autistic traits (Robinson et al., 2011; Lundström et al., 2012), we adopted a similar approach in the present sample. Participants with AQ scores of +0.5 standard deviations were grouped as the high-autistic traits group (n = 37, mean AQ = 24.32, SD = 2.21, range [22; 28]). The rest of the population constituted the low-autistic traits group (n = 66, mean AQ = 16.58, SD = 3.14, range [6; 21]). A chi-square test was

TABLE 4 | Descriptives and comparison of ToM-HCAT scores for high- and low-autistic traits groups.

Score	Low autistic traits (n = 66)					High autistic traits (n = 37)						
	Min	Max	*M*	*SD*	Median	Min	Max	*M*	*SD*	Median	*U*	*p*
Funniness												
N-ToM	10	52	33.67	11.54	34.5	15	63	32.51	12.25	30.0	1105.5	0.427
ToM	30	151	92.79	31.58	97.0	41	140	88.22	28.85	88.0	1106.0	0.429
Meaning-inference												
N-ToM	2	10	6.83	1.94	7.0	3	10	6.38	2.09	7.0	1068.5	0.289
ToM	5	23	18.08	3.68	19.0	11	24	16.97	3.17	17.0	914.5	**0.034**

N-ToM, non-theory of mind; ToM, theory of mind; Min, minimum; Max, maximum; M, mean; SD, standard deviation; CI, confidence interval; LL, lower limit; UL, upper limit. Bold text indicates a statistically significant difference with a p-value less than 0.05.

TABLE 5 | Descriptives and comparison of reaction times for high- and low-autistic traits groups.

Subcategory	Low autistic traits			High autistic traits			95% CI				
	M	*SD*	*n*	*M*	*SD*	*n*	LL	UL	*t*	*df*	*p*
N-ToM	6.78	2.00	66	7.60	2.24	37	−1.70	0.07	−1.85	67.64	0.069
ToM	6.97	1.68	66	7.64	1.80	37	−1.40	0.05	−1.87	70.15	0.066

N-ToM, non-theory of mind; ToM, theory of mind; Min, minimum; Max, maximum; M, mean; SD, standard deviation; CI, confidence interval; LL, lower limit; UL, upper limit.

performed, and no difference was found for gender between high autistic traits (27:39 [m:f]) and low autistic traits (19:18 [m:f]) groups, χ^2 (1, N = 103) = 1.05, p = 0.306. A Mann–Whitney U test indicated that age for the low-autistic traits group (Mdn = 19) was not significantly different from that for the high autistic traits group (Mdn = 20), U = 1119.5, p = 0.472.

The funniness-score and meaning-inference score on each of the two subscales were calculated as the sum of scores on that category. Scores were compared between groups and the means, 95% confidence intervals and comparisons of ToM-HCAT scores can be found at **Table 4**. **Table 4** shows that the meaning-inference score for the ToM category was lower for the high-autistic traits group (Mdn = 17) than for the low-autistic traits group (Mdn = 19); U = 914.5, p = 0.034. There was no difference for N-ToM and ToM funniness scores or N-ToM meaning-inference scores. The Spearman correlation between ToM meaning-inference scores and autistic traits scores was calculated and found to be low and non-significant, $r_s(102) = -0.14$, p = 0.163. To test the robustness of this result the high-end split was further shifted to +1.0 SD. The high-autistic traits group (n = 19, mean AQ = 26.16, SD = 1.50, range [24; 28]) and the low-autistic traits group (n = 84, mean AQ = 17.82, SD = 3.68, range [6; 23]) were compared regarding the respective ToM meaning-inference scores. A Mann–Whitney U test indicated that the meaning-inference score for the ToM category was lower for the high-autistic traits group (Mdn = 17) than for the low-autistic traits group (Mdn = 19); U = 551.0, p = 0.035. Meaning-inference score of the ToM category was compared between females (Mdn = 18) and males (Mdn = 18) and there was no difference for gender U = 1180.0, p = 0.383.

Reaction times for both categories were compared between groups. Descriptives and comparison of reaction times for high- and low-autistic traits groups can be found in **Table 5**. Results showed that the high-autistic traits group had longer reaction times for both subscales; however, no statistical difference existed between the low- and high-autistic traits groups. As reaction times might have been influenced by the number of characters in speech bubbles or by the amount of text in the cartoons, the relationships between text length and reaction time was analyzed. In the N-ToM subscale, 5/10 cartoons had speech bubbles or text. Similarly, in the ToM subscale, there were 25 cartoons, of which 10 featured speech bubbles. A Spearman correlation analysis between character count and reaction time showed a moderate positive correlation, $r_s(34) = 0.39$, p = 0.022. Accordingly, the differences between character counts and reaction times for the N-ToM and ToM subscales were analyzed. A Mann–Whitney U test indicated that character count for the ToM subscale (Mdn = 0) was not significantly different from that for the N-ToM subscale (Mdn = 4.50), U = 120.5, p = 0.872. Similarly, the reaction time for the ToM subscale (Mdn = 7.49) was not significantly different from that for the N-ToM subscale (Mdn = 6.85, U= 120.0, p = 0.872).

DISCUSSION

In this study, we developed and validated a humor test, the ToM-HCAT, to assess humor appreciation and comprehension via the use of cartoons. This test comprises two different subscales: one subscale with ToM content and one subscale without ToM content. This theoretically assumed two-dimensional structure was analyzed by confirmatory factor analysis. The data showed an acceptable-to-good model fit, indicating good construct validity. Reliability measures were good and external validity was evident.

The ToM-HCAT is a performance test consisting of 35 cartoons, and has three outputs: (i) reaction time taken to decide whether the cartoon is funny or not; (ii) funniness score for each cartoon and subscale; and (iii) meaning-inference score for each cartoon and subscale. The reaction time reflects the processing speed of humor appreciation. The funniness score represents humor appreciation, and the meaning-inference score indicates humor comprehension. Within the ToM subscale the meaning-inference score reflects ToM ability by means of humor comprehension.

The test comprises cartoons with or without speech bubbles. In the first subscale, half (n = 5/10) of the cartoons have speech bubbles; in the second subscale, 10 out of 25 cartoons had speech bubbles. The distribution of cartoon types in groups is similar. For all cartoons, the text was limited to a maximum of 70 characters to exclude the effect of reading speed on reaction times. Further, this allows for reduced linguistic demands for comprehension. Although there was a moderate correlation between character counts and reaction time, no difference for character counts between subscales was observed, which makes it possible to compare these. All cartoons are black and white to exclude the confounding effect of color, especially on funniness scores and reaction times. The cartoons were chosen randomly from a large pool. The internet, printed cartoon books of Turkish cartoonists, and yearly books of the "Simavi International Cartoon Competition" (1983–1993) were used. Eighty-five cartoons were used in the study, and the final test consists of 35 cartoons; among these, 16 were published by international cartoonists, which also enables adaptation to other cultures.

Funniness decision consists of both humor comprehension and appreciation processes, although for appreciation it is not always necessary to comprehend (e.g., non-sense humor) (Ruch and Hehl, 1998). On the other hand, meaning-inference involves only the humor comprehension process. Accordingly, funniness and meaning-inference scores of ToM-HCAT should be considered as representing linked but different processes. In the meaning-inference test, participants choose the meaning from four options and it might happen that they comprehend the meaning after seeing the choices. In addition, it is possible that they may not have understood the main meaning of the cartoon in the previous funniness test. In our opinion, this does not decrease the importance of either result. This is because a funniness decision can be independent of comprehension; further, individuals may be unable to comprehend even after seeing the choices. Supporting this hypothesis, in the present study none of the participants achieved the maximum score for the meaning-inference test.

In addition to this, there are advantages to using a forced choice test. Current cartoon sets used in studies use subjective evaluations in which the researcher scores participants' open-ended answers (Happé et al., 1999; Gallagher et al., 2000). The fact that the choices were selected by researchers in our study may be questioned; however, the choices were created after collecting explanations from a pilot group. Furthermore, there is no inter-rater reliability problems in the multiple-choice method. Inter-rater reliability refers to how similar the data collected by different raters are. If raters do not consistently agree in their scoring, then examiner specific factors may contribute unduly to observed score variability (Kline, 2011).

Another output of the test are reaction times for the funniness ratings, which provide the opportunity for evaluating the decision time. Decision time may be affected by cognitive processing speed, serving as a possible indicator of the efficiency of these processes. However, it should be noted that reaction time might have been influenced by numerous factors. For example, the complexity of the cartoons might have influenced the processing time. In the present study, we excluded cartoons for which reaction times were very long as such cartoons might have been overly complex. Another pitfall might have been that participants took a break or were distracted. To prevent this, we excluded reaction times with a very high z-score from the analysis. In our comparison group with autistic traits, reaction times were higher for the high-autistic traits group on both subscales; however, this difference was not significant. Similar results are presented in a cartoon Faux Pas Test. ASD participants took longer than neurotypicals to give their responses independent of cartoon types (Thiébaut et al., 2016). Longer reaction times might be related to the higher detail orientation of individuals with symptoms of autism (Dakin and Frith, 2005; Samson and Hegenloh, 2010). This finding is also an indicator that reaction times may be useful for measuring cognitive processing differences.

In the present study, we showed that individuals with higher autistic traits exhibit poorer humor comprehension if ToM is necessary for understanding the cartoon. It is widely known that the humor response of individuals with ASD differs from the response of neurotypical participants (Van Bourgondien and Mesibov, 1987; Baron-Cohen, 1997; Reddy et al., 2002; Samson and Hegenloh, 2010; Samson et al., 2013). This finding may be interpreted as a result of social communication deficits observed in this disorder (American Psychiatric Association, 2013). Regarding individuals with ASD, the response to humor varies according to the type of humor. Researchers have shown that ASD individuals do not appreciate humor created by socially inappropriate behavior (Reddy et al., 2002), and are unable to readily understand the other person's humorous intention (Baron-Cohen, 1997). High-functioning autistic individuals may make jokes based on lexical or phonological contradictions; however, these tend to be under the age-appropriate level (Van Bourgondien and Mesibov, 1987). In support of our results, a study by Samson and Hegenloh (2010) showed that adults with ASD enjoy visual and semantic pun cartoons at similar levels as neurotypical individuals; however, these individuals exhibit difficulty in understanding ToM cartoons and provide less mentalistic explanations to humor consisting of ToM. In another study, it was shown that adolescents with high-functioning autism or AS performed worse than neurotypical individuals regarding the comprehension of cartoons and jokes (Emerich et al., 2003). We could not show a correlation between meaning-inference scores and AQ; however, this may have arisen from the relatively small sample size. Analysis using a higher number of participants may reveal a significant difference.

As the present population consisted of healthy people with autistic traits without a diagnosis of ASD, this study shows that impairment in ToM and humor extends to the healthy population with autistic traits. In support of this, differences in humor styles and appreciation have been reported among healthy individuals with autistic traits (Eriksson, 2013; Rawlings, 2013). For ToM impairment in healthy individuals, variation was shown by an implicit test: the Reading the Mind in the Eyes test. The test results were negatively correlated with autistic trait scores measured by the AQ (Baron-Cohen et al., 2001a). This result supports the present findings. We found that individuals with higher autistic traits score, as measured by AQ, exhibited poorer comprehension of cartoons with ToM, but not of cartoons without ToM.

To the best of our knowledge, there is no existing test that has the same structure and outputs as our humor test. The most similar test is the 3WD test of humor appreciation (Ruch, 1992): 3WD is another performance test that measures funniness and aversion to cartoons and jokes on a seven-point scale, with 35 items. Three categories of humor are present: non-sense, incongruity-resolution, and sexual. Although the tests are similar in the methods used to measure funniness, there are some differences between the ToM-HCAT and the 3WD test. The most important is that the present test aims to measure ToM processing, and cartoons have accordingly been categorized by their ToM content. To our knowledge, there is no other structured psychometric test that measures ToM ability by humor. A second difference is that we also measured humor comprehension in addition to appreciation.

As our aim was to develop a humor test that measures ToM ability, the finding of lower comprehension scores on ToM subscale cartoons for individuals with high autistic traits supports the validity of our test. Although a difference in funniness scores on the ToM subscale would be expected, we could not find any difference related to autistic traits. In contrast to the present results, a previous study mentioned above reported a significant difference in funniness among AS participants (Samson and Hegenloh, 2010). Although humor comprehension (resolution of incongruity) is considered a prerequisite for humor appreciation (Shultz, 1972; Suls, 1972), it has been suggested that only the detection of incongruity is necessary. This is supported by the appreciation of non-sense or slapstick humor, which does not involve incongruity resolution (Ruch and Hehl, 1998). Therefore, the lack of difference in funniness scores despite the low comprehension scores for the ToM category could be explained by this theory. It is proposed that individuals with high levels of autistic traits find incongruity sufficient for funniness, or that such individuals may perceive a different incongruity and/or resolution. Another difference with the current literature is that we could not show the gender difference in meaning-inference scores for ToM subscale. It is shown that women are superior compared to men in adult ToM tests (Baron-Cohen, 2002). However, in the study by Russell et al. (2007), men showed superior performance compared to women on both physical and mental state cartoons. The results emphasize the hypothesis that the differences in ToM tests could be task specific.

As our starting point was to develop a test to measure variability without a ceiling effect for ToM abilities in the adult healthy population, the findings suggest that our test can detect variability of ToM. Result cannot be attributed to humor ability, because the comprehension scores in the Non-ToM subscale did not show a difference in relation to high or low levels of autistic traits. None of the present participants achieved the perfect score of 25 out of 25 possible points on comprehension for the ToM subscale of the test; further, they scored almost the full range of possible scores of between 5 and 24 points, with a slightly left-skewed distribution. This variation suggests that the ToM-HCAT is sensitive to individual differences in ToM ability. This sensitivity in comparison with other tests may be attributable to the more real-world orientation of cartoons. Cartoons could be regarded as complex social scenarios that require social knowledge, and participants are required to make inferences about their meaning by both explicit mental state reasoning and spontaneous mental state inference. Furthermore, cartoons represent stimuli encountered in daily life.

Limitations

There are several limitations to this study. First, the mean age of participants in the CFA analysis was 21.33 years, with the majority being between 18 and 22 years of age. Second, all participants were undergraduate/graduate students. The use of a more diverse sample is expected to enhance the validity of the current results. In particular, the age range should be wider. Another limitation is the application of the meaning-inference test on paper as a separate test. It may be beneficial to perform this test using a computer to ensure continuity of the entire test. Moreover, reaction times to meaning decisions should be collected, as these may be informative of processing time differences for ToM between individuals with high and low levels of autistic traits. Lastly, studies with larger numbers of participants are required. The present test was not validated using a diagnosed ASD population; this may appear to represent a limitation as this would be a gold standard for ToM disability. However, we validated the ToM-HCAT with autistic traits, which are more subtle than in individuals diagnosed formally with ASD. We showed that the test enables differentiation between these groups, thereby demonstrating its high sensitivity. Moreover, this test was developed to assess variations in ToM ability among the general population. This study, considering the small sample size, should be considered the first step of a scale development process. In future studies, a cross validation phase with a second and larger sample is necessary.

CONCLUSION

In conclusion, a test for assessing ToM involving humor comprehension and appreciation was developed. The item and scale characteristics were good to excellent. The test was externally validated with autistic traits. It has multiple outputs

and is suitable for use in future ToM assessment studies, especially in the healthy population, as it is sensitive to variations in ToM ability among neurotypical individuals. This test is expected to deepen our understanding of differences in ToM ability in the healthy adult population.

AUTHOR CONTRIBUTIONS

SA: design of the study, data collection and analysis, writing and contribution to all parts of the paper. EN: conception and design of the study, supervision of all parts of this project, contribution to all parts of the paper. Both authors contributed to the writing of the manuscript, read it critically, and gave consent to its publication.

ACKNOWLEDGMENTS

The authors thank Derya Öztuna for commenting on statistical analysis of the project. The authors also thank the Aydın Doğan Foundation for reprint permission of the cartoon.

REFERENCES

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* 4, 165–178. doi: 10.1038/nrn1056

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596

Anselmetti, S., Bechi, M., Bosia, M., Quarticelli, C., Ermoli, E., Smeraldi, E., et al. (2009). "Theory" of mind impairment in patients affected by schizophrenia and in their parents. *Schizophr. Res.* 115, 278–285. doi: 10.1016/j.schres.2009.09.018

Arbuckle, J. L. (2012). *Amos (Version 21.0)*. Chicago, IL: IBM SPSS.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S. (1997). Hey! It was just a joke! Understanding propositions and propositional attitudes by normally developing children and children with autism. *ISR. J. Psychiatry Relat. Sci.* 34, 174–178.

Baron-Cohen, S. (2000). Theory of mind and autism: a review. *Int. Rev. Res. Ment. Retard.* 23, 169–184. doi: 10.1016/S0074-7750(00)80010-5

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends Cogn. Sci.* 6, 248–254. doi: 10.1016/S1364-6613(02)01904-6

Baron-Cohen, S., and Hammer, J. (1997). Parents of children with asperger syndrome: what is the cognitive phenotype? *J. Cogn. Neurosci.* 9, 548–554. doi: 10.1162/jocn.1997.9.4.548

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001a). The "Reading the mind in the eyes" test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry* 42, 241–251. doi: 10.1111/1469-7610.00715

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001b). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:1005653411471

Bartlett, M. S. (1954). A note on the multiplying factors for various $\chi 2$ approximations. *J. R. Stat. Soc. Ser. B* 16, 296–298. doi: 10.2307/2984057

Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88, 588–606. doi: 10.1037/0033-2909.88.3.588

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons, Inc. doi: 10.1002/9781118619179

Bora, E., and Pantelis, C. (2013). Theory of mind impairments in first-episode psychosis, individuals at ultra-high risk for psychosis and in first-degree relatives of schizophrenia: systematic review and meta-analysis. *Schizophr. Res.* 144, 31–36. doi: 10.1016/j.schres.2012.12.013

Bozikas, V. P., Kosmidis, M. H., Giannakou, M., Anezoulaki, D., Petrikis, P., Fokas, K., et al. (2007). Humor appreciation deficit in schizophrenia: the relevance of basic neurocognitive functioning. *J. Nerv. Ment. Dis.* 195, 325–331. doi: 10.1097/01.nmd.0000243798.10242.e2

Bralten, J., van Hulzen, K. J., Martens, M. B., Galesloot, T. E., Arias Vasquez, A., Kiemeney, L. A., et al. (2018). Autism spectrum disorders and autistic traits share genetics and biology. *Mol. Psychiatry* 23, 1205–1212. doi: 10.1038/mp.2017.98

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *Br. J. Psychol.* 3, 296–322. doi: 10.1111/j.2044-8295.1910.tb00207.x

Brüne, M. (2003). Theory of mind and the role of IQ in chronic disorganized schizophrenia. *Schizophr. Res.* 60, 57–64. doi: 10.1016/S0920-9964(02)00162-7

Chung, Y. S., Barch, D., and Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophr. Bull.* 40, 602–616. doi: 10.1093/schbul/sbt048

Constantino, J. N., and Todd, R. D. (2003). Autistic traits in the general population. *Arch. Gen. Psychiatry* 60, 524–530. doi: 10.1001/archpsyc.60.5.524

Corcoran, R., Mercer, G., and Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. *Schizophr. Res.* 17, 5–13. doi: 10.1016/0920-9964(95)00024-G

Coulson, S., Urbach, T. P., and Marta, K. (2006). Looking back: joke comprehension and the space structuring model. *Humor – Int. J. Humor Res.* 19, 229–250. doi: 10.1515/humor.2006.013

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.* 52, 281–302. doi: 10.1037/h0040957

Dakin, S., and Frith, U. (2005). Vagaries of visual perception in autism. *Neuron* 48, 497–507. doi: 10.1016/j.neuron.2005.10.018

de los Ángeles Morata-Ramírez, M., and Holgado-Tello, F. (2013). Construct validity of likert scales through confirmatory factor analysis: a simulation study comparing different methods of estimation based on pearson and polychoric correlations. *Int. J. Soc. Sci. Res.* 1, 54–61 doi: 10.11114/ijsss.v1i1.27

de Waal, F. B. M., and Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nat. Rev. Neurosci.* 18, 498–509. doi: 10.1038/nrn.2017.72

Emerich, D. M., Creaghead, N. A., Grether, S. M., Murray, D., and Grasha, C. (2003). The comprehension of humorous materials by adolescents with high-functioning autism and Asperger's syndrome. *J. Autism Dev. Disord.* 33, 253–257. doi: 10.1023/A:1024498232284

Eriksson, K. (2013). Autism-spectrum traits predict humor styles in the general population. *Humor* 26, 461–475. doi: 10.1515/humor-2013-0030

Eysenck, H. J. (1942). The appreciation of humour: an experimental and theoretical study. *Br. J. Psychol. Gen. Sect.* 32, 295–309. doi: 10.1111/j.2044-8295.1942.tb01027.x

Gallagher, H., Happé, F., Brunswick, N., Fletcher, P., Frith, U., and Frith, C. (2000). Reading the mind in cartoons and stories: an fMRI study of "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21. doi: 10.1016/S0028-3932(99)00053-6

Gokcen, S., Bora, E., Erermis, S., Kesikci, H., and Aydin, C. (2009). Theory of mind and verbal working memory deficits in parents of autistic children. *Psychiatry Res.* 166, 46–53. doi: 10.1016/j.psychres.2007.11.016

Greengross, G. (2013). Humor and aging – A mini-review. *Gerontology* 59, 448–453. doi: 10.1159/000351005

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., et al. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. *Brain* 125, 752–764. doi: 10.1093/brain/awf079

Happé, F. G. E. (1995). Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor Symb. Act.* 10, 275–295. doi: 10.1207/s15327868ms1004_3

Happé, F., Brownell, H., and Winner, E. (1999). Acquired "theory of mind" impairments following stroke. *Cognition* 70, 211–240. doi: 10.1016/S0010-0277(99)00005-0

Hempelmann, C. F., and Samson, A. C. (2008). "Cartoons: drawn jokes?," in *The Primer of Humor Research*, ed. V. Raskin (Berlin: Mouton de Gruyter), 609–640. doi: 10.1515/9783110198492.609

Herzog, T. R., and Larwin, D. A. (1988). The appreciation of humor in captioned cartoons. *J. Psychol.* 122, 597–607. doi: 10.1080/00223980.1988.9915534

Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., and D'Esposito, M. (2010). Neural activity during social signal perception correlates with self-reported empathy. *Brain Res.* 1308, 100–113. doi: 10.1016/j.brainres.2009.10.006

Hooper, D., Coughlan, J., and Mullen, M. (2008). Structural equation modeling: guidelines for determining model fit. *EJBRM.* 6, 53–60.

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model. A Multidiscip. J.* 6, 1–55. doi: 10.1080/10705519909540118

IBM Corp. Released (2011). *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY: IBM Corp.

Janssen, I., Krabbendam, L., Jolles, J., and van Os, J. (2003). Alterations in theory of mind in patients with schizophrenia and non-psychotic relatives. *Acta Psychiatr. Scand.* 108, 110–117. doi: 10.1034/j.1600-0447.2003.00092.x

Jöreskog, K. G., and Sörbom, D. (1996). *LISREL 8: User's Reference Guide*. Linconwood, IL: Scientific Software International.

Kaiser, H. F., and Rice, J. (1974). Little jiffy, mark IV. *Educ. Psychol. Meas.* 34, 111–117. doi: 10.1177/001316447403400115

Kleiner, M., Brainard, D. H., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What's new in psychtoolbox-3. *Perception* 36, 1–16

Kline, P. (2000). *The Handbook of Psychological Testing*. London: Routledge.

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modelling*. New York, NY: The Guilford Press.

Kose, S., Bora, E., Erermis, S., and Aydın, C. (2010). Psychometric features of Turkish version of autism-spectrum quotient. *Anatol. J. Psychiatry* 11, 253–260.

Lavoie, M.-A., Plana, I., Bédard Lacroix, J., Godmaire-Duhaime, F., Jackson, P. L., and Achim, A. M. (2013). Social cognition in first-degree relatives of people with schizophrenia: a meta-analysis. *Psychiatry Res.* 209, 129–135. doi: 10.1016/j.psychres.2012.11.037

Losh, M., and Piven, J. (2007). Social-cognition and the broad autism phenotype: identifying genetically meaningful phenotypes. *J. Child Psychol. Psychiatry* 48, 105–112. doi: 10.1111/j.1469-7610.2006.01594.x

Losh, M., Martin, G. E., Klusek, J., Hogan-Brown, A. L., and Sideris, J. (2012). Social communication and theory of mind in boys with autism and fragile x syndrome. *Front. Psychol.* 3:266. doi: 10.3389/fpsyg.2012.00266

Lundström, S., Chang, Z., Råstam, M., Gillberg, C., Larsson, H., Anckarsäter, H., et al. (2012). Autism spectrum disorders and autisticlike traits: similar etiology in the extreme end and the normal variation. *Arch. Gen. Psychiatry* 69, 46–52. doi: 10.1001/archgenpsychiatry.2011.144

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proc. Natl. Inst. Sci.* 2, 49–55

Marjoram, D., Gardner, C., Burns, J., Miller, P., Lawrie, S., and Johnstone, E. (2005). Symptomatology and social inference: a theory of mind study of schizophrenia and psychotic affective disorder. *Cogn. Neuropsychiatry* 10, 347–359. doi: 10.1080/13546800444000092

Marjoram, D., Job, D. E., Whalley, H. C., Gountouna, V.-E., McIntosh, A. M., Simonotto, E., et al. (2006). A visual joke fMRI investigation into theory of mind and enhanced risk of schizophrenia. *Neuroimage* 31, 1850–1858. doi: 10.1016/j.neuroimage.2006.02.011

Martin, R. A. (2007). *The Psychology of Humor: An Integrative Approach*. Burlington, MA: Elsevier Academic Press.

Nilsen, A. P., and Nilsen, D. L. F. (2000). *Encyclopedia of 20th-Century American Humor*. Phoenix, AZ: Oryx Press.

Nowicki, S., and Duke, M. P. (2001). "Nonverbal receptivity: The Diagnostic Analysis of Nonverbal Accuracy (DANVA)," in *The LEA Series in Personality and Clinical Psychology. Interpersonal Sensitivity: Theory and Measurement*, eds J. A. Hall and F. J. Bernieri (Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers), 183–198.

Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*. New York, NY: McGraw-Hill.

Rawlings, D. (2013). Humor preference and the Autism Quotient in an undergraduate sample. *Humor* 26, 411–421. doi: 10.1515/humor-2013-0028

Reddy, V., Williams, E., and Vaughan, A. (2002). Sharing humour and laughter in autism and Down's syndrome. *Br. J. Psychol.* 93, 219–242. doi: 10.1348/000712602162553

Regenbogen, C., Kellermann, T., Seubert, J., Schneider, D. A., Gur, R. E., Derntl, B., et al. (2015). Neural responses to dynamic multimodal stimuli and pathology-specific impairments of social cognition in schizophrenia and depression. *Br. J. Psychiatry* 206, 198–205. doi: 10.1192/bjp.bp.113.143040

Robinson, E. B., Koenen, K. C., McCormick, M. C., Munir, K., Hallett, V., Happé, F., et al. (2011). Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Arch. Gen. Psychiatry* 68, 1113–1121. doi: 10.1001/archgenpsychiatry.2011.119

Rossion, B., and Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. *Perception* 33, 217–236. doi: 10.1068/p5117

Ruch, W. (1992). "Assessment of appreciation of humor: Studies with the 3 WD humor test," in *Advances in Personality Assesment*, Vol. 9, eds C. D. Spielberg and J. N. Butcher (Hilsdale, NJ: Lawrence Erlbaum Associates, 27–75. doi: 10.5167/uzh-77946

Ruch, W. (2007). "Appendix: humor measurement tools," in *The Sense of Humor: Explorations of a Personality Characteristic*, ed. W. Ruch (Berlin: De Gruyter), 405–412. doi: 10.1515/9783110804607-019

Ruch, W., and Hehl, F.-J. (1998). "A two-mode model of humor appreciation: Its relation to aesthetic appreciation and simplicity-complexity of personality," in *The Sense of Humor: Explorations of A Personality Characteristic*, ed. W. Ruch (Berlin: De Gruyter), 109–142. doi: 10.1515/9783110804607.109

Russell, T. A., Tchanturia, K., Rahman, Q., and Schmidt, U. (2007). Sex differences in theory of mind: a male advantage on Happé's "cartoon" task. *Cognt. Emot.* 21, 1554–1564. doi: 10.1080/02699930601117096

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., et al. (2015). Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Mol. Autism* 6:2. doi: 10.1186/2040-2392-6-2

Samson, A. C., and Hegenloh, M. (2010). Stimulus characteristics affect humor processing in individuals with Asperger syndrome. *J. Autism Dev. Disord.* 40, 438–447. doi: 10.1007/s10803-009-0885-2

Samson, A. C., Huber, O., and Ruch, W. (2013). Seven decades after Hans Asperger's observations: a comprehensive study of humor in individuals with Autism Spectrum Disorders. *Humor* 26, 441–460. doi: 10.1515/humor-2013-0026

Samson, A. C., Zysset, S., and Huber, O. (2008). Cognitive humor processing: different logical mechanisms in nonverbal cartoons—an fMRI study. *Soc. Neurosci.* 3, 125–140. doi: 10.1080/17470910701745858

Shamay-Tsoory, S. G., and Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia* 45, 3054–3067. doi: 10.1016/j.neuropsychologia.2007.05.021

Shultz, T. R. (1972). The role of incongruity and resolution in children's appreciation of cartoon humor. *J. Exp. Child Psychol.* 13, 456–477. doi: 10.1016/0022-0965(72)90074-4

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neurosci. Biobehav. Rev.* 30, 855–863. doi: 10.1016/j.neubiorev.2006.06.011

Smith, A. (2006). Cognitive empathy and emotional empathy in human behavior and evolution. *Psychol. Rec.* 56, 3–21. doi: 10.1080/13682820310001615797

Snowden, J. S., Gibbons, Z. C., Blackshaw, A., Doubleday, E., Thompson, J., Craufurd, D., et al. (2003). Social cognition in frontotemporal dementia and Huntington's disease. *Neuropsychologia* 41, 688–701. doi: 10.1016/S0028-3932(02)00221-X

Sommer, M., Döhnel, K., Jarvers, I., Blaas, L., Singer, M., Nöth, V., et al. (2018). False belief reasoning in adults with and without autistic spectrum disorder: similarities and differences. *Front. Psychol.* 9:183. doi: 10.3389/fpsyg.2018.00183

Spearman, C. (1910). Correlation calculated from faulty data. *Br. J. Psychol.* 3, 271–295. doi: 10.1111/j.2044-8295.1910.tb00206.x

Stone, V. E., Baron-Cohen, S., and Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *J. Cogn. Neurosci.* 10, 640–656. doi: 10.1162/089892998562942

Streiner, D. L., and Norman, G. R. (1995). Health *Measurement Scales: A Practical Guide to Their Development and Use*. Oxford: Oxford University Press.

Suls, J. M. (1972). "A two-stage model for the appreciation of jokes and cartoons: an information-processing analysis," in *The Psychology of Humor: Theoretical Perspectives and Empirical Issues*, eds J. H. Goldstein and P. E. McGhee (New York, NY: Academic Press), 81–100. doi: 10.1016/B978-0-12-288950-9.50010-9

Thiébaut, F. I., White, S. J., Walsh, A., Klargaard, S. K., Wu, H. -C., Rees, G., et al. (2016). Does faux pas detection in adult autism reflect differences in social cognition or decision-making abilities? *J. Autism Dev. Disord.* 46, 103–112. doi: 10.1007/s10803-015-2551-1

Turner, R., and Felisberti, F. M. (2017). Measuring mindreading: a review of behavioral approaches to testing cognitive and affective mental state attribution in neurologically typical adults. *Front. Psychol.* 8:47. doi: 10.3389/fpsyg.2017.00047

Uekermann, J., Daum, I., and Channon, S. (2007). Toward a cognitive and social neuroscience of humor processing. *Soc. Cogn.* 25, 553–572. doi: 10.1521/soco.2007.25.4.553

Van Bourgondien, M. E., and Mesibov, G. B. (1987). Humor in high-functioning autistic adults. *J. Autism Dev. Disord.* 17, 417–424. doi: 10.1007/BF01487070

Vrticka, P., Black, J. M., and Reiss, A. L. (2013). The neural basis of humour processing. *Nat. Rev. Neurosci.* 14, 860–868. doi: 10.1038/nrn3566

Wagner, D. D., Kelley, W. M., and Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cereb. Cortex* 21, 2788–2796. doi: 10.1093/cercor/bhr074

Wellman, H. M., and Estes, D. (1986). Early understanding of mental entities: a reexamination of childhood realism. *Child Dev.* 57, 910–923. doi: 10.2307/1130367

Wyer, R. S., and Collins, J. E. (1992). A theory of humor elicitation. *Psychol. Rev.* 99, 663–688. doi: 10.1037/0033-295X.99.4.663

Zaki, J., Ochsner, K. N., and Ochsner, K. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nat. Neurosci.* 15, 675–680. doi: 10.1038/nn.3085

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Aykan and Nalçacı. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

SYSTEMATIC REVIEW
published: 19 June 2018
doi: 10.3389/fpsyg.2018.00890

Humor Assessment and Interventions in Palliative Care: A Systematic Review

Lisa M. Linge-Dahl[1]*, *Sonja Heintz*[2], *Willibald Ruch*[2] and *Lukas Radbruch*[1,3]

[1] Department of Palliative Medicine, University of Bonn, Bonn, Germany, [2] Personality and Assessment, Department of Psychology, University of Zurich, Zurich, Switzerland, [3] Center for Palliative Care, Malteser Hospital Seliger Gerhard Bonn/Rhein-Sieg, Bonn, Germany

Background: The central goal of palliative care is to optimize the quality of life of patients suffering from life-limiting illnesses, which includes psychosocial and spiritual wellbeing. Research has demonstrated positive correlations between humor and laughter with life satisfaction and other aspects of wellbeing, and physiological symptoms can be improved by humorous stimuli.

Objectives: The aim of this review is to evaluate humor interventions and assessments that have been applied in palliative care and to derive implications for future research.

Methods: A systematic review of four databases identified 13 included studies. Criteria for inclusion were peer-reviewed English-language studies on humor interventions or assessments in a palliative care context.

Results: Two studies on humor interventions and 11 studies on humor assessment were included in the systematic review. Most of these studies were about the patients' perspective on humor in palliative care. Findings showed that humor had a positive effect on patients, their relatives, and professional caregivers. Humor was widely perceived as appropriate and seen as beneficial to care in all studies.

Conclusions: Even though humor interventions seem to be potentially useful in palliative care, descriptions evaluating their use are scarce. Overall, research on humor assessment and interventions in palliative care has remained limited in terms of quantity and quality. More research activities are needed to build a solid empirical foundation for implementing humor and laughter as part of regular palliative care activities.

Keywords: humor, intervention, palliative care, end-of-life, systematic review

OPEN ACCESS

Edited by:
Tim Bogg,
Wayne State University, United States

Reviewed by:
Liudmila Liutsko,
Instituto Salud Global Barcelona (ISGlobal), Spain
Konrad Senf,
University of Hohenheim, Germany

***Correspondence:**
Lisa M. Linge-Dahl
lisa.linge-dahl@ukbonn.de

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *29 September 2017*
Accepted: *15 May 2018*
Published: *19 June 2018*

Citation:
Linge-Dahl LM, Heintz S, Ruch W and Radbruch L (2018) Humor Assessment and Interventions in Palliative Care: A Systematic Review. Front. Psychol. 9:890. doi: 10.3389/fpsyg.2018.00890

INTRODUCTION

Rationale

Humor has been subject to research and philosophical reflections for centuries and has also been used for interventions in the health sector (Hulse, 1994). Most research has been conducted in pediatrics (review by Sridharan and Sivaramakrishnan, 2016). Apart from the health sector, humor interventions have also been investigated in the field of positive psychology (Ruch and McGhee, 2014; Ruch and Hofmann, 2017). Some studies in medical settings were conducted with older people in nursing homes (Mathieu, 2008; Goodenough et al., 2012; Low et al., 2013), cancer patients

(Itami, 2000; Venter et al., 2008), veterans (Steinhauser et al., 2000), and patients suffering from depression (Shahidi et al., 2011). Positive correlations have been reported on humor and laughter in relation to life satisfaction outside the health care setting (Wild et al., 2003; Ruch et al., 2010), and there is some evidence of a relationship between humor and health (Martin, 2001, 2004).

The theoretical model of the effect of humor on health has been described by Martin (2008) and Gremigni (2012) extensively, who concluded that humor as a complex psychological phenomenon needs to be differentiated according to the kind of humor and the setting. Hearty laughter, for example, works through different mechanisms than social and interpersonal aspects of humor and results in different effects. Social and interpersonal aspects of humor, such as enhancing personal connections, influence health and wellbeing by increasing one's level of social support, while hearty laughter may predominantly affect health by improving the respiratory, musculoskeletal, vocal, and cardiovascular activity. Each kind of humor requires a specific research setting and will produce specific effects (Martin, 2008).

Society perceives humor to have beneficial effects on health and wellbeing (Boyle and Joss-Reid, 2004). Implementation concepts of humor and the scientific evaluation of their effects (Boyle and Joss-Reid, 2004) have been developed over the last century. These different kinds of interventions range from individualized humor therapy visits via the presentation of humorous movies aligned with patients' humor preferences (Schwartz and Saunders, 2010) to clowns working in the public health sector. Warren and Spitzer (2011) provided a summary of different types of clowns working in health care settings (e.g., elder-clowns and "classical" clowns in hospitals) in various countries and concluded that the application in elder and end-of-life care may not only benefit residents and patients, but health care professionals and family members as well. There are not only different types of clowns in healthcare but also different styles of humor that can be assessed (Craik et al., 1996; Schultes, 1997; Martin et al., 2003; Ruch et al., 2018). One of the few randomized controlled studies on humor interventions with adequate power was carried out in Australia and included 398 residents from nursing homes (Goodenough et al., 2012; Low et al., 2013). The single-blind randomized controlled study evaluated a clown intervention over a period of 9–12 weeks, which showed a significant decrease in agitation in residents compared to the control group receiving usual care. Additionally, so called "LaughterBosses" (staff members in nursing homes) were trained as facilitators with techniques to incorporate humor in between elder-clown visits. Humor also seems to be a relevant coping mechanism in various aspects of patients' lives. In her analysis of posts in an online patient-to-patient cancer forum, Demjén (2016) found that patients make fun of cancer and its consequences in multiple and creative ways to cope with their physical and psychological distress.

Despite these beneficial effects, there has been limited research on humor interventions for patients at the end of life. This might result from the societal perception that death is not supposed to be the object of implementations that included humor (Herth, 1990). Also, certain situations or topics might limit or impede the use of humor; for example, unfamiliarity between the patient and the health care professional (Erdman, 1991) or the fear of ridicule in certain patient groups, such as penile cancer patients (Branney et al., 2014).

However, the limited number of existing studies imply that humor might be beneficial toward the end of life as well (Steinhauser et al., 2000). Cox (1998) explored the effect of humor, art, and music on dying children through a literature review and found that any kind of social support and artistic strategies to process emotions and grief helps children: "[…] to remove the distance to others, find relief for depression, enhance their self-esteem, lower anxiety, fear and other feelings of grief and achieve an improved level of acceptance of reality" (Cox, 1998, p. 416). Cancer patients talk about humor as one of the predominant themes and coping strategies in their lives (Venter et al., 2008). Dean (1997) extrapolated findings from humor research in other health care settings and concluded that humor may be applied in the palliative care setting as well. However, she also noted that in certain situations, like crises and imminent death, humor would not be appropriate. From the perspective of health care professionals, Müller et al. (2012) found that humor is one of the three most powerful resources that protect health care teams from the negative effects of the strain of death and dying.

Kanninen (1998) conducted a review on humor in palliative care, but found only one pilot study that analyzed the effect of humor on 14 patients (Herth, 1990). The remaining articles included in Kanninen's review were anecdotal personal experiences of individuals. Kanninen concluded that research is needed to establish if humor is effective in medicine, especially in palliative care. The present paper reviews the study of Herth (1990) and the research that has been added in the two decades since Kanninen's review. It thus lays the foundation for future research on humor interventions in palliative care, assessing the effects on patients, relatives, and health care professionals.

Objectives

The aim of this review is to synthesize humor interventions and assessments that have been applied in palliative care and to derive implications for future research and applications. The investigated patients were diagnosed with an incurable disease and were at the end of their lives. Study designs and outcomes of interventions and assessment are compared and grouped to facilitate cross-study comparisons.

Research Questions

This systematic review evaluates the effectiveness of humor interventions in a palliative care setting. It also outlines which kinds of humor interventions and assessments have been applied in palliative care until now and the methods, results, and limitations of these studies.

METHODS

Study Design

A systematic literature review of qualitative and quantitative research was undertaken in July 2017.

Participants and Interventions

The target group in the reviewed studies consisted of patients in a palliative care setting who received a humor intervention. Studies assessing the perspective of family caregivers or health care professionals on humor were also included. Different kinds of interventions and assessments were reviewed in a range of patient groups and institutions. All patients had diagnoses of incurable diseases and received end-of-life care.

Systematic Review Protocol

Overall, 336 abstracts were found and reviewed by two authors (LLD and LR), with an agreement rate of >95% regarding the investigated publications. Screenings resulted in 64 abstracts that were rated as potentially relevant for the review. Lack of consensus about inclusion was discussed with another author (SH). Next, 32 articles were analyzed as full-text versions, from which 13 met the inclusion criteria (see **Figure 1**), for further information please access the Supplementary Material. The included studies were published between 1990 and 2017. No older studies have been identified in the literature search. The 17 articles which were not included were an opinion paper (Dean, 1997) or articles that investigated patient groups which did not meet the criteria of palliative care (e.g., Low et al., 2015).

Search Strategy

Three search strings on the topics of humor, intervention, and palliative care connected by Boolean operators were used. The search terms were: {(humor OR humor OR humorous OR clowns OR clown[Title/Abstract]) AND (intervention OR training OR coaching OR visit OR practice OR therapy[Title/Abstract]) AND ("palliative care" OR "hospice care" OR "end-of-life" OR geriatric OR "life limiting illness" OR death OR dying[Title/Abstract])}.

Publications were included if they were published in a peer-reviewed journal, contained original qualitative or quantitative data, applied and/or assessed a humor(ous) intervention, evaluated effects on patients or residents in nursing homes receiving palliative care, and were published in English. The year of the publication of the study was not restricted.

Data Sources and Data Extraction

Four key databases (PsycInfo, PubMed, Google Scholar, and Cochrane Library of systematic reviews) were systematically searched to July 16th 2017. Full-text publications were downloaded via the library of the Medical Faculty of the University of Bonn.

Data Analysis

All included articles were reviewed in depth. The selected studies were divided into (a) studies that investigated humor in palliative care as the main goal of the paper and (b) studies in which humor emerged as an important variable from an initial research question that had not focused on this topic, for example assessing end-of-life wishes (Delgado-Guay et al., 2016). Target groups, participant numbers, publication bias, study methodology, and quality of research were also analyzed using a template. However, the wide range of different conceptualizations of humor in the studies as well as methodological weaknesses prevented meaningful comparison between studies. Results are presented according to target groups and study methodology. Effect sizes were analyzed using Cohen's (1992) guidelines. Potential bias within the studies was identified and discussed.

RESULTS

The 14 included research papers contained data on 13 studies (see **Figure 2**). One study was published in two separate papers, one describing the qualitative results (Kontos et al., 2016) and the other discussing the quantitative results (Kontos et al., 2015). Ten articles were selected because they presented findings of interventions or assessments of humor as the main goal of the paper. Four other publications were included because they dealt with humor, among other variables, as a secondary outcome. Two publications focused on humor interventions and eight mainly on the assessment of patient's perception of humor, while three examined the perspective of caregivers and/or health care professionals. Nine publications described qualitative results (Herth, 1990; Langley-Evans and Payne, 1997; Schultes, 1997; Dean and Gregory, 2004; Adamle and Ludwick, 2005; Richman, 2006; Cain, 2012; Bentur et al., 2014; Kontos et al., 2015), and five articles presented quantitative results (Kissane et al., 2004; Ridley et al., 2014; Delgado-Guay et al., 2016; Kontos et al., 2016; Claxton-Oldfield and Bhatt, 2017). Overall, a total of 759 participants were included in the reviewed studies.

The results are presented in the following order: the two studies that included humor interventions (Schultes, 1997; Kontos et al., 2015, 2016), three studies exploring perception and appropriateness of humor in hospice settings (Herth, 1990; Ridley et al., 2014; Claxton-Oldfield and Bhatt, 2017), followed by five publications that assessed functions and results of humor applications on patients in hospice care (Langley-Evans and Payne, 1997; Dean and Gregory, 2004; Adamle and Ludwick, 2005; Cain, 2012; Delgado-Guay et al., 2016) and one on patients in an oncology ward (Bentur et al., 2014), followed by two studies presenting results from psychotherapists' observations (Kissane et al., 2004; Richman, 2006). Within each of the subsections of the results, the studies are presented in the order of their publication date beginning with the most recent one. At the end of each section, the main information is condensed in a table.

Studies That Included Humor Interventions

Two studies investigated the effects of humor interventions in a palliative care setting (Schultes, 1997; Kontos et al., 2015, 2016), one for patients with advanced dementia in nursing homes, and one for patients being treated by a hospice service at home. Both studies applied humor interventions in a palliative care setting. While one study used clowns (Kontos et al., 2015, 2016), the other study involved nurses using humor with the patient (Schultes, 1997). The outcome measures and the study participants varied strongly, limiting comparability between studies (see **Table 1**).

A Canadian study using so called "elder-clowns" (with a red nose, but minimal make-up and clothing from an earlier era) applied approximately 10 min humor interventions twice a week over a period of 12 weeks to nursing home residents in an advanced stage of dementia (Kontos et al., 2016). No

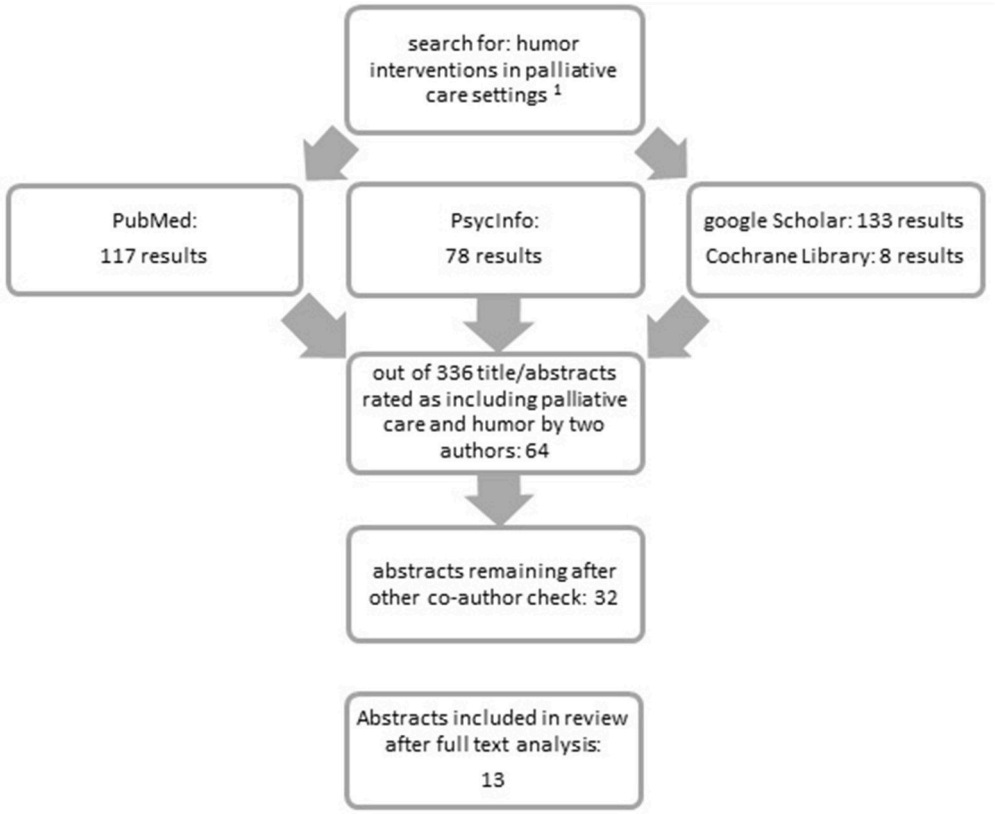


FIGURE 1 | Flowchart of the review protocol. [1]search terms: {(humor OR humor OR humorous OR clowns OR clown[Title/Abstract]) AND intervention OR training OR coaching OR visit OR practice OR therapy[Title/Abstract]} AND "palliative care" OR "hospice care" OR "end-of-life" OR geriatric OR "life limiting illness" OR death OR dying[Title/Abstract].

control group was investigated, so bias cannot be ruled out. The qualitative results of the study were published separately (Kontos et al., 2015). The clowns used improvisations, humor, empathy, song, musical instruments, and dance. Data collection involved video recording the interventions, and the clowns were interviewed afterwards. Several researchers screened and transcribed the videos to assure interrater reliability. The aim of the intervention was to achieve "relational presence," a term that Kontos et al. define as: "[…] the reciprocal nature of engagement during plays, and the capacity of residents to initiate as well as respond to […] creative engagement" (p. 5). To facilitate an appropriately tailored intervention for each participant, so called "census information"—personal preferences, history of the patient and personality—was informally collected from staff or family. With a small number of participants ($N = 23$) a significant improvement was found between the baseline and the end of intervention scores in "behavioral and psychological symptoms of dementia" (from $M = 24.4$; $SD = 12.9$ baseline to $M = 18.6$; $SD = 13.1$ after 12 weeks; scale from 0 to 144; $t = -2.68$, $p = 0.01$; Cohen's $d = -0.45$), quality-of-life (from $M = 0.04$; $SD = 0.51$ baseline to $M = 1.05$; $SD = 0.29$; scale from -5 to 5; after 12 weeks; $F = 23.09$, $p < 0.001$; Cohen's $d = 2.44$) and "occupational disruptiveness" (from $M = 8.09$; $SD = 7.1$ baseline to $M = 4.9$; $SD = 5.2$ after 12 weeks; scale from 0 to 60; $t = -2.58$, $p = 0.02$; Cohen's $d = -0.51$) using questionnaires completed by the nursing staff and family members. Use of psychiatric medication and nursing burden did not change significantly. There was a tendency for decreased agitation/aggression, but this did not reach statistical significance (from $M = 3.3$; $SD = 3.3$ baseline to $M = 2.1$; $SD = 2.0$; scale from 0 to 12; $t = -1.86$, $p = 0.07$; Cohen's $d = -0.44$). The authors report that persons diagnosed with dementia could engage in the humor interventions in different ways even though they were in their last stage of life. This engagement ranged from sharing their sadness to reciprocal play, joy, imaginative exploration, and from recognizing humor to even creating humor on their own initiative.

The second intervention was developed after an analysis of the existing literature on humor in health care. Schultes (1997) evaluated a humor intervention for patients treated by hospice home care nurses. The intervention was guided by humor assessment questions to explore the preferred style of humor (e.g., incongruity, nonsense, ridicule, or slapstick) and instructions for nurses on how to observe humorous behavior. After the assessment procedure, humorous cassettes and movies were shown to the patient according to the preferred humor style.

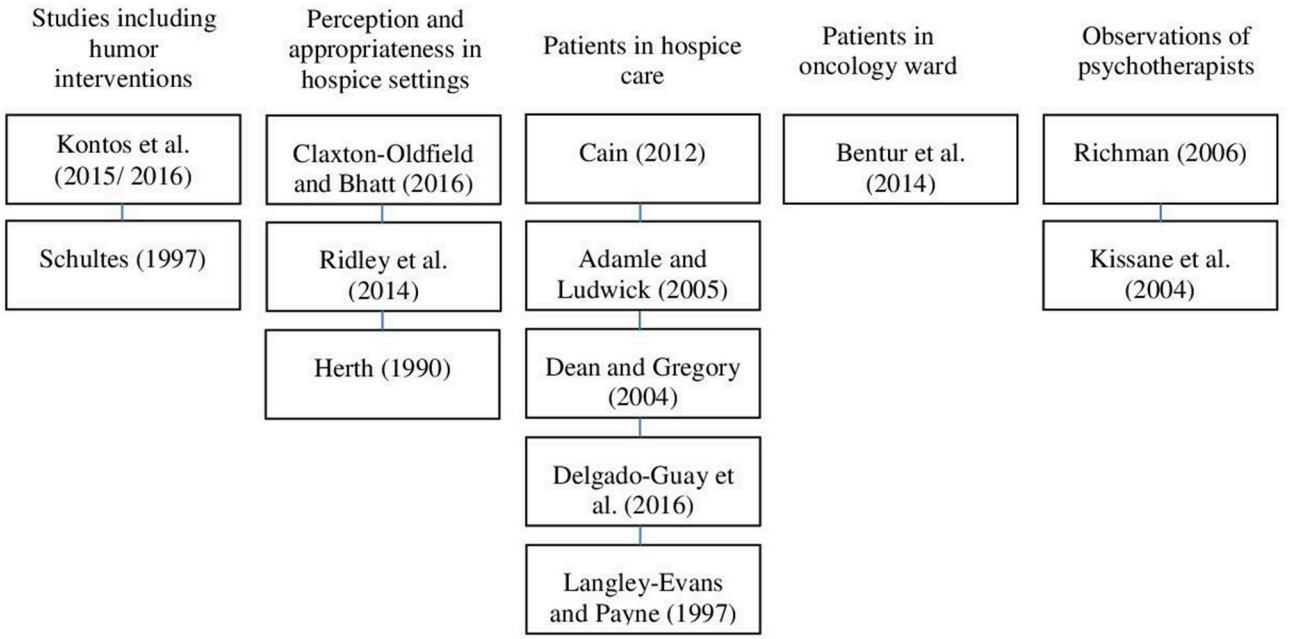


FIGURE 2 | Flowchart of included studies.

The intervention was tested in a clinical case study with a 65-year-old woman suffering from metastatic colon cancer. Data collection on the intervention effects was based on observations of the patient by nurses and informal interviews with the patient's relatives. The results of the case study indicated that humorous interactions and listening to humorous cassettes or watching funny movies made the patient feel better, that she demanded less pain medication and smiled more, and that it also improved the quality of her remaining life. Even after the patient's death, her family reported that they continued to watch the movies, which helped them to feel relieved and to cope better with their grief, and which gave them a sense of power in a situation where they felt weak. However, the authors did not follow up the case report with a humor intervention trial and the lack of an independent researcher in the data collection could have led to biases.

Studies Assessing or Observing Humor in an Explorative Way

Exploring Perception and Appropriateness in Hospice Settings

Three studies assessed the appropriateness of humor as an intervention in hospice settings using qualitative data (Herth, 1990), quantitative data in general (Ridley et al., 2014), and quantitative data on volunteer-patient interactions (Claxton-Oldfield and Bhatt, 2017). Humor was perceived as appropriate or even essential in those settings, though the authors mentioned limitations regarding the use of humor, such as impending death or absence of a sense of humor (see **Table 2**).

Herth's (1990) small study on 14 terminally ill adults receiving hospice care at home explored patients' perceptions of and experiences with humor in structured interviews. Patients explained that humor incorporated the following improvements: connectedness, change of perspective, hope, joy and relaxation including physiological improvements. Also, the majority (12 of the 14 participants) of the interviewees stated a need for humor, indicated by quotes such as "Everyone is so sad," "It just makes it harder, I wish we could lighten up," and "If I ever needed humor it is now" (Herth, 1990, p. 38). The author concluded that terminally ill people appeared to be the ones who benefitted the most from humor interventions. As a coping mechanism, humor becomes essential due to deteriorating body functions, unfamiliar procedures, and physical and emotional suffering. Humor was also described as one of the most powerful coping mechanisms. However, the strong conclusions that the authors drew may be questioned in relation to the small sample size of the study.

Ridley et al. (2014) analyzed whether humor is appropriate in a palliative care setting. They interviewed 100 patients in palliative care units and residential hospices. A standardized questionnaire captured patients' perception toward humor therapy prior to and during their illness (Ridley et al., 2014). Ridley et al. reported a potential "bias inherent to retrospective self-reporting" (2014, p. 474). Most participants valued humor as important prior to (77%) and during (76%) their illness. However, the frequency of laughter in patients who laughed 16 or more times a day declined from 65% prior to the illness to 22% during the illness. Patients who rated humor to be more important than other patients were more likely to consider themselves as funny before ($p < 0.001$) and during ($p = 0.014$) their illness.

The perception of appropriateness, types, frequency, and results of humorous interactions in hospice and palliative care

TABLE 1 | Studies including humor interventions.

Authors (year of publication)	Target group (N)	Intervention/assessment	Main results
Kontos et al., 2015, 2016	Nursing home residents (23)	Elder-clown interventions	Humor interventions of elder-clowns resulted in a decreased level of agitation and a nominal decrease in dosing of psychotropic medications. Observation protocols showed improvements in the patients' expression of joy, sadness, reciprocal play, and co-constructed imagination.
Schultes, 1997	Patients receiving home-care or in hospices (1/case study)	Humor intervention implemented and evaluated by hospice home-care nurses	The case study showed that the patient responded well to the humorous films and a change of the mood was perceived from the family as well. The film material was even used after the patient's death and helped the relatives to cope with the loss.

TABLE 2 | Studies exploring perceptions and appropriateness in hospice settings.

Authors (year of publication)	Target group (N)	Intervention/assessment	Main results
Claxton-Oldfield and Bhatt, 2017	Volunteers working in hospice or palliative care (32)	Humor application during patient visits	Humor was applied after getting to know the patient and following the patients lead (40.7%), depending on his stage of illness (41.4%). 96% of the volunteers stated that there is a place for humor in palliative care and 88.9% stated that humor helped them to cope with the demands of their voluntary work.
Ridley et al., 2014	Patients in palliative care units or residential hospices (100)	Questionnaire about appropriateness of humor in palliative care	Patients valued humor as important prior (77%) and during (76%) their illness. The frequency of laughter declined from 65% of patients who laughed 16 or more times a day prior to the illness to 22% during the illness.
Herth, 1990	Terminally ill adults (14)	Structured interview to explore patients views on humor	Eight participants reported humor to be an important part of their lives prior to the illness. Twelve stated that humor would be helpful in the present situation, but only two indicated a presence of humor in their lives.

patients during their interaction with volunteers was analyzed by Claxton-Oldfield and Bhatt (2017) from a volunteers' ($N = 32$) point of view. A quantitative questionnaire was developed on the basis of an informal discussion with four volunteers. The first part of the questionnaire examined the frequency of humor in patient-volunteer interactions (for example "How often do your patients initiate humor with you during your interactions with them?"). The second part examined the acceptability of humor in interactions. The volunteers visited patients in a range of different settings (hospital, client's home, nursing home, and residential hospice). The authors report a potential bias from nonresponse. More than half of the volunteers rated humor as very or extremely important in interactions with patients. In most cases humor was applied (a) after getting to know the patient and following the patients' lead ($n = 11$; 40.7%) and (b) depending on his/her stage of illness ($n = 12$; 41.4%). Impending death was perceived as a very inappropriate moment for the use of humor. All in all, 96% ($n = 31$) of the volunteers believed that there was a place for humor in palliative care, and 88.9% ($n = 24$) stated that humor helped them to cope with the demands of their voluntary work.

Assessment of Functions and Results of Humor Application

Patients in hospice care

Five studies examined the functions and results of humor applications (see **Table 3**). All of them used observations and interviews as methods of data collection. The results demonstrated that humor was crucial for hospice professionals to cope with the demands of their jobs (Cain, 2012), that it was primarily initiated by patients (Adamle and Ludwick, 2005), and that it helped health care professionals and patients to build relationships and to bear difficult situations. Humor was, moreover, a means to express sensibility (Dean and Gregory, 2004), it represented an important end-of-life wish (Delgado-Guay et al., 2016), and it helped patients to distance themselves from their own death (Langley-Evans and Payne, 1997).

Delgado-Guay et al. (2016) compared four different tools developed to rate end-of-life wishes in a randomized controlled trial (RCT). Hundred patients with advanced cancer in an inpatient palliative care unit in South Texas rated "to keep my sense of humor" as one of the ten most important end-of-life wishes (45% of all participants).

TABLE 3 | Studies assessing functions and results of humor application.

Authors (year of publication)	Target group (N)	Intervention/assessment	Main results
Delgado-Guay et al., 2016	Advanced cancer patients (100)	Assessment of end-of-life wishes	45% of the participants rated "to keep my sense of humor" as one of the ten most important end-of-life wishes.
Cain, 2012	Hospice professionals (41 + 7 informal interviews)	Participating observation, structured and informal interviews	Humor was found to inherit an important role predominantly in the back region of the hospice. Humor was found as an instrument of distancing, to enable professionals to deal with emotional difficult times, and being a resource of strength to continue their job.
Adamle and Ludwick, 2005	Patients in hospice care (160)	Observation of patient-nurse –primary caregiver interactions	Humor was observed during more than 80% of all observed visits. 70% of this humor was initiated by the patients, 17% by the nurses, and 12% by the caregivers. The average number of humorous remarks per visit was three.
Dean and Gregory, 2004	Patients, their families and health care professionals in palliative care (15 interviews)	Participating observation, semi-structured (health care professional) and informal interviews (patients and family)	Humor was found to be pervasive and persistent in palliative inpatient care. Three main functional categories of humor emerged from the data: building relationships, contending with circumstances, and expressing sensibility.
Langley-Evans and Payne, 1997	Patients in a day care palliative care ward (14)	Participant observation and informal interviews	Rather nonverbal humorous nature in this "death talk" enabled the patients to distance themselves from their deaths.

TABLE 4 | Studies with patients in other care settings.

Authors (year of publication)	Target group (N)	Intervention/assessment	Main results
Bentur et al., 2014	Advanced cancer patients (22)	In-depth interviews about coping strategies	Humor was found to be one of five coping strategies that were applied by the patients.
Richman, 2006	Patients receiving psychotherapy (8)	Investigation during psychotherapy	Ten features of humor in psychotherapy with patients at the end of their lives emerged. This included empathy, connectedness, the possibility to mentally distance from death, and the reduction of stress.
Kissane et al., 2004	Women with advanced breast cancer (227)	Supportive expressive group therapy	Amongst other topics, genuine humor was found to be a sign of a healthy functioning group in group therapy.

Cain (2012) analyzed "front" and "back region" personalities of health care professionals; that is, the personality shown in front of patients and relatives on the one hand, and the personality presented in team meetings and with colleagues on the other hand. Data was collected through observations over 1 year by a researcher at the ward and 51 interviews with staff-members. Bias was possible because only one researcher collected the data, so no inter-rater checks were conducted. Among the dynamic and complex interactions of staff and patients, she found that humor fulfilled an important role, predominantly in the "back region" of the hospice staff. It was not only an instrument to distance oneself from negative emotions, but also a resource of strength, which enabled professionals to deal with emotionally difficult times.

Adamle and Ludwick (2005) observed 132 interactions between nurse, patient, and primary caregivers in hospice settings (home care hospice services, inpatient hospice, and hospice care in nursing homes) including 160 participants. They counted the number of occurrences of humor and who initiated them. Potential bias was reported in the selection process of participants. In three different settings, humor was observed in 85% of the 132 observed interactions. In about 70% of the cases, humor was initiated by the patient, and the average number of humor occurrences per visit was three. The lack of humorous occurrences in 15% ($n = 20$) of the observed interactions was due to the cognitive inability or impending death of the patient (nine patients were either in a coma or did not respond physically or mentally to verbal cues, and five patients were dying).

In another study, Dean and Gregory (2004) focused on the circumstances, functions, and appropriateness of humor in an inpatient palliative care unit using participant observation plus informal and structured interviews with 15 health care professionals. Detailed field notes and transcribed interviews were analyzed. Humor was reported to be "pervasive and persistent" (p. 140) and had the following key functions:

(a) building relationships (making connections, humor as attraction, discovering hidden verbal messages, energizing, nurturing community, neutralizing status differences), (b) bearing the situation (humor as respite, humor as survival, humor as tension relief/lightening the heaviness, maintaining perspective/providing support), and (c) expressing sensibility (preserving dignity, acknowledging personhood).

In their ethnographic investigation, Langley-Evans and Payne (1997) studied how patients in a palliative day care unit think and talk about their condition and death, using participant observation over a period of 7 weeks and evaluating field notes and documentary information from health care professionals. One theme that emerged from the qualitative data analyses was the rather nonverbal humorous nature of this "death talk," which enabled patients to distance themselves from their own deaths.

Patients in Other Settings

Three studies examined patients in other settings (see **Table 4**). Bentur et al. (2014) analyzed coping strategies at the end of life in 22 advanced cancer patients in an Israeli daycare oncology clinic. The interviews were transcribed verbatim and analyzed afterwards. Humor was described as one of the five applied coping strategies. One participant stated on the use of humor "maybe it helped me ease the burden" (Bentur et al., 2014, p. 4).

Two studies in a psychotherapeutic setting with end-of-life care patients extracted data from participant (therapist) observations and showed humor as an unplanned result of an explorative observation. Richman (2006) discussed the functions of humor in psychotherapy. Ten features of humor were developed by Richman based on eight patients, at the end of their lives, receiving psychotherapy. There is a risk of bias due to an unclear selection process of the patients. Skills in the use of humor were found to be necessary for psychotherapists treating patients at the ends of their lives or facing the topic of death. The ten features of humor were: (1) emerges spontaneously, (2) timing is essential, (3) fosters social cohesion, (4) power to reduce stress, (5) enforces feeling of community, (6) permits to distance from death, (7) the content of humor can be negative, (8) communication is essential, (9) requires a healing therapist with empathy, and (10) feeling of commonness.

In a large RCT study on 227 women with metastatic breast cancer, the topics and facilitating aspects of a weekly supportive-expressive group therapy were qualitatively analyzed (Kissane et al., 2004), indicating that genuine humor was a sign of a healthy functioning group. Furthermore, notes of the co-therapists were cross-checked by the main therapists and analyzed qualitatively, resulting in five categories: (1) the structure of supportive-expressive group therapy, (2) the role of therapists, (3) key themes, (4) group transformation, and (5) anti-group phenomena.

DISCUSSION

Summary of Main Findings

By systematically reviewing the state of the art of humor in palliative care two decades after the review of Kanninen (1998), which included only one study on a humor intervention, we were able to include 13 studies in this review. Study results suggested that humor is an appropriate and useful resource in palliative care, but only two studies evaluated humor interventions in palliative care, and only one of the two was a RCT. Most of the reviewed publications explored and observed humor in different settings. There was no consensus on a definition of humor, on types of intervention, or on the assessment of effects that would allow comparisons of the published trials. Thus, studies were difficult to compare due to a different understanding of what humor interventions should look like, what they should accomplish, and which group of professionals should implement these interventions. Still, some conclusions about the benefits of humor can be derived from the reviewed studies.

One of the key benefits of humor in health care, which was reported in several trials, was an increased pain tolerance (Weisenberg et al., 1995; Zweyer et al., 2004). This finding was also in line with Herth's study (1990) in a palliative setting. However, it needs to be stated that RCT studies would be necessary to show whether the increase in pain tolerance (cold pressure test) was really due to the humorous stimuli or related to distraction or other factors.

Konradt et al. (2012) demonstrated the effect of a humor therapy group on older patients suffering from depression, which led to lower levels of seriousness and higher satisfaction with life scores in comparison to the control group. The study by Kontos et al. (2015) also highlighted the positive impact of clown interventions on physical and psychological well-being, demonstrating the benefits of the holistic approach. These statements need to be interpreted very carefully in relation to the small sample sizes that have been examined. The SMILES model for the implementation of humor in palliative care (Borod, 2006) was developed on the basis of a literature review about uses of humor and was modeled on the SPIKES model for the delivery of bad news in health care (Baile et al., 2000). SMILES aims at facilitating the use of humor in patient-physician interactions. The categories of this model are "**s**mile" (enter patient room with a smile), "**m**ake eye contact" (look and actually see the patient), "**i**ntuition and imagination" (sense appropriate cues for humor introduction), "**l**ook for, listen to, and Leap at the Opportunity" (get the real meaning of patients statements, so register subtle cues), "**e**lephants never forget" (remember exchanges with the patient and use them in following interactions) and "**s**ensitive to situation" (be aware of appropriateness of humor due to the situation). All these categories were illustrated by examples and aim at the application of humor in an appropriate and successful way. The success was not evaluated and bias in the selection of categories is possible.

But how does humor compare to other interventions in terms of well-being? Wellenzohn et al. (2016) tested the effect of different online humor interventions against a control group that reported early childhood memories and found humor to be efficacious. It needs to be noted though that this study included only healthy adults, and humor interventions would thus have to be tested in hospital patients at the end of their lives to provide conclusions for the target group of the present systematic review. Auerbach et al. (2016) were able to show that clinic clowns can induce more positive emotions than a circus clown and a nurse

interaction by assessing the patients' current emotional state. Lacking in the literature is a comparison of humor interventions with other interventions such as music interventions, relaxation, yoga, or art therapy in palliative care (Koch et al., 2016). These controlled studies should include humor interventions as well as active control groups, including comparable interventions like music and art interventions and groups receiving usual care or additional nursing care to determine which beneficial effects are due to humor and laughter, and which ones are due to indirect factors (such as increased positive emotions or more interpersonal contacts). Future studies should also investigate whether humor interventions (in comparison to control groups) can lead to a decrease in the consumption of analgesics as well as a decrease in self-reported pain intensity. In addition, a longitudinal study setting would be preferable for future research as generalizations are limited for the results of cross-sectional studies.

However, there are discrepancies concerning the aim of the humor intervention. While Kontos et al. (2015) stressed that sadness and frustration need to receive sufficient attention and space, Schwartz and Saunders (2010) stated that the aim of humor therapy is to make patients laugh. Kontos et al. further emphasized that the aim of humor interventions is not to make the resident laugh, but to ease his/her state of mind and work with whatever is possible at that very moment. Similarly, the American Cancer Society (quoted in Schwartz and Saunders, 2010, p. 554) defined humor therapy as "[. . .] the relief of physical or emotional pain and stress and as a complementary method to promote health and cope with illness". Apart from different definitions and concrete applications of humor, the consent of all investigated studies was that humor is not only valuable, but an important component of palliative care: "[. . .] humor is the glue that helps to put the connection together [. . .] and as Palliative Care is all about relationships [. . .] it would be incomplete" (Dean and Gregory, 2004, p. 141).

Not losing one's sense of humor was rated as an important spiritual end-of-life need (Delgado-Guay et al., 2016). These results might differ significantly in other cultural and spiritual settings, but we found no publications on the use of humor outside the Western-European cultural setting.

It has been stated that the sense of humor remains intact in people and even increases toward the end of life (Ruch et al., 2010). Thus, humor interventions are meaningful throughout the whole lifespan, including the end of life. Conducting humor interventions with patients in palliative care makes sense with the limitation that a sense of humor needs to be present in those individual patients taking part (Ruch and Hofmann, 2012; Auerbach, 2017), and the participants should not suffer from gelotophobia (the fear of being laughed at; Ruch et al., 2014).

There were several approaches to assess the patients' preferred kind of humor and whether they perceive humor as appropriate in their individual situation. Asking patients whether they consider themselves to be funny might be used as a screening question to identify people who find humor in their interactions with care providers appropriate (Ridley et al., 2014). However, humor production (being funny) is different from humor appreciation (perceiving humor as appropriate and helpful). Additionally, this kind of question needs to be used with care and considering the patient's actual emotional and spiritual situation. Adamle and Ludwick (2005) suggested that humor should occur without cues or prompting, enabling spontaneous humor. This would require an emotional atmosphere in the palliative care setting that allows the expression of humor from the patient's point of view. However, there were also critical voices that point to the use of off-color humor (gallows humor) amongst health care professionals (Piemonte, 2015). Self-disparagement related to functional defects was found to be predominant in elderly care, but should be initiated by the residents themselves, as otherwise it could be counterproductive (Keltner and Bonanno, 1997). To understand the benefits and limitations of the use of humor in palliative care, researchers need to conceptualize humor as a continuous rather than a binary concept (to have or not have a sense of humor), and they need to consider different facets of humor, ranging from benevolent humor to mockery (see Craik et al., 1996; Ruch et al., 2018). Both the "flavor" of humor (e.g., supportive, critical) as well as the targets (who jokes about whom) need to be taken into account, because it might heavily influence the impact of the use of humor. As a result, humor in palliative care settings should be social, benevolent, and supportive for the patient and his/her family.

The positive effects of humor on mourning relatives reported by Schultes (1997) has also been assessed by Keltner and Bonanno (1997) in a more structured way using questionnaires and structured interviews. However, family caregivers of patients receiving palliative care have not yet been included in a study in a structured and adequate way to comprehensively assess the effect of humor interventions with them.

In the field of professional caregivers and volunteers, humor was observed to be a valuable resource. Cain (2012) recorded statements of hospice workers saying that former colleagues, who quit their jobs because they could not handle the emotional burden, supposedly did so because they had lacked a sense of humor. This implies that humor is an ingredient to successful performance in this field (Müller et al., 2012). Measurement tools for assessing individual differences in humor could also be useful in the area of palliative care (for reviews see Ruch, 2007; Ruch et al., 2014). Critical aspects of humor such as sarcasm and cynicism could be potentially detrimental in the area of palliative care and thus need to be analyzed in more detail (Ruch et al., 2018). Importantly, assessing humor might put more strains on palliative patients (e.g., in terms of concentration, comprehension, and effort) than on healthy adults, for which humor measures were usually developed and tested. Thus, existing instruments might likely need to be adapted and pre-tested to ensure that the measurement is feasible and ethical in palliative patients. For example, short and/or simplified versions might need to be employed, or the items might need to be read to the patients. This need for short assessment tools has become clear in an unpublished pilot test of our research group.

Attrition numbers are an important component when analyzing the effects of humor interventions, because it is possible that certain people are more likely to remain in this kind of study setting. Low et al. (2013) reported a dropout of 16 residents from the initially 414 people that have been assessed for eligibility. Of

those 16 residents, six did not give consent to participate in the study and 10 died or were transferred to a different location. Kontos et al. (2016) reported screening 45 residents, from which 23 were recruited. No information was provided on the selection process. The authors stated that during the intervention, 10 residents received all treatments, whereas 13 missed an average of 2.3 of the 24 visits. It needs to be taken into account that this kind of dropouts needs to be analyzed carefully in future research to explore potential differences in humor-related traits (such as gelotophobia or the sense of humor) of people who stay in humor intervention studies and those who drop out or decline to participate in the first place. Identification of potential responders might be difficult though, as data from people who decline to participate in a study usually is scarce. The study of Wellenzohn et al. (2016) gave detailed information on a 25% dropout rate from all four investigated groups. The dropouts were younger, with a predominance of men, yet they did not differ from other participants in their baseline levels of happiness or in depressive symptoms.

Limitations

Our search strategy focused on publications in peer-reviewed journals and English language, and thus some interesting and potentially relevant results published in dissertations or in other languages could not be included. Overall, the search strategy might have been too restrictive with its focus on palliative care, as results from other areas of medicine might be transferred to the palliative care setting. However, the cognitive and physical impairment of patients with advanced life-limiting diseases and the high prevalence of depression in these patients put this comparability into question. It is also possible that studies have been published in nonmedical or psychological journals that were not included in the databases chosen for the present systematic review. However, any of these expansions would have gone beyond the scope of this paper.

The findings of the analyzed studies were often based on either self-reports or observations. To ensure the validity of the findings, multi-method studies, such as the study by Kontos et al. (2015, 2016), would be worthwhile. Ideally, these studies should combine for example self-reports, other-reports, physiological measures, and behavior observations, and they should include the perspectives of patients, caregivers, and health care professionals alike.

The small effect sizes of the quantitative studies need careful interpretation. Due to the small sample sizes, the effect sizes, according to Cohen's guidelines (1992), were not interpretable as representative results. Larger samples would be needed to demonstrate the efficiency of the interventions in the studies of Kontos et al. (2016), Claxton-Oldfield and Bhatt (2017), and Adamle and Ludwick (2005). Limitations of studies with small sample sizes (Ioannidis, 2005; Maxwell et al., 2008) also imply that for the study of Kontos et al. (2016) a careful calculation of sample size and power analysis would have been required to improve the quality of results. Using multiple comparisons (e.g., Kontos et al., 2016) would also require corrections for alpha error accumulation, if appropriate to the design (Armstrong, 2014).

The risk of bias has been assessed, and no bias has been found due to mutual cross-checks of the selection of articles between two authors. A publication bias may have affected the published literature because studies with significant positive results are more likely to be published than those without significant results.

A documentation template had been developed for our review, but with only scarce information on the quality of research and details on effect sizes, the scheme did not deliver usable results. A different template with a lower focus on study quality might have been more suitable. In general, the quality of the included studies was not as high as would have been desirable for a systematic review. RCTs of the field are needed. These should include humor interventions as well as other comparable interventions such as music and art interventions as well as a control group receiving usual care. Consensus should be sought for evaluating instruments and study settings for the different types of humor in order to provide meaningful data for comparisons and meta-analyses (Martin, 2008).

It needs to be noted that conducting research in palliative care settings needs to be designed with caution to avoid adding to the burden of patients and relatives with assessment and data collection. Also thorough coordination with nursing staff, physicians, relatives, other research staff and the patients themselves is crucial.

CONCLUSION

The review of the literature has shown that 20 years after the first systematic review, there is still only limited research available on the use of humor interventions and assessments in palliative care. Researchers from different fields agree that humor is not only a valuable resource for patients, but also for health care professionals working with patients at the end of life. A few studies have looked at the effect of humor interventions in this group of patients, mostly with promising results. Still, improved quality of life, better communication and sense of connectedness to staff and family members, the ability to distance oneself from the problems and burdens of the illness, and sometimes enabling a decreased perception of pain have been demonstrated. However, there is no consensus on a definition of humor, on types of interventions, or on the best method to assess the effects that would allow comparisons between published trials. Clearly, more research on the use of humor in palliative care is needed. Advancements in outlining the field of humor (Craik et al., 1996; Ruch et al., 2018) and the evaluation of standardized humor interventions (the Humor Habits Program; McGhee, 2010) might be fruitful for the context of palliative care as well.

Future research should use widely agreed definitions of humor and validated assessment instruments. Data from RCTs with humor interventions from different palliative care settings are needed. In addition, training interventions for palliative care teams would be useful, teaching them to use humor as a resource to prevent burnout, but also fostering an emotional atmosphere that allows patients to express humor in their interactions with staff. This would be an efficient way to introduce humor on a structural level with members of staff. By doing so, humor could

be implemented in palliative care with a long-term perspective rather than within the restricted setting of a clinical trial. Providing this kind of evidence will allow humor interventions to become part of the palliative care toolbox, to help lightening the burden of patients, caregivers, and health care professionals.

AUTHOR CONTRIBUTIONS

The study design and search strategy were conceived by LL-D, LR, and SH. LL-D performed the literature search and screened the search results. Publications were reviewed by LL-D and LR. The manuscript was prepared by LL-D with support from SH, WR, and LR. All authors critically reviewed and contributed to the manuscript and approved the final version.

ACKNOWLEDGMENTS

We would like to thank Nancy Preston, Catherine Joyce Letcher Lazo, Gülay Ates, and Birgit Jaspers for their support with conducting the systematic review and preparation of the manuscript. We would also like to thank Eckart von Hirschhausen for providing very helpful comments and for advising on earlier versions of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00890/full#supplementary-material

REFERENCES

Adamle, K., and Ludwick, R. (2005). Humor in hospice care: who, where, and how much? *Am. J Hospice Pall Med.* 22, 287–290. doi: 10.1177/104990910502200410

Armstrong, R. (2014). When to use the Bonferroni correction. *Ophthal. Physio. Opt.* 34, 502–508. doi: 10.1111/opo.12131

Auerbach, S. (2017). Are clowns good for everyone? The influence of trait cheerfulness on emotional reactions to a hospital clown intervention. *Front. Psychol.* 8:1973. doi: 10.3389/fpsyg.2017.01973

Auerbach, S., Ruch, W., and Fehling, A. (2016). Positive emotions elicited by clowns and nurses: an experimental study in a hospital setting. *Transl. Issues Psychol. Sci.* 2, 14–24. doi: 10.1037/tps0000055

Baile, W., Buckman, R., Lenzi, I., Glober, G., Beale, E., and Kudelka, A. (2000). SPIKES-a six-step protocol for delivering bad news: application to the patient with cancer. *Oncologist* 5, 302–311. doi: 10.1634/theoncologist.5-4-302

Bentur, N., Stark, D., Resnizky, S., and Symon, Z. (2014). Coping strategies for existencial and spiritual suffering in Israeli patients with advanced cancer. *Isr. J. Health Policy Res.* 3:21. doi: 10.1186/2045-4015-3-21

Borod, M. (2006). SMILES–toward a better laughter life: a model for introducing humor in the palliative care setting. *J. Cancer Educ.* 21, 30–34. doi: 10.1207/s15430154jce2101_11

Boyle, G., and Joss-Reid, J. (2004). Relationship of humour to health. A psychometric investigation. *Br. J. Health Psychol.* 9, 51–66. doi: 10.1348/135910704322778722

Branney, P., Witty, K., Braybrook, D., Bullen, K., White, A., and Eardley, I. (2014). Masculinities, humour and care for penile cancer: a qualitative study. *J. Adv. Nurs.* 70, 2051–2060. doi: 10.1111/jan.12363

Cain, C. (2012). Integrating dark humor and compassion: identities and presentations of self in the front and back regions of hospice. *J. Contemp. Ethnogr.* 41, 668–694. doi: 10.1177/0891241612458122

Claxton-Oldfield, S., and Bhatt, A. (2017). Is there a place for humor in hospice Palliative Care? Volunteers Say "Yes"! *Am. J. Hospice Palliat Med.* 34, 417–422. doi: 10.1177/1049909116632214

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155

Cox, G. (1998). Using humor, art, and music with dying and bereaved children. *Illness Crisis Loss* 6, 408–420. doi: 10.2190/IL6.4.d

Craik, K. H., Lampert, M., and Nelson, A. J. (1996). Sense of humor and styles of everyday humorous conduct. *Humor* 9, 273–302. doi: 10.1515/humr.1996.9.3-4.273

Dean, R. (1997). Humor and laughter in palliative Care. *J. Palliat. Care* 13, 34–39.

Dean, R., and Gregory, D. (2004). Humor and laughter in palliative care: an ethnographic investigation. *Palliat. Support. Care* 2, 139–148. doi: 10.1017/S1478951504040192

Delgado-Guay, M., Rodriguez-Nunez, A., De la Cruz, V., Frisbee-Hume, S., Williams, J., Wu, J., et al. (2016). Advanced cancer patients' priorities regarding wishes at the end of life: a randomized controlled study. *J. Clin. Oncol.* 33, 4273–4281. doi: 10.1007/s00520-016-3260-9

Demjén, Z. (2016). Laughing at cancer: humour, empowerment, solidarity and coping. *J. Pragmat.* 101, 18–30. doi: 10.1016/j.pragma.2016.05.010

Erdman, L. (1991). Laughter therapy for patients with cancer. *Oncol. Nurs. Forum* 18, 1359–1363. doi: 10.1300/J077V11N04_04

Goodenough, B., Low, L.-F., Casey, A.-N., Chenoweth, L., Fleming, R., Spitzer, P., et al. (2012). Study protocol for a randomized controlled trial of humor therapy in residential care: the Sydney Multisite Intervention of LaughterBosses and ElderClowns (SMILE). *Int. Psychogeriatr.* 24, 2037–2044. doi: 10.1017/S1041610212000683

Gremigni, P. (2012). *Humor and Health Promotion.* Hauppage, NY: Nova Science Publishers.

Herth, K. (1990). Contributions of humor as perceived by the terminally ill. *Am. J. Hosp. Care* 7, 36–40. doi: 10.1177/104990919000700108

Hulse, J. (1994). Humor: a nursing intervention for the elderly. *Geriatr. Nurs.* 15, 88–90. doi: 10.1016/S0197-4572(09)90023-4

Ioannidis, J. (2005). Why most published research findings are false *PLoS Med.* 2:e124. doi: 10.1371/journal.pmed.0020124

Itami, J. (2000). A trial psychosomatic treatment for cancer—meaningful life therapy. *J. Int. Soc. Life Inform. Sci.* 18, 162–166.

Kanninen, M. (1998). Humour in palliative care: a review of the literature. *Int. J. Palliat. Nurs.* 4, 110–114. doi: 10.12968/ijpn.1998.4.3.9112

Keltner, D., and Bonanno, G. (1997). A study of laughter and dissociation. Distinct correlates of laughter and smiling during bereavment. *J. Clin. Oncol.* 73, 687–702.

Kissane, D., Grabsch, B., Clarke, D., Christie, G., Clifton, D., Gold, S., et al. (2004). Supportive-expressive group therapy: the transformation of existential ambivalence into creative living while enhancing adherence to anti-cancer therapies. *Psychooncology* 13, 755–768. doi: 10.1002/pon.798

Koch, S., Gruber, H., Kortum, R., Reichelt, S., Martin, L., Warth, M., et al. (2016). Künstlerische Therapien in der Palliativversorgung – ein Review (creative arts therapies in palliative care - a review). *Hospizzeitschrift* 1, 20–26. doi: 10.1055/s-0043-121402

Konradt, B., Hirsch, R., Jonitz, M., and Junglas, K. (2012). Evaluation of a standardized humor group in a clinical setting: a feasibility study for older patients with depression. *Int. J. Geriatr. Psychiatry* 28, 850–857. doi: 10.1002/gps.3893

Kontos, P., Miller, K.-L., Colobong, R., Lazgare, L., Binns, M., Low, L.-F., et al. (2016). Elder-clowning in long-term dementia care: results of a pilot study. *J. Am. Geriatr. Soc.* 64, 347–353. doi: 10.1111/jgs.13941

Kontos, P., Miller, K.-L., Mitchell, G., and Stirling-Twist, J. (2015). Presence redefined: the reciprocal nature of engagement between elder-clowns and persons with dementia. *Dementia* 16, 46–66. doi: 10.1177/1471301215580895

Langley-Evans, A., and Payne, S. (1997). Light-hearted death talk in a palliative day care context. *J. Adv. Nurs.* 26, 1091–1097. doi: 10.1046/j.1365-2648.1997.00422.x

Low, L., Baker, J., Harrison, F., Jeon, Y., Haertsch, M., Camp, C., et al. (2015). The lifestyle engagement activity program (LEAP): implementing social and

recreational activity into case-managed home care. *J. Am. Med. Dir. Assoc.* 16, 1069–1076. doi: 10.1016/j.jamda.2015.07.002

Low, L., Brodaty, H., Goodenough, B., Harrison, F., Jeon, Y.-H., Haertsch, M., et al. (2013). The Sydney Multisite Intervention on Laughter Bosses and ElderClowns (SMILE) study: cluster randomised trail of humor therapy in nursing homes. *BMJ Open* 3:2. doi: 10.1136/bmjopen-2012-002072

Martin, R. (2001). Humor, laughter, and physical health: methodological issues and research findings. *Psychol. Bull.* 127, 504–519. doi: 10.1037/0033-2909.127.4.504

Martin, R. (2004). Sense of Humor, physical health, and well-being at work: a three-year longitudinal study of Finnish police officers. *Humor* 17, 21–35. doi: 10.1515/humr.2004.006

Martin, R. (2008). "Humor and health," in *The Primer of Humor Research,* ed V. Raskin (Berlin: Mouton de Gruyter), 479–522.

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., and Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being. *J. Res. Pers.* 37, 48–75. doi: 10.1016/S0092-6566(02)00534-2

Mathieu, S. (2008). Happiness and humor group promotes life satisfaction for senior center participants. *Act. Adapt. Aging* 32, 134–148. doi: 10.1080/01924780802143089

Maxwell, S., Kelley, K., and Rausch, J. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annu. Rev. Psychol.* 59, 537–563. doi: 10.1146/annurev.psych.59.103006.093735

McGhee, P. E. (2010). *Humor as Survival Training for a Stressed-Out World: The 7 Humor Habits Program*. Bloomington: AuthorHouse.

Müller, M., Pfister, D., Markett, S., and Jaspers, B. (2012). Wie viel Tod verträgt das Team? Eine bundesweite Befragung der Palliativstationen in Deutschland. *Schmerz* 23, 600–608. doi: 10.1007/s00482-009-0845-y

Piemonte, N. (2015). Last laughs: gallows humor and medical education. *J. Med. Humanit.* 36, 375–390. doi: 10.1007/s10912-015-9338-4

Richman, J. (2006). The role of psychotherapy and humor for death anxiety, death wishes, and aging. *Omega* 54, 41–51. doi: 10.2190/D0NX-7V03-W1H0-4614

Ridley, J., Dance, D., and Pare, D. (2014). The acceptability of humor between palliative care patients and health care providers. *J. Palliat. Med.* 17, 472–474. doi: 10.1089/jpm.2013.0354

Ruch, W. (2007). *The Sense of Humor: Explorations of a Personality Characteristic*. Berlin: Mouton de Gruyter.

Ruch, W., Heintz, S., Platt, T., Wagner, L., and Proyer, R. T. (2018). Broadening humor: comic styles differentially tap into temperament, character, and ability. *Front. Psychol.* 9:6. doi: 10.3389/fpsyg.2018.00006

Ruch, W., and Hofmann, J. (2012). "A temperament approach to humor," in *Humor and Health Promotion*, ed P. Gremigni (Hauppauge, NY: Nova Science Publishers), 79–112.

Ruch, W., and Hofmann, J. (2017). "Fostering humor," in *Positive Psychology Interventions in Practice*, ed C. Proctor (New York, NY: Springer), 65–80.

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. (2014). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., and McGhee, P. E. (2014). "Humor intervention programs," in *Handbook of Positive Psychological Interventions*, eds S. Schueller and A. Parks (Oxford: Wiley-Blackwell), 179–193.

Ruch, W., Proyer, R., and Weber, M. (2010). Humor as a character strength among the elderly. Empirical findings on age-related changes and its contribution to satisfaction with life. *Z. Gerontol. Geriatrie* 43, 13–18. doi: 10.1007/s00391-009-0090-0

Schultes, L. (1997). Humor with hospice clients: you're putting me on. *Home Healthc. Nurse* 15, 561–566. doi: 10.1177/104990919100800307

Schwartz, K., and Saunders, J. (2010). Laughter, Leininger, and home healthcare. *Home Healthc. Nurse* 28, 552–557. doi: 10.1097/NHH.0b013e3181f2f312

Shahidi, M., Mojtahed, A., Modabbernia, A., Mojtahed, M., Shafiabady, A., Delavar, A., et al. (2011). Laughther yoga versus group exercise program in elderly depressed women: a randomized controlled trial. *Int. J. Geriatr. Psychiatry* 25, 322–327. doi: 10.1002/gps.2545

Sridharan, K., and Sivaramakrishnan, G. (2016). Therapeutic clowns in pediatrics: a systematic review and meta-analysis of randomized controlled trials. *Eur. J. Pediatr.* 175, 1353–1360. doi: 10.1007/s00431-016-2764-0

Steinhauser, K. E., Christakis, N. A., Clipp, E. C., McNeilly, M., McIntyre, L., and Tulsky, J. A. (2000). Factors considered important at the end of life by patients, family, physicians and other care providers. *J. Am. Med. Assoc.* 284, 2476–2482. doi: 10.1001/jama.284.19.2476

Venter, M., Venter, C., Botha, K., and Strydom, M. (2008). Cancer patients' illness experiences during a group intervention. *J. Psychol. Afr.* 14, 549–560. doi: 10.1080/14330237.2008.10820234

Warren, B., and Spitzer, P. (2011). Laughing to longevity - the work of elder clown. *Lancet* 378, 562–563. doi: 10.1016/S0140-6736(11)61280-4

Weisenberg, M., Tepper, I., and Schwarzwald, J. (1995). Humor as a cognitive technique for increasing pain tolerance. *Pain* 63, 207–212. doi: 10.1016/0304-3959(95)00046-U

Wellenzohn, S., Proyer, R. T., and Ruch, W. (2016). How do positive psychology interventions work? A short-term placebo-controlled humor-based study on the role of the time focus. *Pers. Individ. Differ. Adv.* 96, 1–6. doi: 10.1016/j.paid.2016.02.056

Wild, B., Rodden, F., Grodd, W., and Ruch, W. (2003). Neural correlates of laughter and humour. *Brain* 126, 1–18. doi: 10.1093/brain/awg226

Zweyer, K., Velker, B., and Ruch, W. (2004). Do cheerfulness, exhilaration, and humor production moderate pain tolerance. *Humor* 17, 85–119. doi: 10.1515/humr.2004.009

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Linge-Dahl, Heintz, Ruch and Radbruch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

SYSTEMATIC REVIEW
published: 21 September 2018
doi: 10.3389/fpsyg.2018.01745

Investigating Humor in Social Interaction in People With Intellectual Disabilities: A Systematic Review of the Literature

Darren David Chadwick and Tracey Platt*

Psychology, University of Wolverhampton, Wolverhampton, United Kingdom

Background: Humor, both producing and appreciating, underpins positive social interactions. It acts as a facilitator of communication. There are clear links to wellbeing that go along with this form of social engagement. However, humor appears to be a seldom studied, cross-disciplinary area of investigation when applied to people with an intellectual disability. This review collates the current state of knowledge regarding the role of humor behavior in the social interactions of people with intellectual disabilities and their carers.

Method: A systematic review utilizing the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines was completed, which aimed to explore the current state of knowledge and quality of empirical evidence relating to humor in people with intellectual disabilities. Following this, articles were grouped thematically and summarized. A comprehensive search of four electronic databases (1954–2017) and additional search strategies yielded 32 articles which met the final inclusion criteria.

Results: Humor played a significant positive and negative role in the social interactions of people with intellectual disabilities. Research had investigated humor in the classroom and humor expression in different groups including those with autism, Down syndrome, Angelman syndrome, Williams syndrome, and Rett syndrome. Few investigations directly studied humor appreciation and comprehension. Humor comprehension was reportedly supported by gestures. Some groups with intellectual disabilities found non-literal humor (e.g., sarcasm, irony) more difficult to understand, which may affect social relationships. Various types of humor were found to be appreciated. The role of humor in relationship development, social facilitation, creativity, and stigma had all received some limited attention. Humor also played a role for carer groups in coping with and enjoying the caring role. Research varied in quality with few experimental studies and mainly quasi-experimental and well-conducted, qualitative studies.

Conclusions: This review revealed the importance of humor behavior in many aspects of the social lives of people with intellectual disabilities. Limited disparate research exists

OPEN ACCESS

Edited by:
Tim Bogg,
Wayne State University, United States

Reviewed by:
Jill Ann Jacobson,
Queen's University, Canada
Stephanie M. Carpenter,
University of Wisconsin-Madison, United States

****Correspondence:***
Darren David Chadwick
d.chadwick@wlv.ac.uk

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *27 November 2017*
Accepted: *29 August 2018*
Published: *21 September 2018*

Citation:
Chadwick DD and Platt T (2018) Investigating Humor in Social Interaction in People With Intellectual Disabilities: A Systematic Review of the Literature. Front. Psychol. 9:1745. doi: 10.3389/fpsyg.2018.01745

pertaining to humor in this group, suggesting the need for further robust research in this area, including more high quality primary research in the areas of humor production, appreciation, comprehension, and stigma.

Keywords: humor, learning disability, stigma, social support, developmental disabilities, autism spectrum disorders, social interaction

INTRODUCTION

Background and Rationale

The population of people with Intellectual disabilities are extremely heterogeneous. They vary greatly in etiology, support needs, and comorbidities (e.g., health problems, mental health issues and physical, and sensory impairments). The clinical definition of intellectual disabilities provided by The World Health Organization (World Health Organisation, 1992) within the International Classification of Diseases (ICD-10) involves three criteria: (i) impaired cognitive functioning; (ii) Challenges to adaptive functioning in at least two key areas (i.e., Communication, self-care, domestic skills, social skills, self-direction, community, academic skills, work, leisure, health, and safety); and (iii) Early developmental onset (<18 years). Clinical definitions embody the medical model approach to intellectual disabilities (Chappell et al., 2001), which focus on individual differences and are primarily deficit and pathology focused, considering disability to be the product of the individual's impairments. However, more recent conceptualizations have highlighted that the purpose of identification of impairments and challenges faced by people with an intellectual disability is to identify necessary supports, which are typically provided by paid support staff or family carers. This is to help ensure that these people maximize their life chances, participation, and inclusion (Van Loon et al., 2010).

Nomenclature has varied across time, geographies, and cultures, with terminology often co-opted and naturalized within society as terms of derision (e.g., idiot, retard etc.), which serves to societally disempower, stigmatize, and devalue this group of people (Siperstein et al., 2010). Alternative social model perspectives focus on the ways in which societal, social and environmental factors disable people with cognitive impairments (Chappell et al., 2001). Thus, intellectual disability is also considered a socially constructed term, both historically and culturally bound. People are labeled as intellectually disabled because they differ from a culturally defined idea of "normal" or "typical" intellectual functioning (Manion and Bersani, 1987), facing societal disadvantage as a result.

Although there has been a concerted effort since the 1980s to remove social and physical barriers and moves toward equal citizenship and inclusion, individuals with intellectual disabilities still face numerous challenges in many aspects of their daily lives. From human rights issues, to experiencing the intolerance of others, they often face social, as well as physical exclusion (Amado et al., 2013). These issues can, and do, lead to social isolation (Abbott and McConkey, 2006). This exclusion extends into the world of research where areas studied extensively in the typically developing majority are often seldom touched upon in people with intellectual disabilities; the study of humor appears to be one such area. Moreover, exclusion may also occur due to perceived additional challenges and effort involved in identifying, classifying, and targeting those with an intellectual disability for study recruitment. Possibly due to the adaptations needed to enable people with differing support needs to participate. Hence this paper aims to collate and summarize the existing state of knowledge around humor in people with intellectual disabilities.

The Role of Humor in Lifestyle and Wellbeing

There is evidence that eliciting positive emotions, such as fun and amusement are key components of positive social engagement. Therefore, it is also relevant for those with an intellectual disability. By its very nature, when spontaneous laughter, a non-verbal vocalized expressive communication signal of amusement occurs, it alters the state of consciousness and allows for *"care, trouble, and even physical pain"* to be forgotten (Hall and Allin, 1897, p. 8). One way that spontaneous laughter can be elicited is through humor. A myriad of situations can be deemed humorous. Humor appreciation goes along with individual differences but falls into three main areas, non-sense, incongruity resolution, or sexual (Ruch, 1992). Most individuals will find some aspects of such situations, or jokes, funny. When we share or engage others, in humorous situations, it serves a number of social functions. For example, Brown and Levinson (1987) suggest that jokes are positive politeness strategies for minimizing face threatening situations. Further evidence of the social function of humor in interpersonal relationships has been demonstrated by Holmes (2006), who showed that within the workplace, humor fosters collegiality and is also used to both construct and maintain good relations. As many people with intellectual disabilities require support outside of the inner family circle (MacTavish et al., 2007), there is a need for a better understanding of the role that humor plays within different relationships.

However, in order to fully understand this dynamic, one also needs to consider that individual differences will play a role. Being high in trait and state cheerfulness, low in seriousness and bad-mood relates to the temperamental basis for a sense of humor (Ruch et al., 1996). Those high in either state or trait cheerfulness will more readily be influenced by exhilarants (stimuli that elicits laughter and amusement)—one of those being the propensity to engage in humor. Cheerfulness has been shown to correlate negatively with both seriousness and bad mood, whereas seriousness and bad mood positively correlate

(Ruch, 1997). Being cheerful allows for a lower threshold for engaging in humorous behaviors and finding things amusing. Clearly, cheerfulness relates to positive affect and extroversion (Ruch, 1995), and thus those with the propensity to be cheerful and engage in humorous behaviors, are those we orientate toward.

As well as being linked to aspects of relationship building and maintaining, humor directly links to positive affect and enhanced quality of life (Kuiper et al., 1992). Consistently, positive affect has been shown to be associated with good health, well-being, and with health protective responses (Pressman and Cohen, 2005). Fredrickson (2000) suggested that the cultivating positive affect can optimizes health and well-being which has the potential to be induced or elicited by, among other things, humor (Baron et al., 1990).

Defining Humor

Moran (2013) noted "humor" is a term with a multitude of meanings. That it is both a "cognitive style"; a term for a stimulus, as well as the response to it (e.g., laughter). She also stated that humor is a term relating to complex interactions between individuals, or for a broader social process; a "personality trait," or an inherent characteristic; an ability to generate a response, produce a response, or detect/observe the two. Finally, she added that the complexity was compounded by the notion of "comedy" which has its own set of interpretations and expectations.

When discussing the positive benefits of humor in all aspects of social interactions it is essential to define this complex construct, as many theories exist and not all may have the functions being discussed in this paper. For example, laughing at someone, or mocking them, is a form of humor interaction but will neither foster good relationships nor elicit positive affect in the target of that mockery. However, teasing, which also relates to "play" laughing at a target, serves a pro-social function, and even seen as part of flirting behavior (Keltner et al., 2001). So how humor is perceived, will depend on the both the context, and the disposition of the actors within the interaction.

One classification was proposed by Schmidt-Hidding (1963) (see Ruch et al., 2017 this volume for an overview). These eight styles of the comic: *humor, fun, nonsense, wit, irony, satire, sarcasm,* and *cynicism* are a useful way of determining the differences. It should be noted in this context that humor is unique from the seven other comic styles and is classified as "coming from the heart" (see **Table 1**).

Table 1 shows how Schmidt-Hidding defined humor. The goal of this definition of humor is to raise our understanding of the incongruities of life, while remaining sympathetic for the human condition. This form of humor holds an understanding for the other, and any humorous judgment will benignly include oneself, rather than maliciously being directed at a target (e.g., when laughing at). The opposing dimension of this benevolent humor would be ridicule or mockery (see Ruch et al. this issue), where those deemed as being weaker or as being from an out-group become the object or target of derision and this fine-grained definition of the forms of humor would be important when investigating humorous interactions of and with people with intellectual disabilities.

TABLE 1 | Schmidt-Hidding comic style for humor.

Characteristics	Humor
Intention, Goal	To arouse sympathy and an understanding for the incongruities of life
Object	Creation in all its forms; human and real issues
Attitude of the agent as subject	Distant, affirmative, conciliatory, tolerant, love of the individual creation
Behavior toward the next	Understanding, benignly including oneself in judgments
The ideal audience	Jovial, relaxed, contemplative
Method	Realistic observation
Linguistic peculiarities	Ambiguous, without punch line; first-person Narration preferred; dialects, and professional jargon

Humor and Intellectual Disability

Little is known about humor in relation to people with intellectual disabilities. The development of the sense of humor is well established and broadly depends on cognitive, social, and individual difference variables. For verbal humor, such as joking, a greater cognitive capacity is required (McGhee, 1979), for example. McGhee (1980), also found that humor develops, from among other things, physical and verbal assertiveness and dominance. Due to the cognitive impairments which characterize intellectual disabilities, it is probable that people with intellectual disabilities may experience challenges in cognitively processing, comprehending, and appreciating humor. Moreover, physical and assertive dominance is likely to be more limited due to the limitations in self-determination, autonomy, and expressive communication.

As the participation in humorous interactions requires both en/and decoding of the play signals, associated craniofacial differences may affect the expressed enjoyment, which may be prohibitive of sustained interactions where humor is exchanged. Conversely, the genetic condition Angelman syndrome includes, as part of its behavioral phenotype, frequent expressions of smiling, and laughter. Though not always the case (see Oliver et al., 2002), these facial and vocalized expressions being displayed may simply occur when no stimulus is present (Nirenberg, 1991) or be disassociated from the context (Bower and Jeavons, 1967). Therefore, breeching the rules of communication that make interactions more difficult to establish and maintain.

Objective

This review aims to investigate the state-of-the-art in the existing empirical evidence regarding the interactional and experiential aspects of humor for people with intellectual disabilities, and those who support them. To this end a systematic review was conducted of the extant literature to address the following questions.

Research Questions

1. In what ways has humor and laughter been empirically explored and investigated in the lives of people with intellectual disabilities?
2. Is humor behavior a significant component of the social interaction of people with intellectual disabilities?
3. What is the quality of empirical evidence regarding humor in people with intellectual disabilities?

METHOD

Study Design

This systematic review study is underpinned by transformative and positive psychology epistemological perspectives, aiming to provide knowledge which can be used to improve the lives of people with intellectual disabilities. It collates and synthesizes literature underpinned by postpositivist, phenomenological, and constructivist epistemologies. From this framework, it aimed to highlight the emergent themes around humor interactions and the experiences of people with intellectual disabilities and their interaction partners (e.g., carers and family members). We predict that humor will play an important role in the social interactions of people with intellectual disabilities.

Systematic Review Protocol

This systematic review employed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009) as a guide to help ensure rigor. Also, a 5-step approach was utilized in implementing the systematic review as outlined by Khan et al. (2003) as follows:

Framing Questions

From the existing expertise of the authors and an initial scoping perusal of the extant literature it appeared that literature focusing on humor and people with intellectual disabilities was scant.

Identifying Relevant Publications

Search strategy

For this literature review a search in the Web of Science (SCI-EXPANDED, SSCI, and A and HCI) and EBSCO (British Education Index, Child Development and Adolescent Studies, Cinahl, Education Research Complete, ERIC, Humanities International Complete, Medline, Psychology, and behavioral sciences collection, PsycINFO and SocINDEX) databases was conducted in April 2017 (Search dates ranged between 1954 until 2017) and subsequently updated in September 2017. All English language papers containing the terms "Intellectual disability" or "learning disability" and "humor" or "humor" with the searches combined terms for humor and intellectual disabilities with the Boolean operator "and" in the title or abstract were identified (Note: the search engines also identified and included related terms in the searches). An example of database specific search terms (Psychinfo) is given in **Appendix 1**.

The titles of these studies (see **Figure 1**) were then inspected to ascertain whether they were likely to contain information, which could aid in answering the questions developed for this review. Once a primary list of articles had been identified a secondary review of the title and abstracts was conducted. Full texts were then gathered and reviewed for inclusion (see below for criteria). Reference lists of these identified studies which met the inclusion criteria (see below) were searched to identify further papers for inclusion. Full texts of salient articles identified this way were then gathered and full reviews conducted for inclusion.

In addition, in March 2017 a request for information on research relevant to humor and people with ID was sent to members of the International Association for the Scientific Study of Intellectual and Developmental Disabilities (IASSIDD) Quality of Life Special Interest Research Group and the Intellectual Disability UK Research mailing list, with the request subsequently being published in the TAC Bulletin in October 2015 (www.teamaroundthechild.com). Furthermore, the same enquiry was made to the International Society for Humor Studies (ISHS) members as well as listserve questions asking for relevant information. Finally, a paper presentation was given at the Annual ISHS conference in Montreal (July, 2017) where requests for information on relevant papers was made.

The authors subsequently identified and reviewed English language studies, focusing on humor interactions by people with intellectual disabilities. Contextually and due to the literature gathered, this paper is written from a UK perspective, but also incorporates research from North America, Asia, Australasia, and other parts of Europe (see **Appendix 2**).

Inclusion criteria

Studies were required to meet all of the following criteria: Collection of empirical data; peer reviewed; English language full text; published between 1950 and 2017. Inclusion criteria germane to the focus of the review were as follows:

(a) Studies included had to include as participants or be focused upon people with intellectual disabilities and/or those with developmental disabilities where intellectual disability is a core component e.g., people with Autistic Spectrum Disorder (autism), Rett syndrome.
(b) Core papers had to focus on humor and laughter in terms of either: (i) it being the primary focus of the paper; or (ii) it being a key finding from the empirical work.
(c) Particular attention was given to studies investigating or presenting findings which centered on the interactional components of humor in the lives of people with intellectual disabilities.

As we were also interested in how humor had been conceptualized and studied in the lives of people with intellectual disabilities, we also included some papers outside of these inclusion criteria which focused on analysis of secondary data in an area of study considered important to the lives of people with intellectual disabilities but seldom investigated in terms of primary data (i.e., the relationship between humor and stigma) or focused on carers and professionals who supported people with intellectual disabilities.

Exclusion criteria

The following exclusion criteria were applied: not peer reviewed or where the peer review status was deemed unclear;

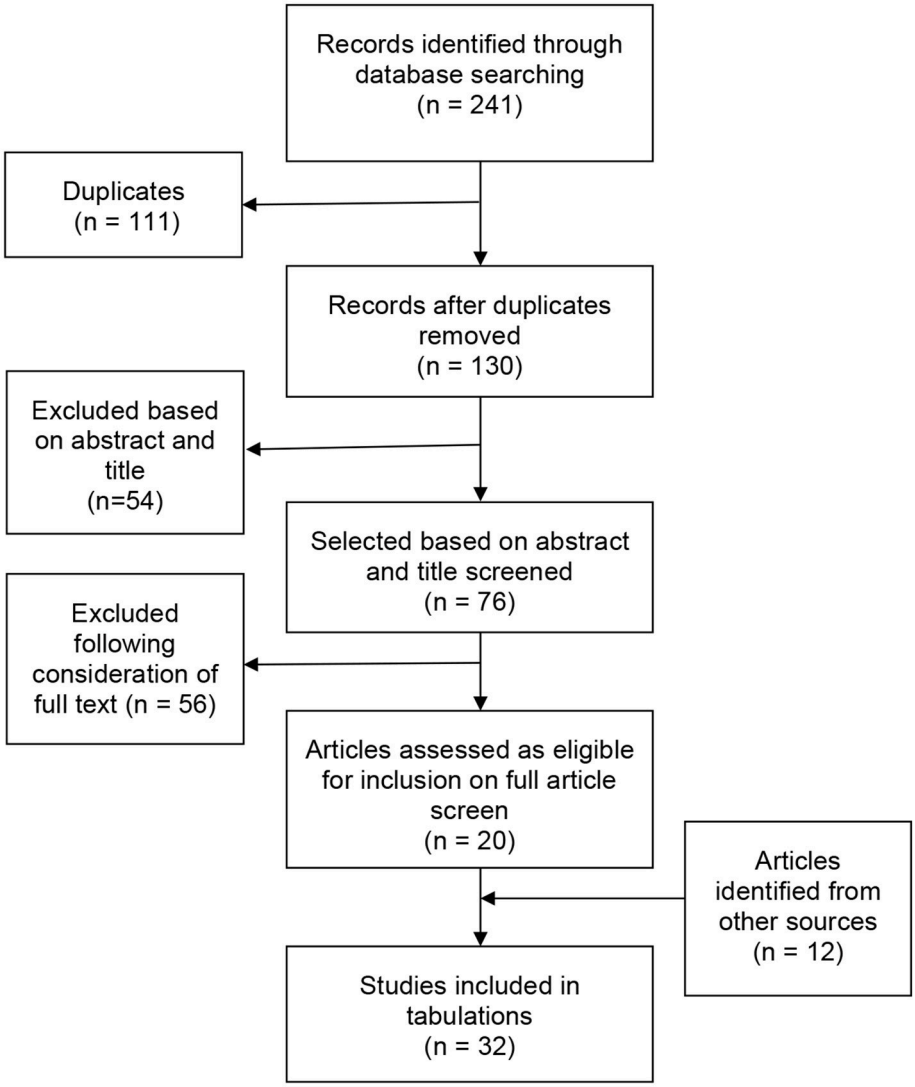


FIGURE 1 | Flowchart of study identification.

reviews, letters, commentaries, editorials, meeting, or conference abstracts; study relates solely to infants (less than 1 year of age). Those articles that did not relate sufficiently to either humor or intellectual disabilities were excluded. We also excluded papers which focused on people with developmental disabilities where intellectual disabilities are not a principal component (i.e., specific developmental disorders, attention deficit hyperactivity disorder, Asperger syndrome etc.).

Following secondary screening by title and abstract, we included two new exclusion criteria. First, any article that focused purely on phenotypic aspects where humor was not a central consideration but a description associated with the phenotype. Second, papers where the focus is on fun and enjoyment as ways of eliciting engagement rather than specifically focusing on interactional and experiential aspects of humor.

Summarizing the Evidence

The findings were summarized in two key ways. Firstly, tabulation of the papers pertinent to humor in people with intellectual disabilities that help shed light on main areas of research. This was supplemented by a thematic organization of the papers which developed from the extraction of data on the foci and findings from the studies in accordance with the specified research questions addressed. Meta-analysis was precluded by heterogeneity across studies.

Assessing Study Quality

Critical appraisal of the quality of the studies and risk of bias for the retained articles was conducted using the QualSyst quality appraisal tool for quantitative studies (Kmet et al., 2004) to allow comparability across studies. Authors independently generated quality scores of "yes," "no" or "partially" for each article on

each quality indicator (14 for quantitative and 10 for qualitative studies). To ensure replicability and objectivity, the goals of this systematic and conceptual review were registered on PROSPERO the International prospective register of systematic reviews, prior to the research being conducted (https://www.crd.york.ac.uk/PROSPERO/), registration number CRD42017070222.

Interpreting the Findings

Finally, based on the number and quality of studies reviewed, conclusions were drawn in relation to the questions posed at the outset of the review.

RESULTS

Summary of Included Papers

Electronic database searches identified 241 references, with 214 remaining after removal of 27 duplicates. Following an initial screen of the peer-reviewed papers based on the paper title and abstract, 138 references were excluded with 76 remaining for further screening. After examination of full text and the addition of studies cited within these, 32 studies met the inclusion criteria. These are summarized in **Appendix 2**. A flow diagram of the process undertaken is presented in **Figure 1**.

Of the 29 papers that met the inclusion criteria 13 were qualitative, 19 quantitative. With regard to the methodology employed within the qualitative papers: Two papers involved description of educational or mentoring programs, with only one of these qualitatively evaluating the program; Four studies used face-to-face interviews with either carers ($N = 3$), or people with intellectual disabilities ($N = 1$); Four conducted qualitative analysis of observational interactional data; finally, three papers, pertaining to humorous media representations of disability, involved analysis of comments on media or analysis, and reflection on film portrayals of intellectual disability. One of the papers (Johnson et al., 2012) analyzed both observational and interview data. Two of the qualitative papers considered humor specifically within online contexts (YouTube/Facebook).

Considering the methodologies used within the quantitative investigations: Three were descriptive studies using survey or observational methods, a fourth descriptive study used a cross-sectional design gathering data using specifically devised materials to ascertain comprehension/appreciation of humor; Two were longitudinal cohort studies. In addition, many studies employed quasi-experimental approaches ($N = 11$): One was a case series utilizing an ABAB design; Ten were comparative ex-post facto design studies taking either cohort ($N = 5$) or case-control ($N = 7$, two studies included both elements) approaches. The first considered potential differences across different cohorts with intellectual disabilities (i.e., People with autism, Down syndrome, Williams syndrome etc.), whilst the case-control studies compared people with intellectual disabilities to typically developing controls, who were often matched on age; Finally, two were true experiments, being small scale, within group studies of humor expression in people with Angleman syndrome observed under different interactional conditions.

With respect to the different subgroups of people with intellectual disabilities recruited into the studies, many ($N = 18$), unsurprisingly, had no specified etiology or diagnosis reported, with six papers being unclear about the extent to which people with intellectual disabilities were included in the study participant group. More specificity was evident in some papers ($N = 12$) with studies including people with autism ($N = 5$), Down syndrome ($N = 5$), Angelman syndrome ($N = 3$), Williams syndrome ($N = 3$), Prader Willi syndrome ($N = 1$), and Rett syndrome ($N = 1$) all encountered in the review, with some studies ($N = 6$) focusing on more than one group. Finally, paid and family caregivers or educationalists as key stakeholders in the social lives of people with intellectual disabilities were key participants in nine of the studies. With regard to the level of intellectual disabilities of participants in the studies, again for over a third of studies ($N = 12$) this was not specified, with roughly equivalent numbers focusing on people with borderline/mild ($N = 8$), moderate ($N = 8$), severe ($N = 6$), and profound ($N = 8$) intellectual disabilities. Again, many studies ($N = 8$) included people with different levels of intellectual disability. Finally, almost two-thirds of the studies ($N = 18$) focused on children and adolescents, with 11 papers focusing on adults, two papers spanned both age groupings and four studies did not specify the age group of the participants.

Synthesized Findings

Eight themes were determined from the post-scrutinized papers. Humor was studied in different and competing ways in the identified literature. To facilitate interpretation of data, findings are organized and presented by these emergent themes.

Humor Comprehension and Appreciation Among People With Intellectual Disabilities

It emerged that five studies explored humor comprehension and preferences in people with intellectual disabilities. This supports the notion that this important communication behavior is a neglected area of study in adults with these disabilities, requiring further investigation. Findings suggest that young people with intellectual disabilities show appreciation of humor (Degabriele and Walsh, 2010). However, humor comprehension was poorer in people with DS and WS compared with age matched to typically developing controls (Krishan et al., 2017). The authors suggest no association between humor appreciation and theory of mind in these participants. Difficulties in social problem-solving and incongruity understanding may impede humor comprehension in children with intellectual disabilities (Short et al., 1993). There is a tentatively indication that comprehension of non-literal humor, including irony and sarcasm, might be reduced in people with Williams syndrome (Godbee and Porter, 2013). This may be due to differential development of linguistic and cognitive systems, which may impact on their social interactions and relationships. Ironic jokes were misclassified as lies by adolescents with Williams syndrome and Prader Willi syndrome (Sullivan et al., 2003). With regard to support for humor comprehension, gestures may potentially be a useful support for humor comprehension in young people with intellectual disabilities (Degabriele and Walsh, 2010).

Degabriele and Walsh (2010) investigated the development of humor [appreciation (and) comprehension] in nine children

with intellectual disabilities aged between 7 and 11 in the Republic of Ireland. Participants with intellectual disabilities rated short video cartoon scenes and found that physical (85%) and visual (84%) humor scenes were more greatly appreciated by school aged children with intellectual disabilities compared with non-specific scene from a cartoon where no humor was evident (74%). Verbal humor was not appreciated significantly more than the non-specific scenes. The non-specific scene in the cartoon were highly appreciated too. Degabriele and Walsh (2010) in their study also investigated comprehension of humor and found that jokes supported by gestures (rather than pictures or acting) were significantly more understood by the young people with intellectual disabilities. Phonological jokes were best understood by participants but other joke forms (lexical, syntactic, and semantic) were also understood.

One study by Short et al. (1993) analyzed the humor skills of elementary school students, investigating those with and without intellectual disabilities on the dimensions of humor comprehension, but also included production and appreciation. This rather early study used common American terminology for their groups, which would not be understood world-wide. For example, they had an achieving normally group, a group with learning disabilities and a group with developmental handicaps. The group with learning disabilities often showed no difference to the achieving normally group and this is most likely due to this group included children with IQ >85. The children with developmental disabilities lacked differential sensitivities to cartoons, which the authors suggest is down to their social problem-solving deficiencies or ability to represent the problem to understand the incongruity and the process of resolution. However, the authors did not take other forms of humor appreciation (e.g., sexual/scatological and non-sense humor) into consideration in their conclusions, which limits this more comprehensive and insightful study.

Utilizing a comparative study to investigate aspects of humor comprehension and its connection to aspects of Theory of Mind, Sullivan et al. (2003) had groups of adolescents, one with William syndrome, another with Prader-Willi syndrome, and a group which had non-specific intellectual disabilities try to distinguish between different forms of non-literal language used in stories that ended in either a lie, or an ironic joke. To do this, the authors manipulated the structural differences in the child's second-order belief about the adult's knowledge of the truth of the situation. This research found that almost none of the participants in any of the groups were able to correctly classify the ironic jokes, instead judging them to be lies because they did not correspond to reality. Their errors were similar to those made by younger normally developing children, but contrasted with those made by brain-damaged adults. The authors state that the consequences of this inability to distinguish between intentionally false utterances, intended as ironic jokes vs. those intended to deceive, may seriously impairs these adolescent's ability to relate to others in everyday social situations.

A similar study by Godbee and Porter (2013) pursued two aims in their study. They aimed to investigate the comprehension of sarcasm, metaphor and simile in people with Williams syndrome compared to neuro-typical controls, secondarily, they aimed to examine the association between non-literal language comprehension and a range of other cognitive abilities, both in Williams syndrome and in the neuro-typical population. Matching both chronological and mental aged groups, all participants listened to randomly selected stories. After each comment from a story character, the participant was asked what the character meant by their comment. The comments were coded for whether the reply demonstrated correct understanding of the non-literal meaning of the comment; otherwise, they were given a zero score. Several types of responses were awarded a score of 0, including: literal explanation; ambiguous explanation; irrelevant explanation; no explanation; recognition of non-literal language without interpretation (e.g., he doesn't mean it); and supply of another non-literal comment without interpretation. For the comprehension of non-literal language, the individuals with Williams syndrome performed significantly below typically developing chronological age matched controls. However, they did not demonstrate significant differences to typically developing mental age matched controls. For the typically developing controls, each of the cognitive measures was strongly correlated with each of the measures of non-literal language comprehension. The same relationships were not always found for participants with William syndrome. In particular, sarcasm comprehension in participants with William syndrome was not significantly correlated with any of the assessed cognitive abilities. The expressive vocabulary was not significantly correlated with any measure of non-literal comprehension. The pattern of correlations between non-literal comprehension and cognitive abilities in the group with WS, relative to the control group suggests that perhaps the linguistic and cognitive systems that underpin non-literal language comprehension in neuro-typically developing individuals interact and integrate in different ways to individuals with Williams syndrome.

A further study conducted by Krishan et al. (2017) investigated humor comprehension and use of mental state language in groups of individuals with Williams syndrome and Down Syndrome relative to each other and to a neuro-typical control group. These groups were chosen for the link of humor to Theory of Mind (ToM) to fill the gap in the literature which focuses on those with ToM deficits such as those with autism. Relative to the control group, both groups of participants with intellectual disabilities had poor humor comprehension. The William Syndrome and Down Syndrome groups had comparable performance to each other, as well as to a mental age matched control group, differing only in physical emotion words, where those with William Syndrome used fewer. The use of cognitive words was less for both groups with intellectual disabilities. The authors also suggest that humor appreciation is not associated with theory of mind in people with Williams syndrome and Down syndrome.

Humor, Social Facilitation and Social Capital

Studies reported findings where humorous exchanges, in particular banter and sharing of humor, were identified as significant, enjoyable components in the facilitation, development and maintenance of social relationships, and capital. They also identified how humor served to enhance

social closeness facilitating intimate shared connection between people with intellectual disabilities and those supporting them. Attunement of those providing support to those with more significant cognitive impairments was highlighted as positive components of social interaction, including attuning of the type of humor (e.g., slapstick).

Griffiths and Smith (2016) aimed to identify the process that regulates communications of people with profound and multiple learning disabilities (PMLD) with others. They used fine grained (second-by-second and frame-by-frame) qualitative analysis of video-recorded observational data from two dyads of people with PMLD and carers, in a developmental disability center for young adults in Ireland. Glasserian grounded theory was the analytic approach used to develop a theory of attuning. This theory asserts that communication takes place in the context of a physical setting. The setting influences the state of mind of those within the interaction. In turn, this influences the stimuli they present which may or may not be attended to by the communication partner. Attending to stimuli is also affected by the setting in which the interaction takes place. Engagement occurs when one player attends to the stimuli of another, the determining factor is the process of attuning. Attuning affects and reflects the feeling of the communication partner in terms of whether they offer stimulus to their partner, attend to the other, engage with the other and act. All of these processes feedback to each communication partner to influence their state of mind (being). Thus, attuning is an implicit, cognitive process that is not observable in of itself, but there are behaviors which are observable and which indicate attuning is taking place. Here, humor is evident in the example data used to illustrate the theory. Humor is described as an indicator of empathic harmony and pro attuning and a manifestation of: (i) close psychological contact via a smile; (ii) shared amusement via a smile or laughter. In a sister paper focusing on the same data set, Griffiths and Smith (2017) briefly mention joking as an exemplar of solidarity in a group situation which could foster an intense level of attuning between people with PMLD and their carers. Although this evidence may seem less substantial, it is a good indicator of the importance of how this form of humorous banter facilitates in-group cohesion.

Johnson et al. (2012) similarly studied the lives of six people with severe intellectual disability, with symbolic but non-linguistic communication skills, and their interactions with others. In this Australian study, they observed interactions between people with severe intellectual disabilities and others and interviewed interaction partners and again analyzed via constructivist grounded theory. Social interactions took place when dyads and groups "shared the moment" this central theme was characterized by hanging out and having fun together. The latter of these involved both routines, utilizing activities such as mimicry, rhythmic play, games, songs, and comedy. Comedic interactions observed comprised several different forms of humor including vulgarity, pranks, jests, and banter. The exert of involvement and initiation differed both across the types of humor (banter occurring more often between support staff but involving people with intellectual disabilities) and participants (three participants were observed to initiate humor, whilst the other three adopted the role of active respondents and joined in with humorous interactions). More vulgar humor was sometimes supported and encouraged and other times discouraged. The humorous interactions were described as animating and enjoyable for the parties participating, fostering a sense of belonging. It is hypothesized within the paper that visual humor (i.e., slapstick) may be enjoyed more by participants because it relies less on verbal skills. Teasing was also observed and was noted to be used by familiar staff to improve the mood of people with severe intellectual disabilities.

Chadwick and Fullwood (2017) conducted a small UK and Ireland based qualitative, phenomenologically focused, study of the online lives of eleven people with mild to moderate intellectual disabilities. Two had Down syndrome and five had autism. They identified two global themes around the online lives of these participants (i) Online relatedness and sharing; (ii) Online agency and support. For the former theme, one basic theme 'coming together on social media with friends and family to chat and share' related to sharing online life and being connected to significant others which supported maintenance and development of social capital with family and friends. One important component of these interactions referred to by four of the eleven participants was humor, which took the forms of playing practical jokes, banter, and 'taking the Mick out of each other' and these interactions were viewed positively by participants as the most enjoyable online activities they engaged in.

Classroom Humor and Laughter

Four papers focused on humor in the classroom and one on changing behavior in pre-school children. Schnitzer et al. (2007) investigated the Feuerstein's Instrumental Enrichment Program (FIEP) as a means of increasing social, cognitive function. Here the comprehension of humor, even complex humor, was one goal of the experimental group who had the FIEP intervention. As part of the GOLD program, designed to support children who were gifted (defined as having IQ potential determined by a screening committee) and also had intellectual and/or developmental disabilities, Bees (1998) highlighted that humor was encouraged and having time for laughter was a way of helping the children relax. These papers highlight the conflicting perceptions of laughter and humor within the classroom context. Unabashed, shrill laughter, was not a welcomed behavior, yet prescribed moments of humor and laughter were seen as beneficial. However, laughter and humor are, by their very nature, organic and as beneficial as allowing for moments of hilarity are, maybe these benefits flourish more when not so prescribed? This idea was reiterated by the study of Jones and Goble (2012), who investigated effective campus mentors in partnerships with students with intellectual disabilities. They identified the key components for effective mentoring partnerships. One of those was of prioritizing fun and socializing, which, they suggest, should happen spontaneously. An afterschool program was designed to enhance character trait development. It utilized high school and college mentors to both introduce the program's curriculum and to help build friendships (Muscott and O'Brien, 1999). This component was key to the program's success, as the

outcome was that the children with intellectual disabilities had found that learning about character was fun and the program rewarding.

The play behaviors of school age children with intellectual disabilities were assessed by the observational Assessment of Ludic Behavior instrument which measured three dimensions: play interests, play abilities and play attitude (Messier et al., 2008). The findings of this study showed that the sense of humor (as well as enjoyment of challenge) were less present than other elements of the test. A component of the ludic attitude dimension, the sense of humor factor, was scored when the child was deemed to show a sense of humor, an understanding of comical situations, and laughs. The authors argue that this deficit is due to humor requiring a complex cognitive ability. Yet, they also raised the point that studies have shown these are often in conflict to parent's observations, who reveal higher scores than therapists do. This demonstrates the benefits of having mentors who build friendships and those close parental ties to the children, as they can often better see and attribute the subtle differences.

Humor and Creativity

People with intellectual disabilities have been shown to be creative in their humor use (Johnson et al., 2012). People with autism and intellectual disabilities, who have been found to display less playful pretending (Hobson et al., 2009), have demonstrated the ability, with prompting, to enhance their humorous creativity (Gagić et al., 2015).

Johnson et al. (2012) discussed the issues around the language skills of adults with severe intellectual disability and how they are limited and this impacted on the range of humorous forms. However, they found that the participants of their study did demonstrate "creativity and variety" (pp. 338) in their attempts at humorous social interaction.

This creative use of humor in social interaction was very different for those children with autism, for example. Hobson et al. (2009), measured both spontaneous and modeled symbolic play, in those with and without autism. They predicted that play for children with autism would lack social-developmental markers. Speculating that this form of play with an investment in the symbolic meanings given to play materials, creativity, and fun. They found that children with autism displayed less playful pretending and investing in symbolic meaning of the items given to play with. However, the study did not have ratings for the produced observed creativity with the play, which would be required, given the low expressivity of children with autism.

(Gagić et al., 2015) used humorous content as an indicator of the expression of creative ability in a drawing task. They used a method of prompting to encourage creative thinking around the art and showed an increase in the humor within the work, post prompting. This kind of prompting and engagement with play may be a way of engaging those who seems to be limited in the social aspects of creative play, such as children with autism.

Play, Humor and Laughter in Children With Autism and Down Syndrome

Some of the identified papers and themes, focused on specific groups of people with intellectual disabilities associated with specific syndromes and how humor is understood, expressed and used in these groups. Diagnoses including Autism, Down syndrome, Angelman syndrome, Williams syndrome, Prader Willi syndrome and Rett syndrome were studied, here we collate research focusing on the first two of these groups.

Four papers investigated the play, the humor and laughter of children with autism and, in one instance, compared them with children with Down syndrome. Hobson et al. (2009) testing pretend play abilities in children with autism and children with learning and developmental delays but without autism, found that although both groups were similar in the mechanics of play, the children with autism showed lesser qualities of playful pretend meaning the awareness of self as creating meanings, investment in symbolic meanings, creativity, and fun. Although this paper focuses on the deficits relating to autism, conversely it highlights that the children with the intellectual disabilities in this sample do not lack these qualities of play.

Reddy et al. (2002), interviewed parents who reported on specific incidents relating to their child's humor. Interview questions focusing to the type of things the child normally finds funny or laughs at, the attempts to join in with others' laughter, repeating others' laugh events (clowning), and teasing by the child or parent were compared in a group of children with autism and a matching group with Down syndrome. Significant differences were found that the majority of parents of children with Down syndrome reported their child tried to join in when others are laughing, whereas only five of the 18 children with autism had such behavior noted by their parents. Similar differences were reported for trying to make others laugh and teasing conditions. Group differences were observed by coding laughter episodes of videoed play sessions. No group differences were found in the frequency of laughter episodes or the rate per hour of laughter started by the children or in interactive situations. This study highlighted that the children with Down syndrome displayed all typical infant development of humor whereas the children with autism only showed some aspects.

Focusing on the vocal expressions of laughter, produced by children with and without autism, Hudenko et al. (2009) recorded laughter during play involving age appropriate humor stimuli that was based on ideas of humor development by McGhee (1979). The children with autism only exhibited one type of laughter compared to the comparison group, who produced two types. Other variables (fundamental frequency, duration, and number of laugh bouts etc.) did not show group differences. The authors argue that their findings indicate that the laughter of children with autism are responses to internal positive states, whereas those children without autism also utilize laughter to negotiate social interactions.

The remaining study was conducted by St. James and Tager-Flusberg (1994). They investigated the cognitive developmental, social and intentional aspects of naturalistic humor in two groups of six children, one with autism and the other Down syndrome. The children were filmed when interacting with their mothers in twice monthly, 1 h long, video-taped sessions. The authors report that the group of children with autism produced less humor overall and less humor that involved non-verbal incongruity. The only two jokes observed were created by children with

Down syndrome. As with the other studies, deficits in the social-cognitive aspects of humor were highlighted for the children with autism.

Laughter as Disruptive, Unelicited, or Inappropriate Social Behavior

In addition to being a means of facilitating social closeness, supporting learning and creativity, research had also focused on laughter as an unwanted, disruptive, unelicited and/or inappropriate, social behavior. Some studies focused on reducing such behavior via corrective intervention, others investigated the trajectory of unwanted laughter as people age, whist other considered whether laughing behavior was unelicited or a response to social and environmental stimuli.

Reducing Disruptive Laughter

A paper by Schieltz et al. (2011) investigated a dedicated program which was designed to target disruptive social behavior in pre-school children. Schieltz and colleagues evaluated functional communication training as a means of correcting destructive and disruptive behaviors, one of the non-targeted disruptive behaviors was *shrill laughter*. Despite the lack of targeting post intervention all undesirable behaviors, including the shrill laughter, reduced.

Night Laughing in People With Rett Syndrome

Rett syndrome is a rare neurodevelopmental disorder which usually affects females. It is associated with a mutation in the MECP2 gene (Amir et al., 1999). Sleep problems have been noted as common in this group and are incorporated into the diagnostic criteria (Kaufmann et al., 2010). These problems manifest as night laughing or night screaming in young children (Hagberg, 2005) and linked to immature sleep patterns (Nomura, 2005) and can negative affect parental relationships and social activities (McDougall et al., 2005).

Wong et al. (2015) studied sleep disorders in this group in Australia in a longitudinal cohort study gathering data at 6 time points over 12 years. They found that more than 80 per cent had sleep problems, but prevalence decreased with increasing age. Night laughing was frequently evident. It occurred in 77 per cent when younger and those with a larger gene deletion had higher prevalence of night laughing. They found that behavioral and pharmacological treatments were associated with a 1.7 per cent reduction in risk of further sleep problems.

Laughter in People With Angelman Syndrome

Angelman syndrome occurs in 1 in 10–12,000 live births and is associated with various degrees of intellectual disabilities (though typically severe to profound cognitive impairment) and greater impairment of expressive over receptive speech (Steffenburg et al., 1996). Physical signs of Angleman syndrome include ataxic gait, craniofacial differences, hand flapping, and hypopigmentation. The behavioral phenotype includes elevated levels of smiling and laughing (Adams et al., 2015), with early studies describing smiling and laughing in this population as excessive and occurring without stimuli. A body of research work has been conducted by Oliver and associates incorporating humor related behaviors (Laughing/smiling) and exploring the role of social and environmental influences on these behaviors. Due to the rarity of this condition these investigations involved small numbers of participants.

Oliver et al. (2002) in a case series of three people with Angelman syndrome living in the UK and Greece found that smiling and laughing was greatest when enthusiastic interaction was taking place, moderate in instructional interactions and when there were others present but no interaction (proximity condition), and lowest when individuals were alone. This finding disputes the earlier assertion that smiling is inappropriate and is not elicited by environmental stimuli indicating a social function for these behaviors and an interaction between the phenotype and environment.

In 2015 Adams et al. published a brief report on a longitudinal UK based study of laughing and smiling in 12 young people with Angelman syndrome across full interactional (with eye contact), interactional (without eye contact) and proximity conditions. The findings revealed that smiling and laughing reduced with age during full interactions for participants as they move from childhood into/toward puberty/adolescence. Thus, an interaction between behavioral phenotype, environment and aging is apparent from the data. The need to explore further how puberty affects physical, emotional, and social development in people with intellectual disabilities is highlighted here.

Mount et al. (2011) in a study of the effects of familiarity and eye contact on the social behaviors of people with Angelman syndrome found that although they were the most variable social behaviors observed, more laughing/smiling was observed with familiar contacts when eye contact was maintained, though this finding did not reach statistical significance, likely due to the small sample size ($N = 15$) in the study.

Humor as a Coping Strategy for Carers and Support Staff

One of the ways in which humor and shared humor operated as important aspects of the social worlds of people with intellectual disabilities was as a coping strategy carers used to manage and bring enjoyment and value to the caring responsibilities and societal stigma which accompanied their role. This was found in three of the identified articles.

MacDonald et al. (2007), in a cross-sectional descriptive survey study of respite care and coping strategies employed by family carers in Ireland, found that over 80 per cent of both male (81.5%) and female (81.8%) carers reported that 'seeing the funny side of the situation' was employed as a managing meaning coping strategy. Such strategies were frequently employed by carers to enable them to maintain a sense of humor regarding their role. It also reportedly supported them to remind themselves that the person with a learning disability who they supported, was not to blame for their behavior and support needs.

In a qualitative interview based study with eight paid staff members working on a treatment program for sex offenders with intellectual disabilities, Sandhu et al. (2012) investigated the emotional challenges these staff faced. Interpretive phenomenological analysis revealed that humor was, once

again, used as a way of dealing with negative emotions arising from working in this context and with this group of people. Banter and a "sick" sense of humor reportedly helped staff to process negative emotions that they otherwise may carry with them. There was also a sense of sharing and bonding over this "sick" sense of humor that was seemingly viewed by respondents as exclusive to colleagues working in this field. In addition to being a coping strategy to help staff process the stress of work, humor was also interpreted as a defense mechanism which prevented the staff team from exploring the personal and emotional impact of work. The authors viewed this as having potentially negative consequences for the wellbeing of staff and the therapeutic process for clients. Another feature of the narratives from staff was that empathy for the people with intellectual disabilities that they worked with was challenging and complex due to their emotional responses to the offending behavior.

Forster and Iacono (2014) conducted a phenomenological study of the perceptions of communication interaction of three residential support workers who knew one individual well (having worked with them for 2 and 15 years). The study revealed that communication with the person with PMLD comprised: ascription of meaning, attachment, touch, movement away from age-appropriateness, learning to interact, and valuing knowledge and existing skills. With regard to humor, laughing was a valued part of interactions with the person with PMLD, it was viewed as something of a leveler within interactions, as both the support staff and person with PMLD could share laughter on more of an equal footing. It was deemed a positive part of the interactions.

Support staff enjoyed seeing laughing in the person they were supporting and felt that smiles and signs of positive affect made the more negative aspects of the support worker role worthwhile. The staff also valued sharing sad times with the person with PMLD, as well as laughter, indicating that humorous exchanges are only one important component of interactions and relationship building. Interactions involved continual ascription of meaning to the behaviors of the person with PMLD. A strong emotional component was evident in the descriptions of interaction, which also involved physical touch, and built attachment between the person with PMLD and the support staff. This was reportedly somewhat at odds with the professional role of being a carer. The idea of age-appropriate interactions was critically questioned by the phenomenological accounts.

Humor and as an Indicator of Disablist Attitudes and Stigma

Humor was a key component in papers investigating stigma and prejudice directed toward people with intellectual disabilities. Intellectual disability was also investigated as an object of humor and consequentially an indicator of disablist attitudes and stigma. Four papers had this focus within the review. Two investigations focused on representations of people with intellectual disabilities in the media. Goggins (2010) highlighted the complexities and lack of adequate academic debate around the distinction between laughing at and laughing with people with intellectual disabilities. The study used the case of a documentary "Laughing at the disabled" (Later renamed "Down Under Mystery Tour") to explore the challenges around this debate within media and disability studies. It tackles some of the challenges inherent in research with and on people with disabilities and engages with the idea that further work and debate around these issues is needed.

Fudge Schormans et al. (2013) in a co-researched critique of a film featuring a disabled superhero "Defendor" discuss the importance of the film for people with and without intellectual disabilities and the representations of disability therein. They highlight the importance of the film but in one section the point is made that instead of being a positive representation of disability, instead one of the authors believed it would likely lead non-disabled viewers to see his attempts to be a superhero as humorous and funny and would simply laugh at the character. This made it more challenging for this person to relate to the central character within the film and highlights the tension between having positive representations of people with disabilities and the possibility that the non-disabled majority might simply laugh at them.

Johanson-Sebera and Wilkins (2014) wrote a paper investigating the uses and implications of the term "retarded" from its original meaning as a special educational classification, to how it is used now, based on the analysis of the social media platform YouTube. Five themes for where the where and how the term was used was found. Those were (a) the traditional use of the term, (b) in humorous context, (c) to insult or criticize, (d) as a substitute for other words, and (e) as hip hop slang. Although the stigmatizing nature of term is highlighted, for the humorous context theme the word was reportedly repurposed as a positive term, akin to recent changes to the word "sick," being slang for "great," in Western youth culture. Although changes were made so that person first language was adopted in the 1990s in accordance with the Individuals with Disabilities Education Act, it is clear that the general use of the term remains complex and holds negative connotations and is therefore stigmatizing for those with intellectual disability. This is especially poignant when one considers that people with intellectual disabilities may not be as able to "reclaim" the word, as other marginalized populations have with related abusive terms.

Only one cross sectional UK survey by Ali et al. (2016) collected primary data on stigma and considered humor as an operationalized aspect of stigma. This investigation found that older males with moderate intellectual disabilities were more likely to report stigma (being treated differently, like children and made fun of) compared with females. Additional impairments such as sensory, mobility and speech difficulties did not correlate with reported stigma. Overall across the 229 participants approximately one third of participants with intellectual disabilities responded affirmatively to the items "people laugh at me because of the way I talk (33.19%)/look (31.88%)." The authors highlight the need to tackle stigma at both a societal and at an individual support level.

Quality Assessment of the Literature

The quality of papers selected for inclusion in the review was assessed for all papers by both authors using the standard quality assessment for evaluating primary research papers

(Kmet et al., 2004). Qualitative and quantitative studies were evaluated based on 10 and 14 criteria respectively, which considered design, sampling, methodology, analysis, results, rigor and trustworthiness and conclusions. For each criterion, papers were scored either 2 (good), 1 (partial fulfillment), 0 (not fulfilled) or N/A (not applicable/relevant) with the exception of the qualitative criteria "Use of verification procedure(s) to establish credibility" which was scored as 1 (fulfilled) or 0 (Not fulfilled) (For this item ETA was used as the measure of inter-rater agreement and not Spearman's rho correlation). Dividing by the total possible score resulted in a composite overall score ranging between 0 and 1 (see **Appendix 2**), with <0.5 indicating limited quality, 0.5–0.7 adequate quality, 0.7–0.8 good quality, and >0.8 being indicative of strong quality. Inter-rater agreement of the ratings was within an acceptable range for both the qualitative (N = 10, rho = 0.791–1.00) and quantitative (N = 19, rho = 0.745–1.00) ratings. Following inter-rater agreement analysis, disagreements between raters were discussed until agreement was reached.

A mean score was computed for each article to provide an overall rating of quality (see **Appendix 2**). In addition, a mean score for each of the criteria was used to indicate the relative strengths and limitations across all 32 included studies. Overall the majority of the papers reviewed were rated as strong (N = 18) or good (N = 10) quality. Few papers were rated as adequate (N = 3) or limited (N = 1) quality. For the quantitative papers in the study none were rated as limited quality, three adequate quality, eight good quality, and eight strong quality. For the qualitative papers in the study one was of limited quality, none adequate quality, two good quality, and ten were strong quality papers. Considering mean quality criteria scores across the papers, the quantitative papers strengths lay in well described objectives, participant group descriptions, use of robust outcome measures and detail and sufficiency of results reporting. Weaknesses were evident in the lack of experimental and intervention studies, lack of control for confounding variables and lack of variance estimates (i.e., confidence intervals) presented in study findings. Due to the limited number of intervention studies, partial bias around outcome measurement and intervention description as evaluated in the Kmet quality assessment was only present in one quantitative study, Wong et al. (2015). Bias in description and recruitment of participant groups was more prevalent in the quantitative studies with four having partial bias ratings due to their inadequate description of participant groups. Similarly, for the qualitative investigations the sufficiency of objective explanation and context description, sufficient to allow transferability of findings, were strengths. Weakness included inadequacies in theoretical framework, data collection, and data analysis accounts and a lack of inclusion of reflexivity and credibility verification checks to enhance study trustworthiness. Future studies should be mindful to incorporate aspects lacking in prior studies to enhance the rigor of evidence around humor and intellectual disability. Given the limited number of relevant studies available no exclusions were made based on quality scores.

DISCUSSION

Summary of Main Findings

After scrutinizing the extant literature, this systematic review yielded 32 papers, from which eight themes were extracted. The meanings of humor investigated characterized it as a complex interactional process, a social process, a facilitator of development, a response to social and interactional stimuli, and an inherent characteristic. This is in line with the complexity and varying conceptualizations and meanings previously assigned to humor (Moran, 2003; Coogan and Mallett, 2013). Humor was found to be a significant aspect of the social interactional lives of people with intellectual disabilities and those who provide them with support, though the extant literature reviewed was currently limited and diverse in both focus and quality.

The Role and Functions of Humor in the Social Lives of People With Intellectual Disabilities

Humor comprehension and preference had not been extensively studied in the literature. The few studies that had explored this area revealed that humor comprehension can be supported by gestures. People with Williams syndrome found non-literal humor (e.g., sarcasm, irony) more difficult to understand which may impact on their social relationships. People with intellectual disabilities appreciated many various types of humor.

Research findings evident in the reviewed studies highlighted the utility and value of benevolent humor in facilitating social relationships, social closeness, carer coping and carer value, and enjoyment of the caring role. Despite this, there were few studies that specifically focused on the utility of humor in developing relationships and social closeness. Two studies highlighted the importance of shared humor for good interactions of people who do not use formal means of communication (i.e., people with PMLD). Humor was found to be an important component of online interactions for people with mild to moderate cognitive impairment and those with autism, Down syndrome and intellectual disabilities. For people with complex support needs and more severe cognitive impairments (e.g., those with Angelman syndrome), humor was also found to be a response to familiar interactional stimuli. Given the importance of humor in these contexts, it would behoove future research to consider humor as more of a key variable in interactions between people with intellectual disabilities and significant others across a variety of contexts.

Benevolent humor and sharing of social moments were key in fostering relationships, serving important social functions of humor in the lives of people with intellectual disabilities. Humor interactions, are by their very nature, complex. They can relate to laughing along together, while experiencing a shared moment (Chapman, 1983). Or perhaps, be playful, pro-social teasing or bantering, which uses fake scorn and derision to help build trust within groups or social interaction partners (Keltner et al., 2001). Humor can also be a means of trying to correct others who are deemed to be breaching social norms of a group, as satirists do to politicians (Ruch and Heintz, 2016). However, humor too can be malicious and hurtful (Billig, 2005). Mockery and ridicule serves the purpose of socially excluding the target. How we

determine the intent of the humor depends on many things. At an interacting group level, it may depend on whether one is the target, the bystander/observer or the active humor protagonist. It may also depend on your general disposition or the momentary state you are in (Ruch et al., 1996).

A relationship was identified between humor and stigma. Stigma has been found to be linked to negative evaluative beliefs about the self, experiences of feeling different; with this internalizing experienced stigma negatively affecting the psychological wellbeing of people with intellectual disabilities (Dagnan and Waring, 2004). Only one study gathering primary data addressed the role of humor in stigmatizing people with intellectual disabilities, with the majority of studies gathering secondary data or involving media related case studies. A large body of more discursive literature exists focusing on critical aspects of humor and disability (e.g., Coogan and Mallett, 2013). This identifies humor as disability activism serving entertainment, societal education and re-appropriating functions (e.g., Shain, 2013). However, to date, this literature has seldom focused on humor and people with intellectual disabilities, instead primarily focusing on disability where cognitive impairment is not present.

Humor was also explored in educational settings with a focus on its role as a facilitator of development and learning. However, laughter was considered an unwanted, disruptive or inappropriate behavior in some studies too, with a small number of investigations attempting to unpick the factors which elicit laughter. Further exploration of context and differing conceptualizations of humor are clearly needed. Humor was rarely studied as a component of creativity amongst people with intellectual disabilities and autism with only one study investigating it in this way. Others highlighted the creativity inherent in the humorous expression of people with intellectual disabilities and that creativity may differ between children with and without autism. However, creativity was not always operationalized adequately within these studies.

Humor and play literature focused only on children with autism and Down syndrome and revealed that despite evidence of deficits in the social-cognitive aspects of humor, with some reductions in scope and complexity of expression, young people did demonstrate humor in their play. Although hinted at in some of the investigations of the social communication (i.e., humorous banter), there is a need for further exploration of play in adulthood in people with intellectual disabilities given its positive association with wellbeing (Proyer, 2013).

For those providing support, humor served a bonding function between carers sharing similar challenging circumstances and facilitated coping. Observed expressions of humor and joy in people with ID and shared humor between carers and those supported enabled carers to maintain a sense of satisfaction, worth and joy in their caring role, despite the difficult times they may experience.

Evaluation of the Reviewed Literature

Currently, there exists limited literature focusing on humor in the lives of people with intellectual disabilities. In the literature that does exist a range of methods have been employed. In the main, studies adopted descriptive, survey, qualitative observation or interview based methods, with a number of quasi-experimental ex post facto design investigations and very few true experiments. The quality of the reviewed papers was, in the main good, with a few exceptions, in particular the qualitative research reviewed was well conducted. Nevertheless, there were few studies providing direct empirical investigation of humor appreciation and comprehension of people with intellectual disabilities. Some studies, especially those focusing on specific syndromes, were small scale, underpowered and lacked statistical analysis, however this is understandable given the rarity of these conditions.

Within the papers included in the review, humor was often incorporated, not as a primary variable, but instead as a descriptive secondary variable or illustrative of a wider field of study (i.e., social interaction/communication) or emerged as a finding not initially sought in the study. Seldom was humor the primary variable under investigation ($N = 6$). There may be a number of reasons for this. The first relates directly to the issue of the ubiquitous nature of humor within social exchange. This common oversight is well evidenced in the humor literature (Martin, 2010). Coupled with this the difficulties recruiting and designing studies to include people with intellectual disabilities and the social and research disenfranchisement of people with intellectual disabilities may also contribute to the current lack of literature. Where humor did emerge as an important variable, it was primarily highlighted for its facilitative nature in supporting relationships, development and psychological wellbeing and because it was illustrative of positive social interactions.

Limitations and Future Directions for Research

Given the positive and negative impacts on wellbeing, the ubiquitousness of humor as part of the human experience and the varied conceptualizations of humor evident, there does appear to be a need for more research specifically focusing on humor and intellectual disabilities. More high-quality, primary, empirical research appears to be needed. In particular, future studies are needed in the areas of humor comprehension, representation and stigma, with greater clarity and specificity needed around the meaning and measurement of humor under scrutiny. Moreover, no study directly explored the relationship between humor and wellbeing in people with intellectual disabilities, which is a notable oversight and needs addressing in future research endeavor. Due to the potential negative effects on psychological wellbeing, the role of humor as a manifestation of societal stigma is also in need of further robust empirical investigation.

Although the search terms for this study were representative of and aligned with the review aims, other search terms may have been overlooked. This may have yielded relevant literature omitted from this review. Definitional difference in nomenclature (i.e., the term learning disabilities equating to intellectual disabilities in the UK whilst in the US and Canada it more typically equated to specific learning difficulties and developmental disabilities) made identification of papers where the participant group was people with intellectual disabilities

more challenging. Alongside this, some papers did not adequately describe or define the participants which may have led to the inclusion of some papers which may not have been as directly relevant to people with intellectual disabilities (e.g., Bees, 1998; Muscott and O'Brien, 1999).

Finally, due to the novelty of the area of investigation the review presented is, by necessity, broad and multidisciplinary in scope in terms of the range of people with intellectual disabilities included. It does not focus on one specific group of people with intellectual disabilities with a range of methodologies employed in the selected studies. We aimed to explore the current state of knowledge in this field and so people with intellectual disabilities from different age groups, and their carers, were all included to provide more comprehensive and valuable insights into this unexplored area. Hence, we did not feel it appropriate to incorporate more specificity into inclusion/exclusion criteria for this initial review. Despite this, we would urge future empirical research and reviews to specify the distinct stakeholder and age groups and the particular etiology of participants. This will enable a corpus of research to be developed which can be synthesized in future meta-analysis and qualitative synthesis research. Moreover, many of the themes identified had only a handful of papers investigating them so the themes identified in this review are tentative. Further work is needed to bolster the existing evidence base and to fully explore many of the areas identified in this review. In particular the themes when developed from the review did not conform to a humor production / appreciation thematic structure as might be expected. Future research should prioritize work to better understand humor appreciation and production in people with intellectual disabilities to help achieve research parity and, more importantly, to enable more efficacious and positive support to occur through dissemination of this research work to key stakeholders and support staff. Finally, research endeavor should also be mindful to conduct humor research which is of importance to people with intellectual disabilities themselves via more inclusive and participatory strategies integrated into the research endeavor so that the work does not remain remote from the lives of people.

CONCLUSIONS

Humor is an important aspect of the social interactional lives of people with intellectual disabilities and their carers serving important social, developmental, and emotional wellbeing functions. In particular it can serve an equalizing function in terms of interactional power fostering the experience of shared moments and building of social capital. On the other hand, humor can also be a manifestation of negative attitudes and derogation of people with intellectual disabilities, serving as a source of source of stigma and emotional harm. However, the literature as it stands is limited with the need for further methodologically robust investigations where humor is a central variable of interest. Such work will enable the ways in which humor serves both positive and negative functions in people's lives to be better understood, fostered and combatted.

AUTHOR CONTRIBUTIONS

TP and DC contributed to the conceptualization of the review. DC was first screener of the papers focusing on people with Rett and Angelman syndrome and papers relating to carers and stigma. TP reviewed the humor in play, creativity and classroom papers, and those papers focusing on people with Down syndrome and autism. Authors shared preliminary reviewing of the core humor appreciation, comprehension and social facilitation papers. Both authors independently completed the quality reviews on all selected papers. TP and DC contributed equally to the writing of the study.

ACKNOWLEDGMENTS

We would like to thank Dr. Wendy Nicholls for her advice on the quality assessment processes of conducting a systematic review.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01745/full#supplementary-material

REFERENCES

Abbott, S., and McConkey, R. (2006). The barriers to social inclusion as perceived by people with intellectual disabilities. *J. Intellect. Disabil.* 10, 275–287. doi: 10.1177/1744629506067618

Adams, D., Horsler, K., Mount, R., and Oliver, C. (2015). Brief report: a longitudinal study of excessive smiling and laughing in children with Angelman syndrome. *J. Autism Dev. Disord.* 45, 2624–2627. doi: 10.1007/s10803-015-2404-y

Ali, A., King, M., Strydom, A., and Hassiotis, A. (2016). Self-reported stigma and its association with socio-demographic factors and physical disability in people with intellectual disabilities: results from a cross-sectional study in England. *Soc. Psychiatry Psychiatr. Epidemiol.* 51, 465–474. doi: 10.1007/s00127-015-1133-z

Amado, A. N., Stancliffe, R. J., McCarron, M., and McCallion, P. (2013). Social inclusion and community participation of individuals with intellectual/developmental disabilities. *Intellect. Dev. Disabil.* 51, 360–375. doi: 10.1352/1934-9556-51.5.360

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* 23, 185–188. doi: 10.1038/13810

Baron, R. A., Fortin, S. P., Frei, R. L., Hauver, L. A., and Shack, M. L. (1990). Reducing organizational conflict: the role of socially-induced positive affect. *Int. J. Conf. Manag.* 1, 133–152. doi: 10.1108/eb022677

Bees, C. (1998). The GOLD Program: a program for gifted learning disabled adolescents. *Roeper Rev.* 21, 155–161. doi: 10.1080/02783199809553951

Billig, M. (2005). *Laughter and Ridicule: Towards a Social Critique of Humour*. London: Sage Publications.

Bower, B. D., and Jeavons, P. M. (1967). The" happy puppet" syndrome. *Arch. Dis. Child.* 42, 298–302. doi: 10.1136/adc.42.223.298

Brown, P., and Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage,* Vol. 4. Cambridge: Cambridge University Press.

Chadwick, D. D., and Fullwood, C. (2017). An online life like any other: Identity, self-determination, and social networking among adults with intellectual disabilities. *Cyberpsychol. Behav. Soc. Netw.* 21, 56–64 doi: 10.1089/cyber.2016.0689

Chapman, A. J. (1983). "Humor and laughter in social interaction and some implications for humor research," in *Handbook of Humor Research Volume 1: Basic Issues,* eds P. E. McGhee and J. H. Goldstein (New York, NY: Springer-Verlag), 135–157.

Chappell, A. L., Goodley, D., and Lawthom, R. (2001). Making connections: the relevance of the social model of disability for people with learning difficulties. *Br. J. Learn. Disabil.* 29, 45–50. doi: 10.1046/j.1468-3156.2001.00084.x

Coogan, T., and Mallett, R. (2013). Introduction: disability, humour and comedy. *J. Lit. Cult. Disabil. Stud.* 7, 247–253. doi: 10.3828/jlcds.2013.22

Dagnan, D., and Waring, M. (2004). Linking stigma to psychological distress: testing a social–cognitive model of the experience of people with intellectual disabilities. *Clin. Psychol. Psychother.* 11, 247–254. doi: 10.1002/cpp.413

Degabriele, J., and Walsh, I. P. (2010). Humour appreciation and comprehension in children with intellectual disability. *J. Intellect. Disabil. Res.* 54, 525–537. doi: 10.1111/j.1365-2788.2010.01277.x

Forster, S., and Iacono, T. (2008). Disability support workers' experience of interaction with a person with profound intellectual disability. *J. Intellect. Dev. Disabil.* 33, 137–147. doi: 10.1080/13668250802094216

Forster, S., and Iacono, T. (2014). The nature of affect attunement used by disability support workers interacting with adults with profound intellectual and multiple disabilities. *J. Intellect. Disabil. Res.* 58, 1105–1120. doi: 10.1111/jir.12103

Fredrickson, B. L. (2000). Why positive emotions matter in organizations: lessons from the broaden-and-build model. *Psychol. Manag. J.* 4:131. doi: 10.1037/h0095887

Fudge Schormans, A., Renwick, R., Barker, D., Chasi, E., Smith, B., McWilliam, L., et al. (2013). Why *can't* we be superheroes? Researchers' with and without intellectual and developmental disabilities thoughts on *Defendor*. *Media Rev. J. Dev. Disabil.* 19, 109–112.

Gagić, S., JapundŽa-Milisavljević, M., and Durić-,Zdravković, A. (2015). Examples from visual surroundings as an incentive for children with mild intellectual disability to express their creativity in the art domain. *Hrvatski Casopis za Odgoj i Obrazovanje,* 17, 41–64. doi: 10.15516/cje.v17i0.1073

Godbee, K., and Porter, M. (2013). Comprehension of sarcasm, metaphor and simile in Williams syndrome. *Int. J. Lang. Commun. Disord.* 48, 651–665. doi: 10.1111/1460-6984.12037

Goggins, G. (2010). "Laughing with/at the disabled": the cultural politics of disability in Australian universities. *Discourse Stud. Cult. Polit. Educ.* 31, 469–481. doi: 10.1080/01596306.2010.504363

Griffiths, C., and Smith, M. (2016). Attuning: a communication process between people with severe and profound intellectual disability and their interaction partners. *J. Appl. Res. Intellect. Disabil.* 29, 124–138. doi: 10.1111/jar.12162

Griffiths, C., and Smith, M. (2017). You and me: the structural basis for the interaction of people with severe and profound intellectual disability and others. *J. Intellect. Disabil.* 21, 103–117. doi: 10.1177/1744629516644380

Hagberg, B. (2005). Rett syndrome: long-term clinical follow-up experiences over four decades. *J. Child Neurol.* 20, 722–727. doi: 10.1177/08830738050200090401

Hall, G. S., and Allin, A. (1897). The psychology of tickling, laughing, and the comic. *Am. J. Psychol.* 9, 1–41. doi: 10.2307/1411471

Hobson, R. P., Lee, A., and Hobson, J. A. (2009). Qualities of symbolic play among children with autism: a social-developmental perspective. *J. Autism Dev. Disord.* 39, 12–22. doi: 10.1007/s10803-008-0589-z

Holmes, J. (2006). Sharing a laugh: pragmatic aspects of humor and gender in the workplace. *J. Pragmat.* 38, 26–50. doi: 10.1016/j.pragma.2005.06.007

Hudenko, W. J., Stone, W., and Bachorowski, J. A. (2009). Laughter differs in children with autism: an acoustic analysis of laughs produced by children with and without the disorder. *J. Autism Dev. Disord.* 39, 1392–1400. doi: 10.1007/s10803-009-0752-1

Johanson-Sebera, B., and Wilkins, J. (2014). The uses and implications of the term "retarded" on YouTube. *Rev. Disabil. Stud. Int. J.* 6, 47–61.

Johnson, H., Douglas, J., Bigby, C., and Iacono, T. (2012). Social interaction with adults with severe intellectual disability: having fun and hanging out. *J. Appl. Res. Intellect. Disabil.* 25, 329–341. doi: 10.1111/j.1468-3148.2011.00669.x

Jones, M. M., and Goble, Z. (2012). Creating effective mentoring partnerships for students with intellectual disabilities on campus. *J. Policy Pract. Intellect. Disabil.* 9, 270–278. doi: 10.1111/jppi.12010

Kaufmann, W. E., Glaze, D. G., Christodoulou, J., Clarke, A. J., Bahi-Buisson, N., Leonard, H., et al. (2010). Rett syndrome: revised diagnostic criteria and nomenclature. *Ann. Neurol.* 68, 944–950. doi: 10.1002/ana.22124

Keltner, D., Capps, L., Kring, A. M., Young, R. C., and Heerey, E. A. (2001). Just teasing: a conceptual analysis and empirical review. *Psychol. Bull.* 127:229–248. doi: 10.1037/0033-2909.127.2.229

Khan, K. S., Kunz, R., Kleijnen, J., and Antes, G. (2003). Five steps to conducting a systematic review. *J. R. Soc. Med.* 96, 118–121. doi: 10.1177/014107680309600304

Kmet, L. M., Lee, R. C., and Cook, L. S. (2004). *Standard Quality Assessment Criteria for Evaluating Primary Research Papers From A Variety of Fields (Vol. 22).* Edmonton, AB: Alberta Heritage Foundation for Medical Research.

Krishan, S., Batchelor, J., and Porter, M. (2017). Humour comprehension and use of mental state language in William Syndrome and Down Syndrome. *Global J. Intellect. Dev. Disabil.* 2:555594. doi: 10.19080/GJIDD.2017.02.555594

Kuiper, N. A., Martin, R. A., and Dance, K. A. (1992). Sense of humour and enhanced quality of life. *Pers. Individ. Diff.* 13, 1273–1283. doi: 10.1016/0191-8869(92)90169-P

MacDonald, E., Fitzsimons, E., and Walsh, P. N. (2007). Use of respite care and coping strategies among Irish families of children with intellectual disabilities. *Br. J. Learn. Disabil.* 35, 62–68. doi: 10.1111/j.1468-3156.2006.00399.x

MacTavish, J. B., MacKay, K. J., Iwasaki, Y., and Betteridge, D. (2007). Family caregivers of individuals with intellectual disability: perspectives on life quality and the role of vacations. *J. Leis. Res.* 39, 127–155. doi: 10.1080/00222216.2007.11950101

Manion, M. L., and Bersani, H. A. (1987). Mental retardation as a western sociological construct: a cross-cultural analysis. *Disabil. Handicap Soc.* 2, 231–245. doi: 10.1080/02674648766780301

Martin, R. A. (2010). *The Psychology of Humor: An Integrative Approach*. London: Elsevier.

McDougall, A., Kerr, A. M., and Espie, C. A. (2005). Sleep disturbance in children with Rett syndrome: a qualitative investigation of the parental experience. *J. Appl. Res. Intellect. Disabil.* 18, 201–215. doi: 10.1111/j.1468-3148.2005.00220.x

McGhee, P. E. (1979). *Humor, its Origin and Development*. San Francisco, CA: W.H. Freeman.

McGhee, P. E. (1980). *Development of the Creative Aspects of Humor. Children's Humour*. Chichester: John Wiley and Sons.

Messier, J., Ferland, F., and Majnemer, A. (2008). Play behavior of school age children with intellectual disability: their capacities, interests and attitude. *J. Dev. Phys. Disabil.* 20, 193–207. doi: 10.1007/s10882-007-9089-x

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 6:e1000097. doi: 10.1371/journal.pmed.1000097

Moran, C. (2003). Beyond content: does humor help coping? *Disabil. Stud. Q.* 23.

Moran, C. C. (2013). "Humor as a moderator of compassion fatigue," in *Treating Compassion Fatigue*, ed C. R. Figley (New York, NY: Routledge), 139–154.

Mount, R., Oliver, C., Berg, K., and Horsler, K. (2011). Effects of adult familiarity on social behaviours in Angelman syndrome. *J. Intellect. Disabil. Res.* 55, 339–350. doi: 10.1111/j.1365-2788.2010.01364.x

Muscott, H. S., and O'Brien, S. T. (1999). Teaching character education to students with behavioral and learning disabilities through mentoring relationships. *Educ. Treat. Child.* 22, 373–390.

Nirenberg, S. A. (1991). Normal and pathologic laughter in children. *Clin. Pediatr.* 30, 630–632. doi: 10.1177/000992289103001103

Nomura, Y. (2005). Early behavior characteristics and sleep disturbance in Rett syndrome. *Brain Dev.* 27, S35–S42. doi: 10.1016/j.braindev.2005.03.017

Oliver, C., Demetriades, L., and Hall, S. (2002). Effects of environmental events on smiling and laughing behavior in Angelman syndrome. *Am. J. Ment. Retard.* 107, 194–200. doi: 10.1352/0895-8017(2002)107<0194:EOEEOS>2.0.CO;2

Pressman, S. D., and Cohen, S. (2005). Does positive affect influence health? *Psychol. Bull.* 131, 925–971. doi: 10.1037/0033-2909.131.6.925

Proyer, R. T. (2013). The well-being of playful adults: Adult playfulness, subjective well-being, physical well-being, and the pursuit of enjoyable activities. *Eur. J. Humour Res.* 1, 84–98. doi: 10.7592/EJHR2013.1.1.proyer

Reddy, V., Williams, E., and Vaughan, A. (2002). Sharing humour and laughter in autism and Down's syndrome. *Br. J. Psychol.* 93, 219–242. doi: 10.1348/000712602162553

Ruch, W. (1992). Assessment of appreciation of humor: studies with the 3 WD humor test. *Adv. Pers. Assess.* 9, 27–75.

Ruch, W. (1995). Will the real relationship between facial expression and affective experience please stand up: the case of exhilaration. *Cogn. Emot.* 9, 33–58. doi: 10.1080/02699939508408964

Ruch, W. (1997). State and trait cheerfulness and the induction of exhilaration. *Eur. Psychol.* 2, 328–341. doi: 10.1027/1016-9040.2.4.328

Ruch, W., and Heintz, S. (2016). The virtue gap in humor: exploring benevolent and corrective humor. *Transl. Issues Psychol. Sci.* 2, 35–45. doi: 10.1037/tps0000063

Ruch, W., Heintz, S., Platt, T., Wagner, L., and Proyer, R. T. (2017). *Broadening humor: comic styles differentially tap into temperament*, character, and ability. *Front. Psychol.* 9:6. doi: 10.3389/fpsyg.2018.00006

Ruch, W., Köhler, G., and Van Thriel, C. (1996). Assessing the "humorous temperament": construction of the facet and standard trait forms of the State-Trait-Cheerfulness-Inventory—STCI. *Humor Int. J. Humor Res.* 9, 303–340.

Sandhu, D. K., Rose, J., Rostill-Brookes, H. J., and Thrift, S. (2012). 'It's intense, to an extent': a qualitative study of the emotional challenges faced by staff working on a treatment programme for intellectually disabled sex offenders. *J. Appl. Res. Intellect. Disabil.* 25, 308–318. doi: 10.1111/j.1468-3148.2011.00667.x

Schieltz, K. M., Wacker, D. P., Harding, J. W., Berg, W. K., Lee, J. F., Dalmau, Y. C. P., et al. (2011). Indirect effects of functional communication training on non-targeted disruptive behavior. *J. Behav. Educ.* 20, 15–32. doi: 10.1007/s10864-011-9119-8

Schmidt-Hidding, W. (1963). *Europaische Schlusselworter. Band 1, Humor und Witz. (Humour and Wit)*. Munich: Hueber.

Schnitzer, G., Andries, C., and Lebeer, J. (2007). Usefulness of cognitive intervention programmes for socio-emotional and behaviour problems in children with learning disabilities. *J. Res. Spec. Educ. Needs* 7, 161–171. doi: 10.1111/j.1471-3802.2007.00093.x

Shain, A. (2013). Comment from the field: perspectives on comedy and performance as radical disability activism. *J. Lit. Cult. Disabil. Stud.* 7, 337–346. doi: 10.3828/jlcds.2013.28

Short, E. J., Basili, L. A., and Schatschneider, C. W. (1993). Analysis of humor skills among elementary school students: comparisons of children with and without intellectual handicaps. *Am. J. Ment. Retard.* 98, 63–73.

Siperstein, G. N., Pociask, S. E., and Collins, M. A. (2010). Sticks, stones, and stigma: a study of students' use of the derogatory term "retard". *Intellect. Dev. Disabil.* 48, 126–134. doi: 10.1352/1934-9556-48.2.126

Steffenburg, S., Gillberg, C. L., Steffenburg, U., and Kyllerman, M. (1996). Autism in Angelman syndrome: a population based study. *Pediatr. Neurol.* 14, 131–136. doi: 10.1016/0887-8994(96)00011-2

St. James, P. J. S., and Tager-Flusberg, H. (1994). An observational study of humor in autism and Down syndrome. *J. Autism Dev. Disord.* 24, 603–617. doi: 10.1007/BF02172141

Sullivan, K., Winner, E., and Tager-Flusberg, H. (2003). Can adolescents with Williams syndrome tell the difference between lies and jokes? *Dev. Neuropsychol.* 23, 85–103. doi: 10.1080/87565641.2003.9651888

Van Loon, J., Claes, C., Vandevelde, S., Van Hove, G., and Schalock, R. L. (2010). Assessing individual support needs to enhance personal outcomes. *Exceptionality* 18, 193–202. doi: 10.1080/09362835.2010.513924

Wong, K., Leonard, H., Jacoby, P., Ellaway, C., and Downs, J. (2015). The trajectories of sleep disturbances in Rett syndrome. *J. Sleep Res.* 24, 223–233. doi: 10.1111/jsr.12240

World Health Organisation (1992). *International Classification of Diseases. 10th Revision*. Geneva: WHO.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Chadwick and Platt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 April 2018
doi: 10.3389/fpsyg.2018.00577

Feasibility of a Humor Training to Promote Humor and Decrease Stress in a Subclinical Sample: A Single-Arm Pilot Study

Nektaria Tagalidou[1], Viola Loderer[1], Eva Distlberger[1] and Anton-Rupert Laireiter[1,2]*

[1] Department of Psychology, University of Salzburg, Salzburg, Austria, [2] Faculty of Psychology, University of Vienna, Vienna, Austria

The present study investigates the feasibility of a humor training for a subclinical sample suffering from increased stress, depressiveness, or anxiety. Based on diagnostic interviews, 35 people were invited to participate in a 7-week humor training. Evaluation measures were filled in prior training, after training, and at a 1-month follow-up including humor related outcomes (coping humor and cheerfulness) and mental health-related outcomes (perceived stress, depressiveness, anxiety, and well-being). Outcomes were analyzed using repeated-measures ANOVAs. Within-group comparisons of intention-to-treat analysis showed main effects of time with large effect sizes on all outcomes. *Post hoc* tests showed medium to large effect sizes on all outcomes from pre to post and results remained stable until follow-up. Satisfaction with the training was high, attrition rate low (17.1%), and participants would highly recommend the training. Summarizing the results, the pilot study showed promising effects for people suffering from subclinical symptoms. All outcomes were positively influenced and showed stability over time. Humor trainings could be integrated more into mental health care as an innovative program to reduce stress whilst promoting also positive emotions. However, as this study was a single-arm pilot study, further research (including also randomized controlled trials) is still needed to evaluate the effects more profoundly.

Keywords: humor training, subclinical, coping humor, cheerfulness, perceived stress, single-arm

OPEN ACCESS

Edited by:
René T. Proyer, Martin Luther University Halle-Wittenberg, Germany

Reviewed by:
Jennifer Hofmann, Universität Zürich, Switzerland
Ofra Nevo, University of Haifa, Israel

****Correspondence:***
Nektaria Tagalidou
nektaria.tagalidou@sbg.ac.at

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *14 December 2017*
Accepted: *05 April 2018*
Published: *24 April 2018*

Citation:
Tagalidou N, Loderer V, Distlberger E and Laireiter A-R (2018) Feasibility of a Humor Training to Promote Humor and Decrease Stress in a Subclinical Sample: A Single-Arm Pilot Study. Front. Psychol. 9:577. doi: 10.3389/fpsyg.2018.00577

INTRODUCTION

Increased levels of stress are highly prevalent (Wiegner et al., 2015) and entail serious physical and mental health problems. Prolonged stress increases the risk of acute myocardial infarction (Rosengren et al., 2004), weakens the immune system (Segerstrom and Miller, 2004), and is related to depression, exhaustion (Wiegner et al., 2015), and reduced quality of life (Golden-Kreutz et al., 2005).

Adaptive appraisal and coping can be good ways to handle stress more effectively and so diminish the negative impact of it on health outcomes (Lazarus and Folkman, 1984). However, not everybody possesses the ability to cope with stress adequately and suffers from its consequences. Due to that, stress prevention and stress reduction programs receive growing attention, as they teach the use of adaptive coping mechanisms and therefore help handle prolonged stress (Jaremko and Meichenbaum, 2013).

A new and promising strategy to handle and reduce stress, whilst furthermore promoting also mental health and well-being, is the use of humor. Humor has already been recognized as an effective stress moderator (Martin, 2011): Research shows that the use of humor is an adaptive emotion regulation strategy in the short-term (Strick et al., 2009; Samson and Gross, 2012; Kugler and Kuhbandner, 2015) and also serves as a coping strategy against negative and stressful life situations in the longer-term (Martin and Lefcourt, 1983; Labott and Martin, 1987; Overholser, 1992; Sliter et al., 2014). Furthermore, using humor does not only downregulate negative emotions but elicits also positive emotions, such as amusement (Herring et al., 2011), which are important for promoting resilience and well-being as stated by the broaden-and-built theory (Fredrickson and Joiner, 2002; Fredrickson, 2004).

Due to its basic working mechanisms (downregulation of negative emotions and upregulation of positive emotions) humor is positively related to life satisfaction, positive affect, and well-being (Martin et al., 2003; Martinez-Marti and Ruch, 2014), and has positive effects in various aspects of life (Martin, 2011).

To profit from the diverse positive effects of humor, latest research has focused on improving humor through various interventions, especially humor trainings. Humor trainings can differ in their structure and grounded theory; however, they all aim at the same targets: to promote positive emotions, longer-lasting positive mood states such as cheerfulness (Ruch et al., 1996), and most importantly coping humor, which is defined as the ability to use humor to cope with stress (Ruch and Hofmann, 2017). A widely known humor training program was developed by McGhee (1996, 2010) called the "7 Humor Habits program." It has already proven its efficacy in increasing/decreasing several mental health outcomes like positive affect, life satisfaction, depression, and anxiety (Sassenrath, 2001, unpublished; Beh-Pajooh et al., 2010; Crawford and Caltabiano, 2011; Ruch et al., 2018). Also, most importantly, perceived stress and stress levels can indeed be reduced by the participation at the training (Crawford and Caltabiano, 2011).

It is important to note that the studies just mentioned used only healthy participants in their designs. Research on humor trainings in clinical settings is scarce, although people suffering from mental disorders would profit from them. Mental disorders can entail various humor- and stress-related deficits. Regarding humor, difficulties in cognitive and affective components of humor or difficulties experiencing cheerfulness have been reported (Uekermann et al., 2007; Uekermann et al., 2008; Falkenberg et al., 2011). Regarding stress, people with mental disorders show serious deficits in emotion regulation and coping processes (Aldao et al., 2010). The training could help affected people improve their sense of humor, so that they are able to use it for coping with stress and negative affectivity in daily life. Some studies already tried to investigate the effects of humor trainings in clinical populations and found promising results. Falkenberg et al. (2010), for example, could demonstrate an increase in coping humor for depressed inpatients and Cai et al. (2014) found improvements in symptomatology, depression, and sense of humor for schizophrenics. Tagalidou et al. (in press) found improvements in coping humor and cheerfulness for people of a routine care institution suffering from schizophrenia, personality disorders, anxiety, or depression. All studies used McGhee's humor training. Further studies, using an alternative humor training program for depressed elderly inpatients, could also show positive results like changes in resilience, cheerfulness, or satisfaction with life (Hirsch et al., 2010; Konradt et al., 2013).

As can be seen, research on manualized humor trainings is relatively new. There have been studies conducted with healthy or clinical samples which show promising results. However, studies including participants with stress-related and subclinical, yet burdening, symptoms like increased levels of stress, depressiveness, or anxiety have not yet been published at all, although it would be reasonable to concentrate on this population, too. Subclinical problems can easily grow up to clinical symptoms which have to be treated with psychotherapy or psychotropic drugs, burdening affected people and the health care system (Vigo et al., 2016). A low-threshold, preventive offer like a short humor training could help decrease subclinical symptoms and stress and furthermore promote cheerfulness. By integrating both these aspects at an early stage of symptom development, ideally it would be possible to diminish incidence rates of mental disorders.

Aims and Research Questions

The study investigates feasibility of a humor training for people with subclinical symptoms like increased levels of stress, depressiveness, or exhaustion and tries to narrow the gap opened by lack of research in this area. The main focus is the evaluation of the training as a low-threshold, preventive program against everyday life stress and hassles.

Although different stress preventive programs have already been developed to reduce stress by now, most of them concentrate mainly on the reduction of stress-related symptoms. The promotion of positive aspects like well-being, resilience, and personal strengths remains rather neglected. If the training appears to be as feasible as already broadly implemented programs [see e.g., "Mindfulness-Based Stress Reduction" (MBSR); Grossman et al., 2004; Khoury et al., 2015], it is conceivable to integrate humor trainings in future health-care systems as additional and alternative prevention programs against stress and mental disorders. Interventions like this one could help decrease the incidence rates of mental symptoms as they intervene already at early stages of symptom development and thereby also promote cheerfulness and well-being.

The study focuses on three main objectives: the first one is to evaluate if the humor training can improve humor related outcomes. The training mainly promotes using humor under stress (coping humor), therefore coping humor was chosen as a humor-related outcome. Furthermore, cheerfulness was assessed as a longer-lasting positive mood state.

The second aim is to evaluate if the training can improve mental health and well-being. To test this hypothesis, several mental health-related outcomes were included in the study's design. As the training primarily tries to decrease stress, perceived stress was included as an outcome variable. Further, depressiveness and anxiety were included to test if the training can improve subclinical forms of depressiveness and anxiety.

Lastly, well-being was included as a positive outcome of the training.

The third aim of the study is to evaluate applicability of the training based on the feedback of participants. Evaluation of feedback in humor training studies is scarce, so we want to emphasize this aspect more to get a broader overview about the feasibility of the training.

MATERIALS AND METHODS

Design

The design of the study was a single-arm trial to explore feasibility of the humor training for a subclinical population currently experiencing increased stress, depressiveness, or exhaustion. The within-factors design had three measurement time points: 1 week before treatment, 1 week after treatment, and a 1-month follow-up after treatment.

The training took place in the outpatient clinic of the University of Salzburg. The study's protocol was approved by the ethics commission of the University of Salzburg (44/2016) and registered in the German Clinical Trials Register (DRKS00013480).

Participants

Calculated by G*Power 3.1 (Faul et al., 2007), the required sample size to find a medium effect of $f = 0.25$ with a power of $\beta = 0.80$ and an α level of 0.05 was 28 (a medium effect was assumed for calculation based on reviewing already existent research on humor trainings). However, to also cover potential dropouts, a higher N than 28 is needed. McDermut et al. (2001) reported an average attrition rate of 18.6% for group therapies of mood disorders. Considering their result, a total N of 33 was assumed to be needed. In the end, the final number of participants was 35. They were recruited via an advertising article in the local newspaper of Salzburg ("Salzburger Nachrichten"), which contained a report about the planned humor training. The article addressed people who currently experienced stressful situations, depressiveness, or exhaustion in their daily lives and it was explained how humor and the experience of cheerfulness can help in coping better with negative situations and emotions. Everyone who was interested in humor and experienced stress and hassles in daily life was invited to participate in the study. Participants received the humor training for free; in return they were asked to fill in evaluation questionnaires. It was not explained which outcomes are being measured and analyzed. Participants were only told that the humor training helps coping better with daily life stress.

The inclusion criterion for participating in the study was the subclinical experience of symptoms like increased stress, exhaustion, depressiveness, or anxiety. Thus, if someone showed clinical symptoms and fulfilled the criteria for any current mental disorder, he or she was excluded from the study. Only one exception was made for people with a recurrent depressive disorder currently in remission (ICD 10: F33.4, DSM-IV: 296.36). These people do not show any symptoms of a current depressive episode; however, they generally experience subclinical depressive symptoms frequently and have high risk of recurrences (Solomon et al., 2000). So, to help them build up preventive strategies against forthcoming relapses, they were invited to participate in the study. Further inclusion criteria for the study have been good German language skills and no cognitive deficits like dementia.

A total of 111 people were interested in the training and submitted registration. Of these, 105 (94.6%) could be contacted for the telephonic pre-screening. Seventy-six (68.5%) participated in the face-to-face diagnostic interview and finally 35 (31.5%) persons met inclusion criteria and started training. **Figure 1** depicts the complete selection process.

Procedure

If interested, people registered online on the training's homepage and submitted contact details so they could be contacted telephonically. During the phone call, general organizational information was communicated and a quick pre-screening, concerning interest and motivation for participation, was conducted. If the interested persons appeared to suit in the study's design, they were invited to take part in a face-to-face diagnostic interview. The interview was conducted using the Structured Clinical Interview for DSM-IV, I, and II (Wittchen et al., 1997). Only people who showed no current mental disorder or a recurrent depressive disorder currently in remission were allowed to participate in the training. The interviews were conducted by employees of the outpatient clinic who have been in training

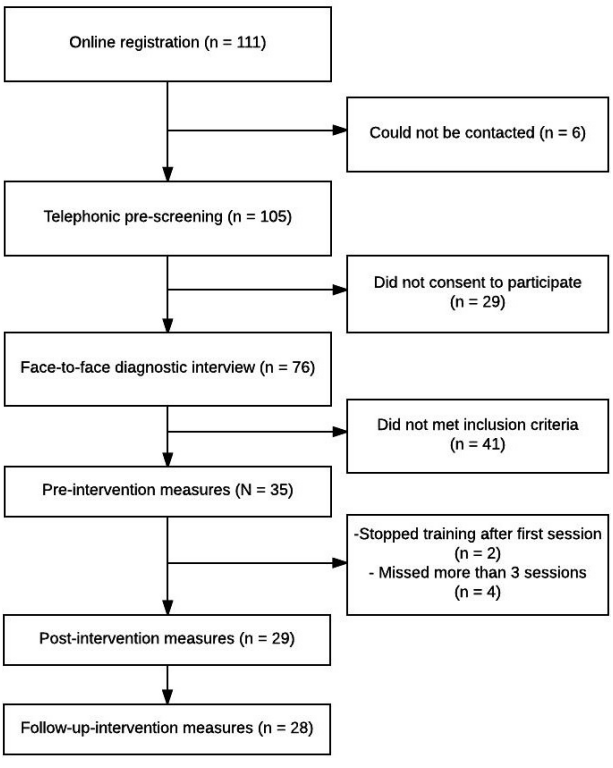


FIGURE 1 | Flowchart of the study.

as clinical psychologists. They got a regular training and had sufficient experience with diagnostic interviews in general, and the SCID I and II manual in specific.

Finally, people who fulfilled inclusion criteria were invited to participate in the training. They were assigned to a training group and written informed consent and a non-disclosure agreement were signed.

There were four humor training groups which started consecutively. Each group was led by two group leaders. In total, six group leaders conducted trainings. They were employees of the outpatient clinic and in training as clinical psychologists or trained master's students at the end of their studies. They all had clinical experience and were extensively introduced to and trained in the humor training program. Assignment of trainers to the groups was random.

Humor Training

The humor training is based on the German manual of Falkenberg et al. (2013). It is a 7-week program to promote cheerfulness and humor in everyday life and based on McGhee's (1996, 2010) "7 Humor Habits Program". Special attention is paid to the improvement of coping humor abilities, so that participants can use humor as a protective factor against personal stressful situations. The manual of Falkenberg et al. (2013) was developed specifically for people with mental disorders. However, we still used it for our subclinical population as the contents can easily be transferred also to people not suffering from a mental disorder. Furthermore, we slightly modified the manual with our own ideas so that it was more suitable for our sample. The training contained psychoeducational elements, which were combined with various exercises like role plays, games, and discussions. Every session addressed one specific humor topic like finding humor in everyday life, promoting playfulness, and finding a benevolent attitude toward personal weaknesses. Every session lasted 90 min. Additionally, participants had to do homework to implement the learned better in everyday life. **Table 1** summarizes the seven sessions and their associated content.

To get a better overview of the sessions' structure, session 3, "laughter" will be explained in more detail. The session starts with an opening game to activate the participants and get them into positive mood. After that, homework is discussed and the last session briefly summarized. Beginning with the psychoeducational part of the session, the positive effects of laughter on physical and mental health and the concept of real (Duchenne) vs. fake (non-Duchenne) smiles are explained. People then participate in a quiz, where they have to detect Duchenne or non-Duchenne smiles on their own. After the general information about laughter and smiling, participants have an imagination exercise about laughter and group work, where they have to make their partner laugh with funny grimaces or jokes. In the end, homework is discussed and a funny closing game played.

TABLE 1 | Topics of the humor training.

Session	Content
1. Session	• Topic: Introduction • Meet and greet • Definitions of humor • The own sense of humor
2. Session	Topic: Seriousness vs. playfulness • The effects of seriousness on everyday life • The function of play and playful behavior
3. Session	Topic: Laughter • Physical and mental benefits of laughter • Duchenne vs. non-Duchenne laughter • Laughter exercises
4. Session	Topic: Creating verbal humor • Telling jokes • Ambiguousness • Exaggeration
5. Session	Topic: Finding humor in everyday life • Change of perspectives • Searching for humor consciously in daily situations
6. Session	Topic: To laugh about oneself • Humorous perspective on personal weaknesses • Pros/cons of laughing at oneself
7. Session	Topic: Finding humor in stressful situations • Definition of stress • Effects of humor on stress • Feedback

Measures

All outcomes were measured online using self-report questionnaires. Additionally, a feedback questionnaire was included after training, which could be filled in voluntarily and anonymously by participants.

Humor-Related Outcomes

Coping humor was measured using the *Coping Humor Scale (CHS)* by Martin and Lefcourt (1983). It is an economical 7-item scale which assesses the amount of humor someone uses to cope with stressors. The 4-point Likert scale ranges from 1 to 4. Internal consistency (Cronbach's alpha) was between $\alpha = 0.75$ and 0.80 for the three measurement time points.

Cheerfulness was measured using the *State-Trait-Cheerfulness Inventory (STCI)* – state version, which assesses short-term changes of exhilaration (Ruch et al., 1996, 1997). The questionnaire has three subscales: cheerfulness, seriousness, and bad mood with 10 items each and a 4-point Likert scale (1–4). Internal consistencies were between $\alpha = 0.90$ and 0.93, $\alpha = 0.62$ and 0.81, and $\alpha = 0.88$ and 0.93 for the three scales and measurement time points respectively.

As an additional humor-related outcome, which will be analyzed only descriptively, gelotophobia was assessed before treatment using the *Gelotophobia Questionnaire (GELOPH-15)* by Ruch and Proyer (2008a,b). Gelotophobia is defined as the fear of being laughed at by others (Ruch and Proyer, 2008a,b) and the questionnaire was included to get a more detailed picture about the characteristics of the sample. As the training contains numerous situations with laughter and cheerfulness and would stress people with gelotophobia, it is interesting to explore how many people with gelotophobic fears would in fact register for humor training. Ruch and Proyer (2008a,b) have defined three cut-off criteria for gelotophobia: A mean $\geq$2.50 indicates a slight

degree, a mean ≥3.00 a marked degree, and a mean ≥3.50 an extreme degree of gelotophobia. Fifteen items with a 4-point Likert scale (1–4) show an internal consistency of α = 0.87 at pre-treatment.

Mental Health-Related Outcomes

Perceived stress was assessed with the German version of the well-established *Perceived Stress Scale (PSS)* by Klein et al. (2016). Ten items with a 5-point Likert scale (0–4) show an internal consistency between α = 0.76 and 0.81 for the three measurement time points.

To evaluate the changes in depressive symptoms, the German *Center for Epidemiological Studies Depression Scale Revised (CESD-R)* was used. It was developed by Hautzinger et al. (2012) and includes 15 items with a 4-point Likert scale (0–3). Internal consistency was between α = 0.74 and 0.87 for the three measurement time points.

Anxiety was measured using the German translation of the *State-Trait-Anxiety Inventory (STAI)* in the state version (Laux et al., 1981). It consists of 20 items with a 4-point Likert scale (1–4) and its internal consistency was between α = 0.89 and 0.94 for the three measurement time points.

The German version of the economic *WHO-5 Well-Being Index (WHO-5)* was used as a screening tool for subjectively perceived well-being (Brähler et al., 2007). The five items have a 6-point Likert scale (0–5) and an internal consistency between α = 0.73 and 0.79 for the three measurement time points.

Evaluation of Applicability

A feedback questionnaire with 14 quantitative items and three qualitative items was constructed by the authors to evaluate the general satisfaction and applicability of the training. The quantitative items range from 1 to 5 except the last question "Would you recommend the training?" which ranges from 1 to 4. All items were analyzed separately on the item level.

Statistical Analyses

The statistical software used was JASP 0.8.4 (JASP, 2017) and all analyses were calculated based on the intention-to-treat technique (ITT). Missing data at post and follow-up were imputed with the last observation carried forward method (LOCF). Outcomes were analyzed using repeated-measures ANOVAs. Time main effects and *post hoc* tests with Bonferroni correction (pre–post and pre–follow-up) were calculated. Furthermore, effect sizes of time main effects and *post hoc* tests were analyzed to evaluate the effects of the training more profoundly. Cohen's *d* was chosen as effect size for the time main effects which was converted from eta squared (η^2) based on the formula of Cohen (1988). Effect sizes of *post hoc* tests with 95% confidence interval are also reported in Cohen's *d* based on the formula of Gibbons et al. (1993). Cohen (1988) defines $d = 0.2$ as small effect, $d = 0.5$ as medium effect, and $d = 0.8$ as large effect. Feedback was analyzed descriptively for quantitative items and qualitatively for open-format items.

RESULTS

Sample Characteristics

Table 2 summarizes demographic characteristics of the sample. Generally, participants had a mean age of 51.9 years (*SD* = 9.67), were predominantly female (*n* = 26, 74.3%), and Austrian (*n* = 33, 94.3%). They were mainly well educated (*n* = 26, 74.3%) and employed (*n* = 22, 62.9%).

With regard to inclusion criteria, 28 persons (80.0%) reported subclinical symptoms without depressive episodes in the past and 7 people (20.0%) reported subclinical symptoms with a recurrent depressive disorder currently in remission. One person was in psychotherapy (2.9%) and six took psychotropic drugs (*n* = 6, 17.1%). While participating in the study, two people (5.7%) had changes in their medication. One person changed dose and one person discontinued medication. Gelotophobic fear has been small in the sample: only one person reported a slight and another person a marked degree of gelotophobia. Thirty-three people were below the cut-off score of 2.50.

At the end, dropout of the study was similar compared to the average attrition rate of 18.6% (McDermut et al., 2001). Two persons (5.7%) stopped training after the first session and 4 persons (11.4%) missed three or more sessions and were therefore classified as non-completers (total attrition rate: 17.1%).

TABLE 2 | Demographic characteristics of the sample (*N* = 35).

	***M* (*SD*) or *n* (%)**
Age, *M* (*SD*)	51.9 (9.67)
Gender, *n* (%)	
Female	26 (74.3%)
Male	9 (25.7%)
Nationality, *n* (%)	
Austrian	33 (94.3%)
German, Romanian, Swiss, Lithuanian (1 in each case)	4 (11.6%)
Education, *n* (%)	
≥9 years of education (compulsory school)	9 (25.7%)
≥12 years of education (A level)	11 (31.4%)
≥ any tertiary education (e.g., university)	15 (42.9%)
Employment	
Currently employed	22 (62.9%)
Retirement	6 (17.1%)
Parental leave/educational leave	3 (8.6%)
Partial retirement	2 (5.8%)
Student	1 (2.9%)
Not specified	1 (2.9%)
Subclinical symptoms	
Without *F* 33.4/ 296.36	28 (80.0%)
With *F* 33.47 296.36	7 (20.0%)
Treatment	
Psychotherapy	1 (2.9%)
Psychotropic drugs	6 (17.1%)
Gelotophobia	
No degree (<2.50)	33 (94.3%)
Slight degree (≥2.50)	1 (2.9%)
Marked degree (≥3.00)	1 (2.9%)

TABLE 3 | *M*, *SD*, and effect sizes (pre–post and pre–follow-up) for the ITT analysis of outcome measures ($N = 35$).

	Pre *M* (*SD*)	Post *M* (*SD*)	Follow-up *M* (*SD*)	Pre–post effect size (Cohen's *d*)	Pre–follow-up effect size (Cohen's *d*)
Coping humor (CHS)	2.47 (0.58)	2.80 (0.52)	2.77 (0.57)	0.88 [0.48–1.27]***	0.63 [0.26–0.99]**
Cheerfulness (STCI)	2.13 (0.54)	2.62 (0.61)	2.72 (0.62)	0.75 [0.37–1.12]***	0.88 [0.49–1.28]***
Seriousness (STCI)	2.99 (0.36)	2.64 (0.44)	2.54 (0.52)	0.80 [0.42–1.18] ***	0.91 [0.51–1.30]***
Bad mood (STCI)	2.07 (0.66)	1.66 (0.53)	1.59 (0.64)	0.61 [0.24–0.96]**	0.65 [0.28–1.01]***
Perceived stress (PSS)[a]	19.46 (5.47)	14.80 (4.63)	14.26 (4.06)	1.05 [0.63–1.46]***	1.09 [0.66–1.50]***
Depressiveness (CES-D)[a]	12.63 (5.49)	8.03 (4.11)	8.43 (5.93)	0.90 [0.51–1.29]***	0.81 [0.42–1.19]***
Anxiety (STAI)[a]	45.54 (9.61)	39.26 (8.33)	37.49 (10.21)	0.65 [0.28–1.01]***	0.62 [0.26–0.98]**
Well-being (WHO-5)[a]	15.06 (3.13)	17.49 (3.06)	17.26 (3.31)	1.05 [0.63–1.46]***	0.67 [0.30–1.03]***

*[a]Sum scores; 95% confidence intervals of effect sizes in square brackets; CHS, Coping Humor Scale; STCI, State-Trait-Cheerfulness Inventory – state version; PSS, Perceived Stress Scale; CES-D, Center for Epidemiological Studies Depression Scale; STAI, State-Trait-Anxiety Inventory – state version; WHO-5, WHO-5 Well-Being Index; *$p < 0.05$, **$p < 0.01$, ***$p \leq 0.001$.*

Table 3 summarizes mean values, standard deviations, and effect sizes for the outcomes at pre, post, and follow-up.

Aim 1: Improving Humor-Related Outcomes

Coping humor [$F_{(2,68)} = 14.21$, $p \leq 0.001$, $d = 1.29$], cheerfulness [$F_{(2,68)} = 19.05$, $p \leq 0.001$, $d = 1.50$], seriousness [$F_{(2,68)} = 19.01$, $p \leq 0.001$, $d = 1.50$], and bad mood [$F_{(1.69,57.52)} = 11.35$, $p \leq 0.001$, $d = 1.15$] showed significant main effects of time with large effect sizes. *Post hoc* tests with Bonferroni correction were also significant and effects for pre–post ranged from medium to large ($d = 0.61$ [0.24–0.96] to 0.88 [0.48–1.27]). Effects remained stable for pre–follow-up with medium to large effect sizes ($d = 0.63$ [0.26–0.99] to 0.91 [0.51–1.30]) too.

Aim 2: Improving Mental Health-Related Outcomes

Similar results were found for mental health related outcomes: Perceived stress [$F_{(1.35,46.01)} = 36.05$, $p \leq 0.001$, $d = 2.06$], depressiveness [$F_{(2,68)} = 16.00$, $p \leq 0.001$, $d = 1.37$], anxiety [$F_{(1.67,56.60)} = 10.81$, $p \leq 0.001$, $d = 1.13$], and well-being [$F_{(1.70,57.62)} = 13.57$, $p \leq .001$, $d = 1.26$] changed significantly over time with large effect sizes. Furthermore, *post hoc* tests with Bonferroni correction were significant on all outcomes (see **Table 3**) and effect sizes ranged from medium to large with stability until follow-up ($d = 0.62$ [0.26–0.98] to 1.09 [0.66–1.50]).

Aim 3: Evaluation of Applicability

Twenty-six participants (74.3%) completed the feedback questionnaire. Satisfaction with training (on a 1 to 5 Likert scale) was very high ($M = 4.46$, $SD = 0.65$). Understandability of the contents was rated highest ($M = 4.85$, $SD = 0.37$), whereas the improvement of symptoms lowest ($M = 3.58$, $SD = 0.81$); however, still ranging in the positive spectrum ("does partly apply" to "does rather apply"). Generally, participants would highly recommend the training (on a 1 to 4 Likert scale: $M = 3.54$, $SD = 0.58$). All means and standard deviations of the feedback items are summarized in **Table 4**.

Qualitative feedback revealed that participants mainly liked the group constellation and informal atmosphere ($n = 11$, 42.3%), the group leaders ($n = 11$, 42.3%), and the practical orientation of the training ($n = 4$, 15.4%). Negative feedback was primarily regarding lack of time ($n = 3$, 11.5%) and dropout/low motivation of other participants ($n = 3$, 11.5%). Furthermore, participants wanted more practical games and exercises ($n = 6$, 23.1%) and more time for the humor training in general ($n = 5$, 19.2%).

TABLE 4 | *M*, *SD* for the quantitative items of the feedback questionnaire ($n = 26$).

	M	*SD*
I was satisfied with the training as a whole.	4.46	0.65
The contents of the training have been well understandable for me.	4.85	0.37
The discussed topics have been interesting for me.	4.42	0.70
The structure of the sessions had a logic and comprehensible order for me.	4.46	0.51
The discussions about the humor topics were interesting.	4.23	0.86
The sharing of humor information was useful for me.	4.54	0.65
I liked the games within the humor training.	4.23	0.77
The mixture of theory and practice was convenient.	4.23	0.77
The location was comfortable.	3.73	1.08
I think I can transfer the learned in everyday life.	4.00	0.75
I can integrate humor after the training more willful in my everyday life.	3.88	0.71
Generally, I experience more cheerfulness after the training.	3.88	0.91
I think my problems have become better because of the humor training.	3.58	0.81
Would you recommend the training?[a]	3.54	0.58

1 = does not apply at all; 2 = does rather not apply; 3 = does partly apply; 4 = does rather apply; 5 = applies completely; [a]1 = no, in no case; 2 = rather not; 3 = rather yes; 4 = yes, in any case.

DISCUSSION

The present study pursued three main targets: to test the effects of a humor training on (1) humor-related outcomes, (2) mental health-related outcomes, and (3) to evaluate the applicability of the training based on the feedback of participants.

Humor-related outcomes (coping humor, cheerfulness, seriousness, and bad mood) evolved in the desired direction with medium to large effect sizes. A similar pattern of results was found for all of the mental health-related outcomes. Perceived stress, depressiveness, anxiety, and well-being showed medium to large effect sizes, too. Regarding applicability, the positive feedback of participants indicated that the training is applicable and accepted as people would recommend the training and were satisfied with it.

In the following, we want to go more into detail regarding interesting outcomes and results. First, perceived stress showed one of the strongest effects compared to all other outcomes. Time main effect, as well as *post hoc* tests, were highly significant with large effect sizes. Compared to a meta-analysis of MBSR for healthy adults (Khoury et al., 2015), the effect sizes of this study (pre–post: d = 1.05, pre–follow-up: d = 1.09) are comparable to the effect-sizes reported in the meta-analysis (pre–post: d = 0.83 [0.58–1.08]). One possible explanation for the strong decrease of perceived stress in this study could be explained by the improved coping humor abilities. Humor has already proven its efficacy as an adaptive way of coping with stress (Martin and Lefcourt, 1983). So, as people practiced this strategy in the training profoundly (and therefore increased in coping humor), it is not surprising that perceived stress simultaneously decreased throughout the course of training. This assumption is in line with the stability until follow-up. Participants might have continued to use coping humor until follow-up and implemented it as a preventive strategy against everyday life stress. Therefore, perceived stress continued to remain low until follow-up. However, further research with more follow-up measurements is needed, to evaluate if the relationship between coping humor and perceived stress can also be seen in the longer term. Another reason for the strong decrease of perceived stress could be due to enhanced laughter. Participants in the training had above average situations of mirth and laughter due to various games and role plays which were also transferred in daily life. Laughter is recognized as a stress-relieving process as it decreases cortisol (Berk et al., 1989), heart rate (Kraft and Pressman, 2012), and muscle tone (Paskind, 1932; Bennett and Lengacher, 2008). It might be possible that participants implemented more laughter in their everyday life and therefore had this strong decrease of stress.

Second, symptomatology like depressiveness and anxiety decreased. The positive effects of different humor interventions on depression have been already reported numerously (Beh-Pajooh et al., 2010; Crawford and Caltabiano, 2011; Gander et al., 2013; Cai et al., 2014; Proyer et al., 2014; Wellenzohn et al., 2016a,b), so the results of this study suit well with existing research and strengthen the assumption that humor can be an effective mechanism against depressive symptoms. Anxiety, however, has not been exhaustively investigated in the context of humor interventions/trainings yet. Research from non-interventional studies have already demonstrated anxiety-relieving effects of short-term induced humor (Yovetich et al., 1990; Szabo, 2003; Berk and Nanda, 2006; Ford et al., 2012). In line with these results, this study additionally shows that humor influences anxiety in the longer term, too. This assumption should be focused more in future research, testing the hypothesis that people suffering from increased anxiety or even anxiety disorders will profit from a humor intervention. There is only one study that confirms this hypothesis. Ventis et al. (2001) proved that a humorous desensitization was equally effective as a traditional systematic desensitization to reduce arachnophobia. More studies are definitely needed to evaluate the effects of humor on anxiety and anxiety disorders.

Summarizing, humor and mental health-related outcomes improved sustainably through humor training. However, one important aspect should be taken into account in future studies:

Moderating variables of humor trainings should be focused more, to be able to create a better person-intervention fit for the participants (Ruch and Hofmann, 2017). As already shown in other studies, inter-individual differences in trait cheerfulness play an important role in the effects of humor interventions (Papousek and Schulter, 2008; Hofmann et al., 2015), so it would be helpful to differentiate between high and low scorers and optimize the interventions based on these outcomes. Another personality variable which may influence the effects of humor trainings may be extraversion as extroverts experience more humor, especially benevolent humor, compared to introverts and may therefore differ in their humor behavior (Deaner and McConatha, 1993; Martin et al., 2003; Vernon et al., 2008).

Beyond inter-individual differences, another important moderator variable which should not be overlooked is group processes within the training. Humor is a social phenomenon (Martin, 2011) and inharmonic group constellations may influence the intervention's outcome as people may not engage in the humor training as intensively as in a harmonic group constellation. Group cohesion has a powerful impact on treatment effects of group interventions (Marziali et al., 1997; Yalom and Leszcz, 2005); therefore, this moderator should be further explored. Studies including group process outcomes are still scarce; however, they are highly interesting and required to create a holistic picture about the efficacy of humor trainings.

Besides humor and mental health-related outcomes, another important part of the study was the evaluation of applicability based on feedback. Generally, participants were very satisfied with the training as 14 persons (53.8%) marked the highest score of 5 on the satisfaction item and the mean ranged between 4 and 5 (M = 4.46, SD = 0.65). Furthermore, nearly everyone, except one person, would recommend the training (n = 25, 92.3%). Content-related topics were evaluated consistently as positive. Especially the understandability of the training's contents was rated highest (M = 4.45; SD = 1.24). Items about subjectively perceived change (in cheerfulness, humor, and symptoms) showed lower scores with mean values ranging from 3.5 to 4; however, they still range on the positive side of the scale.

Regarding dropout, the attrition rate of the study (17.1%) does not exceed the average attrition rate of 18.6% for group therapies of depression (McDermut et al., 2001). Also, compared to dropout in psychotherapy with up to 47%

(Wierzbicki and Pekarik, 1993), the dropout rate in this study can be ranked as rather low.

As can be seen, the training was evaluated consistently positive by the majority of participants. Also, the low attrition rate pleads for the acceptance of the training. In combination with the positive outcome results, there is definitely potential to further investigate humor trainings for subclinical samples and in the longer term, maybe even implement them also in health care systems as stress-preventive programs.

Limitations of the Study

Although results are promising, the limitations of the study should not be overlooked: First, study's design did not include a control condition and sample size was relatively small. Due to that, findings should be interpreted with caution and should not be generalized. Further studies are needed, which include control groups and also more sophisticated designs (as randomized-controlled trials), to be able to make evaluations on the efficacy of the training. Second, follow-up was relatively short (1 month) due to restricted time and resources. But, as the results show, many outcomes continued to remain stable or even improved in the follow-up. Thus, it would be interesting to investigate the effects of the training also in further follow-ups. Third, it has to be noted that some outcomes, as anxiety and depressiveness (Clark and Watson, 1991), highly interrelate and can interact. This interaction can bias the training's effect on outcomes. Fourth, all outcomes were assessed with state-sensitive questionnaires (either asking for the current situation or the past 2 weeks). However, it would be recommendable to assess also trait variables (e.g., trait cheerfulness) to see if the training can be effective in improving habitual outcomes sustainably. Lastly, people were thoroughly scanned in the face-to-face diagnostic interview based on SKID I and II; however, there are no objective cut-off criteria for subclinical symptoms of stress, depressiveness, anxiety, etc. in the interview. If a person reported subclinical symptoms and did not fulfill the conditions of a mental disorder concurrently, the person was invited to take part in the study. Future diagnostics should specify more concrete criteria for subclinical symptoms. However, as the interviews lasted on average at least 1 hour and the interviewers had experience in diagnostics, we assume that the lack of objective criteria for subclinical symptoms carried no big negative weight in the diagnostics because the interviewers nevertheless received a good overview of the participants' problems.

CONCLUSION

The humor training was effective in decreasing perceived stress, depressiveness, and anxiety whilst increasing coping humor, cheerfulness, and well-being in a subclinical sample. The feedback of participants was positive, indicating acceptance of the training.

However, as this was one of the first studies in this field, further research definitely is needed, including also more sophisticated designs like randomized controlled trials. Nevertheless, the results highlight the potential of humor trainings as preventive programs against stress and mental symptoms, and indicate new scopes of application, e.g., mental health care.

AUTHOR CONTRIBUTIONS

NT conceived and designed the work, analyzed and interpreted the data, and drafted the article. NT, VL, and ED carried out the data collection. A-RL critically revised the article.

ACKNOWLEDGMENTS

We would like to thank the group leaders for participating in the training. Also, we acknowledge financial support by the Open Access Publication Fund of the University of Salzburg.

REFERENCES

Aldao, A., Nolen-Hoeksema, S., and Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: a meta-analytic review. *Clin. Psychol. Rev.* 30, 217–237. doi: 10.1016/j.cpr.2009.11.004

Beh-Pajooh, A., Jahangiri, M., and Zahrakar, K. (2010). Humor skills training and reducing depression in patients with spinal cord injury. *J. Iran. Psychol.* 6, 227–236.

Bennett, M. P., and Lengacher, C. (2008). Humor and laughter may influence health: III. Laughter and health outcomes. *Evid. Based Complement. Alternat. Med.* 5, 37–40. doi: 10.1093/ecam/nem041

Berk, L. S., Tan, S. A., Fry, W. F., Napier, B. J., Lee, J. W., Hubbard, R. W., et al. (1989). Neuroendocrine and stress hormone changes during mirthful laughter. *Am. J. Med. Sci.* 298, 390–396. doi: 10.1097/00000441-198912000-00006

Berk, R. A., and Nanda, J. (2006). A randomized trial of humor effects on test anxiety and test performance. *Humor Int. J. Humor Res.* 19, 425–454. doi: 10.1515/humor.2006.021

Brähler, E., Mühlan, H., Albani, C., and Schmidt, S. (2007). Teststatistische prüfung und normierung der deutschen versionen des EUROHIS-QOL lebensqualität-index und des WHO-5 wohlbefindens-index [Testing and standardization of the German version of the EUROHIS-QOL and WHO-5 quality-of life-indices]. *Diagnostica* 53, 83–96. doi: 10.1026/0012-1924.53.2.83

Cai, C., Yu, L., Rong, L., and Zhong, H. (2014). Effectiveness of humor intervention for patients with schizophrenia: a randomized controlled trial. *J. Psychiatr. Res.* 59, 174–178. doi: 10.1016/j.jpsychires.2014.09.010

Clark, L. A., and Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *J. Abnorm. Psychol.* 100, 316–336. doi: 10.1037/0021-843X.100.3.316

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NY: Erlbaum.

Crawford, S., and Caltabiano, N. J. (2011). Promoting emotional well-being through the use of humour. *J. Posit. Psychol.* 6, 237–252. doi: 10.1080/17439760.2011.577087

Deaner, S. L., and McConatha, J. T. (1993). The relation of humor to depression and personality. *Psychol. Rep.* 72, 755–763. doi: 10.2466/pr0.1993.72.3.755

Falkenberg, I., Buchkremer, G., Bartels, M., and Wild, B. (2010). Implementation of a manual- based training of humor abilities in patients with depression: a pilot study. *Psychiatry Res.* 186, 454–457. doi: 10.1016/j.psychres.2010.10.009

Falkenberg, I., Jarmuzek, J., Bartels, M., and Wild, B. (2011). Do depressed patients lose their sense of humor? *Psychopathology* 44, 98–105. doi: 10.1159/000317778

Falkenberg, I., McGhee, P., and Wild, B. (2013). *Humorfähigkeit Trainieren: Manual für die Psychiatrisch-Psychotherapeutische Praxis [Training Humor Ability: Manual for the Psychiatric-Psychotherapeutic Work]*. Stuttgart: Schattauer.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf03193146

Ford, T. E., Ford, B. L., Boxer, C. F., and Armstrong, J. (2012). Effect of humor on state anxiety and math performance. *Humor Int. J. Humor Res.* 25, 59–74. doi: 10.1515/humor-2012-0004

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philos. Trans. R. Soc. B Biol. Sci.* 359, 1367–1378. doi: 10.1098/rstb.2004.1512

Fredrickson, B. L., and Joiner, T. (2002). Positive emotions trigger upward spirals toward emotional well-being. *Psychol. Sci.* 13, 172–175. doi: 10.1111/1467-9280.00431

Gander, F., Proyer, R. T., Ruch, W., and Wyss, T. (2013). Strength-based positive interventions: further evidence for their potential in enhancing well-being and alleviating depression. *J. Happiness Stud.* 14, 1241–1259. doi: 10.1007/s10902-012-9380-0

Gibbons, R., Hedeker, D., and Davis, J. (1993). Estimation of effect size from a series of experiments involving paired comparisons. *J. Educ. Stat.* 18, 271–279. doi: 10.2307/1165136

Golden-Kreutz, D. M., Thornton, L. M., Wells-Di Gregorio, S., Frierson, G. M., Jim, H. S., Carpenter, K. M., et al. (2005). Traumatic stress, perceived global stress, and life events: prospectively predicting quality of life in breast cancer patients. *Health Psychol.* 24, 288–296. doi: 10.1037/0278-6133.24.3.288

Grossman, P., Niemann, L., Schmidt, S., and Walach, H. (2004). Mindfulness-based stress reduction and health benefits: a meta-analysis. *J. Psychosom. Res.* 57, 35–43. doi: 10.1016/S0022-3999(03)00573-7

Hautzinger, M., Bailer, M., Hofmeister, D., and Keller, F. (2012). *Allgemeine Depressionsskala (ADS) [Center for Epidemiological Studies Depression Scale Revised (CESD-R)]*. Göttingen: Hogrefe.

Herring, D. R., Burleson, M. H., Roberts, N. A., and Devine, M. J. (2011). Coherent with laughter: subjective experience, behavior, and physiological responses during amusement and joy. *Int. J. Psychophysiol.* 79, 211–218. doi: 10.1016/j.ijpsycho.2010.10.007

Hirsch, R. D., Junglas, K., Konradt, B., and Jonitz, M. F. (2010). Humortherapie bei alten Menschen mit einer Depression [Humor therapy in the depressed elderly: results of an empirical study]. *Z. Gerontol. Geriatr.* 43, 42–52. doi: 10.1007/s00391-009-0086-9

Hofmann, J., Platt, T., Ruch, W., Niewiadomski, R., and Urbain, J. (2015). The influence of a virtual companion on amusement when watching funny films. *Motiv. Emot.* 39, 434–447. doi: 10.1007/s11031-014-9461-y

Jaremko, M., and Meichenbaum, D. (2013). *Stress Reduction and Prevention*. New York, NY: Springer Science & Business Media.

Khoury, B., Sharma, M., Rush, S. E., and Fournier, C. (2015). Mindfulness-based stress reduction for healthy individuals: a meta-analysis. *J. Psychosom. Res.* 78, 519–528. doi: 10.1016/j.jpsychores.2015.03.009

Klein, E. M., Brähler, E., Dreier, M., Reinecke, L., Müller, K. W., Schmutzer, G., et al. (2016). The German version of the Perceived Stress Scale – psychometric characteristics in a representative German community sample. *BMC Psychiatry* 16:159. doi: 10.1186/s12888-016-0875-9

Konradt, B., Hirsch, R. D., Jonitz, M. F., and Junglas, K. (2013). Evaluation of a standardized humor group in a clinical setting: a feasibility study for older patients with depression. *Int. J. Geriatr. Psychiatry* 28, 850–857. doi: 10.1002/gps.3893

Kraft, T. L., and Pressman, S. D. (2012). Grin and bear it. *Psychol. Sci.* 23, 1372–1378. doi: 10.1177/0956797612445312

Kugler, L., and Kuhbandner, C. (2015). That's not funny! - But it should be: effects of humorous emotion regulation on emotional experience and memory. *Front. Psychol.* 6:1296. doi: 10.3389/fpsyg.2015.01296

Labott, S. M., and Martin, R. B. (1987). The stress-moderating effects of weeping and humor. *J. Hum. Stress* 13, 159–164. doi: 10.1080/0097840X.1987.9936809

Laux, L., Glanzmann, P., Schaffner, P., and Spielberger, C. D. (1981). *State-Trait-Angstinventar (STAI) [State-Trait-Anxiety Inventory (STAI)]*. Weinheim: Beltz Testgesellschaft.

Lazarus, R. S., and Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York, NY: Springer.

Martin, R. A. (2011). *The Psychology of Humor: An Integrative Approach*. Burlington, MA: Elsevier Academic Press.

Martin, R. A., and Lefcourt, H. M. (1983). Sense of humor as a moderator of the relation between stressors and moods. *J. Pers. Soc. Psychol.* 45, 1313–1324. doi: 10.1037/0022-3514.45.6.1313

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., and Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: development of the humor styles questionnaire. *J. Res. Pers.* 37, 48–75. doi: 10.1016/s0092-6566(02)00534-2

Martinez-Marti, M. L., and Ruch, W. (2014). Character strengths and well-being across the life span: data from a representative sample of German-speaking adults living in Switzerland. *Front. Psychol.* 5:1253. doi: 10.3389/fpsyg.2014.01253

Marziali, E., Munroe-Blum, H., and McCleary, L. (1997). The contribution of group cohesion and group alliance to the of group psychotherapy. *Int. J. Group Psychother.* 47, 475–497. doi: 10.1080/00207284.1997.11490846

McDermut, W., Miller, I. W., and Brown, R. A. (2001). The efficacy of group psychotherapy for depression: a meta-analysis and review of the empirical research. *Clin. Psychol. Sci. Pract.* 8, 98–116. doi: 10.1093/clipsy.8.1.98

McGhee, P. (2010). *Humor as Survival Training for a Stressed-out World: The 7 Humor Habits Program*. Bloomington, IN: AuthorHouse.

McGhee, P. E. (1996). *Health, Healing and the Amuse System: Humor as Survival Training*. Dubuque, IA: Kendall/Hunt.

Overholser, J. C. (1992). Sense of humor when coping with life stress. *Pers. Individ. Dif.* 13, 799–804. doi: 10.1016/0191-8869(92)90053-r

Papousek, I., and Schulter, G. (2008). Effects of a mood enhancing intervention on subjective well-being and cardiovascular parameters. *Int. J. Behav. Med.* 15, 293–302. doi: 10.1080/10705500802365508

Paskind, H. A. (1932). Effect of laughter on muscle tone. *Arch. Neurol. Psychiatry* 28, 623–628. doi: 10.1001/archneurpsyc.1932.02240030143007

Proyer, R. T., Gander, F., Wellenzohn, S., and Ruch, W. (2014). Positive psychology interventions in people aged 50–79 years: long-term effects of placebo-controlled online interventions on well-being and depression. *Aging Ment. Health* 18, 997–1005. doi: 10.1080/13607863.2014.899978

Rosengren, A., Hawken, S., Ôunpuu, S., Sliwa, K., Zubaid, M., Almahmeed, W. A., et al. (2004). Association of psychosocial risk factors with risk of acute myocardial infarction in 11 119 cases and 13 648 controls from 52 countries (the INTERHEART study): case-control study. *Lancet* 364, 953–962. doi: 10.1016/s0140-6736(04)17019-0

Ruch, W., and Hofmann, J. (2017). "Fostering humour," in *Positive Psychology Interventions in Practice*, ed. C. Proctor (Cham: Springer), 65–80. doi: 10.1007/978-3-319-51787-2_5

Ruch, W. F., Hofmann, J., Rusch, S., and Stolz, H. (2018). Training the sense of humor with the 7 Humor Habits Program and satisfaction with life. *Humor Int. J. Humor Res.* doi: 10.1515/humor-2017-0099

Ruch, W., Köhler, G., and van Thriel, C. (1996). Assessing the "humorous temperament": construction of the facet and standard trait forms of the State-Trait-Cheerfulness-Inventory – STCI. *Humor Int. J. Humor Res.* 9, 303–340. doi: 10.1515/humr.1996.9.3-4.303

Ruch, W., Köhler, G., and van Thriel, C. (1997). To be in good or bad humour: construction of the state form of the State-Trait-Cheerfulness-inventory—STCI. *Pers. Individ. Dif.* 22, 477–491. doi: 10.1016/S0191-8869(96)00231-0

Ruch, W., and Proyer, R. T. (2008a). The fear of being laughed at: individual and group differences in Gelotophobia. *Humor Int. J. Humor Res.* 21, 47–67. doi: 10.1515/humor.2008.002

Ruch, W., and Proyer, R. T. (2008b). Who is gelotophobic? Assessment criteria for the fear of being laughed at. *Swiss J. Psychol.* 67, 19–27. doi: 10.1024/1421-0185.67.1.19

Samson, A. C., and Gross, J. J. (2012). Humour as emotion regulation: the differential consequences of negative versus positive humour. *Cogn. Emot.* 26, 375–384. doi: 10.1080/02699931.2011.585069

Segerstrom, S. C., and Miller, G. E. (2004). Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychol. Bull.* 130, 601–630. doi: 10.1037/0033-2909.130.4.601

Sliter, M., Kale, A., and Yuan, Z. (2014). Is humor the best medicine? The buffering effect of coping humor on traumatic stressors in firefighters. *J. Organ. Behav.* 35, 257–272. doi: 10.1002/job.1868

Solomon, D. A., Keller, M. B., Leon, A. C., Mueller, T. I., Lavori, P. W., Shea, M. T., et al. (2000). Multiple recurrences of major depressive disorder. *Am. J. Psychiatry* 157, 229–233. doi: 10.1176/appi.ajp.157.2.229

Strick, M., Holland, R. W., van Baaren, R. B., and van Knippenberg, A. (2009). Finding comfort in a joke: consolatory effects of humor through cognitive distraction. *Emotion* 9, 574–578. doi: 10.1037/a0015951

Szabo, A. (2003). The acute effects of humor and exercise on mood and anxiety. *J. Leis. Res.* 35, 152–162. doi: 10.1080/00222216.2003.11949988

Tagalidou, N., Faschingbauer, M., Mussuros, T., Ruch, W., and Laireiter, A.-R. (in press). Effectiveness and feasibility of a humor training in a routine care setting for people suffering from mental disorders. *Humor Int. J. Humor Res.*

JASP Team (2017). *JASP (Version 0.8.4)[Computer Software]*. New York, NY: JASP.

Uekermann, J., Channon, S., Lehmkamper, C., Abdel-Hamid, M., Vollmoeller, W., and Daum, I. (2008). Executive function, mentalizing and humor in major depression. *J. Int. Neuropsychol. Soc.* 14, 55–62. doi: 10.1017/s1355617708080016

Uekermann, J., Channon, S., Winkel, K., Schlebusch, P., and Daum, I. (2007). Theory of mind, humour processing and executive functioning in alcoholism. *Addiction* 102, 232–240. doi: 10.1111/j.1360-0443.2006.01656.x

Ventis, W. L., Higbee, G., and Murdock, S. A. (2001). Using humor in systematic desensitization to reduce fear. *J. Gen. Psychol.* 128, 241–253. doi: 10.1080/00221300109598911

Vernon, P. A., Martin, R. A., Schermer, J. A., and Mackie, A. (2008). A behavioral genetic investigation of humor styles and their correlations with the Big-5 personality dimensions. *Pers. Individ. Dif.* 44, 1116–1125. doi: 10.1016/j.paid.2007.11.003

Vigo, D., Thornicroft, G., and Atun, R. (2016). Estimating the true global burden of mental illness. *Lancet Psychiatry* 3, 171–178. doi: 10.1016/s2215-0366(15)00505-2

Wellenzohn, S., Proyer, R. T., and Ruch, W. (2016b). Humor-based online positive psychology interventions: a randomized placebo-controlled long-term trial. *J. Posit. Psychol.* 11, 584–594. doi: 10.1080/17439760.2015.1137624

Wellenzohn, S., Proyer, R. T., and Ruch, W. (2016a). How do positive psychology interventions work? A short-term placebo-controlled humor-based study on the role of the time focus. *Pers. Individ. Dif.* 96, 1–6. doi: 10.1016/j.paid.2016.02.056

Wiegner, L., Hange, D., Björkelund, C., and Ahlborg, G. (2015). Prevalence of perceived stress and associations to symptoms of exhaustion, depression and anxiety in a working age population seeking primary care - An observational study. *BMC Fam. Pract.* 16:38. doi: 10.1186/s12875-015-0252-7

Wierzbicki, M., and Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Prof. Psychol. Res. Pr.* 24, 190–195. doi: 10.1037/0735-7028.24.2.190

Wittchen, H.-U., Zaudig, M., and Fydrich, T. (1997). *Strukturiertes Klinisches Interview für DSM-IV (SKID I & II) [Structured Clinical Interview for DSM-IV (SCID I & II)]*. Göttingen: Hogrefe.

Yalom, I. D., and Leszcz, M. (2005). *The Theory and Practice of Group Psychotherapy*. New York, NY: Basic Books.

Yovetich, N. A., Dale, J. A., and Hudak, M. A. (1990). Benefits of humor in reduction of threat-induced anxiety. *Psychol. Rep.* 66, 51–58. doi: 10.2466/pr0.66.1.51-58

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Tagalidou, Loderer, Distlberger and Laireiter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 May 2018
doi: 10.3389/fpsyg.2018.00821

Who Benefits From Humor-Based Positive Psychology Interventions? The Moderating Effects of Personality Traits and Sense of Humor

Sara Wellenzohn[1], René T. Proyer[1,2] and Willibald Ruch[1]*

[1] Department of Psychology, University of Zurich, Zurich, Switzerland, [2] Department of Psychology, Martin Luther University of Halle-Wittenberg, Halle, Germany

The evidence for the effectiveness of humor-based positive psychology interventions (PPIs; i.e., interventions aimed at enhancing happiness and lowering depressive symptoms) is steadily increasing. However, little is known about who benefits most from them. We aim at narrowing this gap by examining whether personality traits and sense of humor moderate the long-term effects of humor-based interventions on happiness and depressive symptoms. We conducted two placebo-controlled online-intervention studies testing for moderation effects. In Study 1 (N = 104) we tested for moderation effects of basic personality traits (i.e., psychoticism, extraversion, and neuroticism) in the three funny things intervention, a humor-based PPI. In Study 2 (N = 632) we tested for moderation effects of the sense of humor in five different humor-based interventions. Happiness and depressive symptoms were assessed before and after the intervention, as well as after 1, 3, and 6 months. In Study 2, we assessed sense of humor before and 1 month after the intervention to investigate if changes in sense of humor go along with changes in happiness and depressive symptoms. We found moderating effects only for extraversion. Extraverts benefitted more from the three funny things intervention than introverts. For neuroticism and psychoticism no moderation effects were found. For sense of humor, no moderating effects were found for the effectiveness of the five humor-based interventions tested in Study 2. However, changes in sense of humor from pretest to the 1-month follow-up predicted changes in happiness and depressive symptoms. Taking a closer look, the playful attitude- and sense of humor-subscales predicted changes in happiness and depression for up to 6 months. Overall, moderating effects for personality (i.e., extraversion) were found, but none for sense of humor at baseline. However, increases in sense of humor during and after the intervention were associated with the interventions' effectiveness. Thus, we found humor-based interventions to be equally suited for humorous and non-humorous people, but increases in the sense of humor during the intervention phase could serve as an indicator whether it is worth continuing the intervention in the long-term.

Keywords: happiness, humor, personality, positive psychology, positive psychology interventions

OPEN ACCESS

Edited by:
Martin S. Hagger,
Curtin University, Australia

Reviewed by:
Ursula Beermann,
University of Innsbruck, Austria
Andres Mendiburo-Seguel,
Universidad Andrés Bello, Chile
Michaela Boerner,
Independent Researcher, Bayreuth, Germany

****Correspondence:***
Sara Wellenzohn
sara.wellenzohn@uzh.ch;
sarawellenzohn@msn.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *29 November 2017*
Accepted: *07 May 2018*
Published: *28 May 2018*

Citation:
Wellenzohn S, Proyer RT and Ruch W (2018) Who Benefits From Humor-Based Positive Psychology Interventions? The Moderating Effects of Personality Traits and Sense of Humor. Front. Psychol. 9:821. doi: 10.3389/fpsyg.2018.00821

INTRODUCTION

Positive Psychology is the scientific study of what makes life most worth living (Seligman and Csikszentmihalyi, 2000). It aims at promoting psychological research and practice in areas such as morally positively valued traits (character strengths), positive emotions, and positive institutions and their contribution to well-being. Another core topic of positive psychology is the development of so-called positive psychology interventions (PPIs; i.e., "[. . .] treatment methods or intentional activities that aim to cultivate positive feelings, behaviors, or cognitions"; Sin and Lyubomirsky, 2009, p. 468). Recent meta-analyses by Sin and Lyubomirsky (2009) and Bolier et al. (2013) found support for the notion that they are effective in enhancing happiness and ameliorating depressive symptoms.

One specific variant of PPIs are interventions, which focus on humor. Previous research provides support for the notion that they can enhance well-being in the general population (e.g., McGhee, 2010b; Crawford and Caltabiano, 2011; Gander et al., 2013; Proyer et al., 2014; Wellenzohn et al., 2016b; for an overview see Ruch and McGhee, 2014; Ruch and Hofmann, 2017), but also in clinical samples [e.g., Hirsch et al., 2010; Falkenberg et al., 2011; Konradt et al., 2013; see also Berger et al. (2017)]. There are group-administered training programs for humor that were found to be effective for enhancing emotional well-being, life satisfaction, psychological well-being, subjective health, positive mood, optimism, and lowering depression, feelings of stress or suicidal tendencies (e.g., Papousek and Schulter, 2008; Hirsch et al., 2010; Crawford and Caltabiano, 2011; Falkenberg et al., 2011; Ruch et al., 2018b; Tagalidou et al., 2018, Tagalidou et al., in press; for an overview see McGhee, 2010a,b). Thus, humor-based PPIs are expected to be well-received by the participants and enable a higher commitment to continue practicing and incorporating the activities into daily life. It has been shown that humor induces amusement (Ruch, 2001, 2008, 2009; Auerbach et al., 2016), an important facet of positive emotions (the one that most frequently goes along with laughter; Platt et al., 2013). Given that the elicitation of positive emotions is one of the proposed working mechanisms of PPIs (Sin and Lyubomirsky, 2009), humor seems to be particularly well-suited for incorporation in PPIs. Furthermore, Wellenzohn et al. (2016a) found support for savoring positive emotions serving as a working mechanism in humor-based PPIs.

While evidence for the effectiveness of PPIs is steadily growing, only little knowledge exists on whether (and how) certain personality traits moderate these effects. This is especially of interest from an applied perspective since the person × intervention fit (i.e., the degree to which an intervention matches an individual's preferences and personality) is associated with an intervention's effectiveness (e.g., Schueller, 2010, 2012, 2014; Proyer et al., 2015). We report two studies that are aimed at narrowing this gap in the literature by testing the impact of basic personality traits and sense of humor as defined by McGhee (1999, 2010a) as moderators in humor-based PPIs.

Humor-Based Online Positive Psychology Interventions

Seligman et al. (2005) published the first large-scale online placebo-controlled PPI study. They report findings for three self-administered online PPIs that are effective for up to 6 months in ameliorating depressive symptoms and enhancing happiness in comparison with a placebo control condition: The *gratitude visit*- (i.e., writing and delivering a gratitude letter to a person who has not been thanked so far), *three good things*- (i.e., writing down three good things that happened during the day), and *using signature strengths in a new way*-intervention (i.e., participants complete a character strengths inventory and receive feedback on their five highest strengths and the instruction to apply these strengths in a new way). An advantage of these online programs is that they are more cost effective than programs in group- or individual-settings as they are scalable (i.e., they can be easily distributed and made accessible to a large number of interested users) and can be self-administered using standardized written instructions; both are typically associated with low expenses for the researcher applying and supervising these programs in practice. There is also initial experience with humor-based online interventions. For example, Gander et al. (2013) adapted the *three good things*-intervention to a *three-funny things*-intervention by changing the instruction to include humor as its core component—instead of writing down three good things that happened to the person during the day, participants were asked to write down three *funny* things that happened to them during the day. The authors found the intervention to be effective in enhancing happiness for up to 3 months and ameliorating depressive symptoms up to 6 months after the intervention-week compared to a placebo control condition. Similar effects were recently found for a sample of people aged 50–79 years (Proyer et al., 2014).

A third study by Wellenzohn et al. (2016b) replicated the findings for the *three funny things*-intervention and adapted four other well-established PPIs into 1-week humor-based PPIs (see Wellenzohn et al., 2016b for a more detailed description of the interventions); namely, (a) the *gratitude visit*- (Seligman et al., 2005) was adapted into the *collecting funny things*-intervention (i.e., remembering the funniest things ever experienced and writing them down in as much detail as possible); (b) the *counting kindness*- (Otake et al., 2006) into the *counting funny things*-intervention (i.e., counting all funny things that happen during the day and note the total number); (c) the *using your signature strengths in a new way*- (Seligman et al., 2005) into the *applying humor*-intervention (i.e., noticing the humorous experiences during the day and add humorous activities); and (d) the *one door closes and another door opens*- (Rashid and Anjum, 2008) into the *solving stressful situations in a humorous way*-intervention (i.e., thinking about a stressful experience and how it could have been solved in a humorous way). These newly adapted interventions (self-administered over 1 week) were then tested in an online-setting by comparing their long-term effectiveness with a placebo control condition (early childhood memories as in Seligman et al., 2005). As in earlier studies, the *three funny things*-intervention was effective in increasing well-being, but there

were no effects for depression. Furthermore, two out of the four newly adapted humor-based PPIs enhanced happiness (counting funny things- and applying humor-) and two were effective in ameliorating depressive symptoms (*applying humor-* and *solving stressful situations in a humorous way*-intervention) for up to 6 months. Hence, three out of the five tested interventions were effective in enhancing well-being and ameliorating depression and more research in this area seems warranted.

Who Benefits Most From a Humor-Based Positive Psychology Intervention?

Thus far, only few studies have directly examined the influence of individual difference variables in PPIs, and the findings are mixed. Senf and Liau (2013) showed that higher levels in extraversion and openness contribute to greater increases in happiness after a gratitude-based intervention. Greater extraversion was also associated with a stronger reduction in depressive symptoms following a gratitude- and a strengths-based intervention. Schueller (2012) also found that extraverted participants benefit more from a gratitude-intervention, as well as from a savoring-intervention. However, contrary to the findings by Senf and Liau (2013), Schueller found stronger benefits for introverts from a strengths-based-intervention. Furthermore, he also found introverts to benefit more from an *active-constructive responding-* and a *three good things*-intervention. Extraversion seems to play an important role for the effectiveness of interventions (e.g., when having to interact with others or share experiences with others), this could also be expected by extensive literature that supports robust positive associations of extraversion with well-being (e.g., Pavot et al., 1990; Oerlemans and Bakker, 2014). Ng (2015) tested the role of neuroticism in a gratitude/kindness-intervention and found that participants with low levels in neuroticism demonstrated greater increases in happiness. However, a recent study using a randomized, group-based-design for interventions targeting the components of Seligman's (2002) Authentic Happiness Theory (i.e., the pleasurable, engaged, and meaningful life) has found no moderating effect of personality in the sense of the big five personality traits (Proyer et al., 2016). In the same line, Wang et al. (2017) did not find any moderating effects of personality for a well-being intervention in adolescents (only for the control phase). Hence, several studies suggest that individual difference variables moderate the effectiveness of some PPIs and encourage further research into the person × intervention fit as there seem to be intervention-specific differences in how far personality variables may have an impact. Thus far, no study has tested moderating effects of individual differences variables in humor-based interventions. Based on the existing literature, we expect humor-based PPIs to work better for those higher in extraversion. This hypothesis also receives support from correlational studies showing a positive relation between measures of humor and extraversion (e.g., Köhler and Ruch, 1996).

In addition to basic personality traits, *sense of humor* might be an important moderating variable for humor-based interventions. There are numerous conceptualizations of the sense of humor (for an overview see Ruch, 2007, 2008). McGhee (1999) provides a multi-faceted model that is based on six hierarchically ordered humor-skills or -habits (i.e., enjoyment of humor, laughter, verbal humor, humor in everyday life, laughing at oneself and finding humor under stress). He argues that these humor-skills are malleable in order to increase ones sense of humor (McGhee, 2010a,b). McGhee defines sense of humor as an ability to cope with stressful situations in daily life. He sees playfulness as its basis and argues that humor is a variant of play, namely the play with ideas (for an overview see Ruch and Heintz, 2018). A playful attitude can be seen as a facilitating frame of mind for establishing humor and for successfully processing humorous stimuli along with positive mood. McGhee's (1999) framework seems best-suited for a further exploration in PPI studies as he also developed a measure specifically for usage in intervention studies (i.e., the *Sense of Humor Scale;* McGhee, 2010a). We aim to test Wellenzohn et al.'s (2016b) hypothesis on the moderating role of the sense of humor in humor-based PPIs and its potential in predicting long-term changes in happiness and depressive symptoms.

The Present Studies

Our main aim is to examine the moderating effects of personality and the sense of humor on the effectiveness of humor-based interventions in a set of two studies. In Study 1, we test basic personality traits (i.e., the superfactors of personality *psychoticism, extraversion,* and *neuroticism* in Eysenck's personality model; see e.g., Eysenck and Eysenck, 1985) as moderators for the effectiveness of the *three funny things*-intervention (re-analyzing data from the study by Gander et al., 2013). Based on the existing literature, we expect humor-based PPIs to be more effective for people low in neuroticism and high in extraversion. In Study 2, we examine sense of humor as conceptualized by McGhee (2010a) as a moderator in the *three funny things*-intervention as well as in four further humor-based PPIs (re-analyzing data from the study by Wellenzohn et al., 2016b). Furthermore, we test (a) whether changes in sense of humor from pretest to the 1-month follow-up can predict long-term changes in happiness and depressive symptoms, and (b) whether changes in sense of humor and its sub-components differ in their ability to predict changes in happiness and depressive symptoms. Both studies are placebo-controlled online intervention-studies with happiness and depressive symptoms assessed at pre- and posttest as well as at 1, 3, and 6 months follow-ups.

Those with a higher sense of humor (according to McGhee's conceptualization; McGhee, 2010a) are more often exposed to humorous situations and thus, might come up with funny things to write down more easily (the core of the three funny things-intervention), to remember (as in the collecting funny things-intervention), or also noticing funny things during the day more easily (as in the counting funny things-intervention). Moreover, those with high scores in sense of humor might also find it easier to come up with ideas on how and where to apply humor in a new way (as in the applying humor-intervention), or be more creative in solving stressful situations in a humorous way. Thus, we expect those with higher levels in sense of humor to benefit more from humor-based PPIs. Furthermore, as the sense of humor might be

a trigger of positive emotions, we expect early changes in sense of humor and its sub-components to predict upward changes in happiness and amelioration of depression.

STUDY 1

Method

Participants

The total sample consisted of $N = 104$ women who completed all follow-up assignments in the *three funny things*-intervention ($n = 55$) or the placebo control condition ($n = 49$) in the study[1] by Gander et al. (2013). Their mean age was 45.16 years ($SD = 9.75$), ranging from 19 to 79. The participants were generally well-educated, with 26.9% having a university degree, 17.3% having a degree from an applied university, 22.1% having a certificate that would allow them to attend university, and 33.7% having completed vocational training.

Instruments

The *Eysenck Personality Questionnaire-Revised* (EPQ-R; Eysenck and Eysenck, 1985; German version by Ruch, 1999) consists of 102 items with a yes/no answer-format for the assessment of psychoticism (32 items, $\alpha = 0.63$), extraversion (23 items, $\alpha = 0.79$), and neuroticism (25 items, $\alpha = 0.84$), and additionally a lie scale (22 items, $\alpha = 0.74$) to cover social desirability.

The *Authentic Happiness Inventory* (AHI; Seligman et al., 2005) is a subjective measure for the assessment of overall happiness in the past week. Its reliability and validity, in the original as well as the German version, was supported by a broad range of studies (e.g., Ruch et al., 2010; Proyer et al., 2015). Every item consists of five statements (e.g., from to "Most of the time I feel bored" to "Most of the time I feel fascinated by what I am doing"). In Study 1, a 33-item version was used and in Study 2 a newer, revised version with 24 items was used. Internal consistency at pretest in Study 1 was $\alpha = 0.91$.

The *Center for Epidemiologic Studies Depression Scale* (CES-D; Radloff, 1977; in the German Adaption by Hautzinger and Bailer, 1993) consists of 20-items with a four-point scale ranging from 0 (*Rarely or none of the time [Less than 1 day]*) to 3 (*Most or all of the time [5–7 days]*) and measures the frequency of depressive symptoms in the past week (e.g., "My sleep was restless"). Internal consistency at pretest in Study 1 was $\alpha = 0.92$.

Procedure

The study was advertised as a free strengths-training in leaflets, in newspapers and magazines. The participants registered on a website that was set up for the administration of the program and were randomly assigned to either the *three funny things*-intervention (i.e., writing down three funny things that happened during the day), or the placebo control condition (i.e., writing about early childhood memories; see Seligman et al., 2005; Gander et al., 2013). All participants filled in the basic demographics and baseline-questionnaires (i.e., AHI, CES-D, and EPQ-R). They subsequently received instructions for the intervention and conducted the intervention for the following seven consecutive days. After the intervention-week, as well as 1, 3, and 6 months after the intervention, they logged on to the website and completed the AHI and the CES-D. Participants received an automatically generated personalized feedback on their well-being scores over the course of 6 months at the end of the study.

Results

Preliminary Analysis

Descriptive statistics for the AHI ($M = 2.98$, $SD = 0.49$), the CES-D ($M = 15.56$, $SD = 10.73$) and the EPQ-R as well as correlations between the personality variables and the AHI and CES-D at pretest are presented in **Table 1**. The table shows the expected findings in the cross-sectional analysis. Extraversion was robustly positively correlated with happiness and negatively with depression, while neuroticism demonstrated a negative relation with happiness, but was positively associated with depression.

Moderating Effects of Personality

In order to test potential moderating effects of the three personality dimensions (extraversion, neuroticism, and psychoticism), we computed hierarchical regression analyses. We analyzed interaction effects between each personality dimension and the group-condition on happiness (averaged over the four follow-ups), controlling for the baseline level in

[1]The attrition rate reported there was 37% in the intervention group and 51% in the placebo control group.

TABLE 1 | Descriptive statistics and moderating effects of personality at baseline on happiness and depressive symptoms in the three funny things condition compared to the placebo control-condition for Study 1.

			AHI	CES-D	Happiness					Depression				
	M	*SD*	*r*	*r*	*df*	*B*	*SE B*	*t*	*p*	*df*	*B*	*SE B*	*t*	*p*
Psychoticism	7.68	3.42	−0.02	−0.06	99	−13.67	18.60	−0.74	0.23	99	11.97	9.68	1.24	0.11
Extraversion	11.96	4.46	0.39***	−0.20*	99	23.46	9.88	2.37	0.01	99	−9.19	5.24	−1.75	0.04
Neuroticism	14.00	5.25	−0.53***	0.41***	99	−5.84	9.41	−0.62	0.27	99	0.81	4.73	0.17	0.43
Lie-scale	8.36	3.81	−0.05	0.08	99	−0.34	11.74	−0.03	0.49	99	−1.50	6.12	−0.24	0.41

*N = 104. r = partial correlation with AHI/CES-D at pretest controlled for age. Happiness/Depression = Personality × condition interaction (0 = Placebo control condition, 1 = Three funny things-intervention) as predictor of the happiness/depression scores after the intervention (all follow-ups averaged), when controlling for pretest scores in happiness/depression and personality. *p < 0.05, **p < 0.01, ***p < 0.001 (two-tailed).*

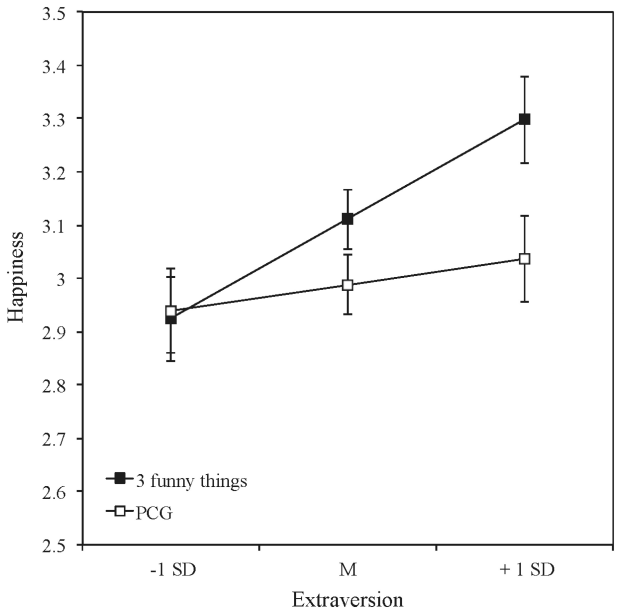


FIGURE 1 | Happiness after the intervention (all time periods averaged, controlled for the pretest scores) for the three funny things condition and the placebo control condition (PCG) for different levels of extraversion.

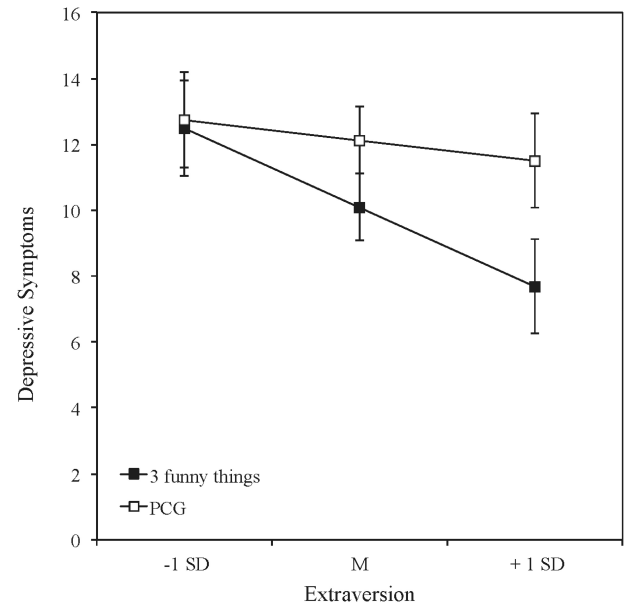


FIGURE 2 | Depressive symptoms after the intervention (all time periods averaged, controlled for the pretest scores) for the three funny things condition and the placebo control condition (PCG) for different levels of extraversion.

the AHI. For this calculation we used the SPSS PROCESS-macro by Hayes (2013). This macro allows analyzing the direct effect (regression controlled for the mediator), the total effect (regression without including the mediator), and the indirect effect. The same analyses were conducted for depressive symptoms (see **Table 1** for the interaction effects). Extraversion moderated the effectiveness of the intervention on happiness and also on depressive symptoms. **Figures 1**, **2** show the direction of the interaction-effects of extraversion for happiness and depressive symptoms. Error Bars represent the standard errors of the group differences.

Higher levels in extraversion went along with greater increases in happiness (**Figure 1**) and greater decreases in depressive symptoms (**Figure 2**) in the *three funny things*-intervention in comparison with the placebo control condition.

While Study 1 has shown that extraversion plays a role for the effectiveness of a humor-based PPI, Study 2 examines the role of individual differences in the sense of humor as an additional moderator in PPIs.

STUDY 2

Method

Participants

Of the 1,472 participants who have started the intervention (thereof, 243 in the placebo control condition), we used a sample of N = 632 adults (117 men and 515 women) who completed all follow-up measurements in the study by Wellenzohn et al. (2016b). The participants' mean age was 47.38 (SD = 11.55) and they were rather well educated with 41.5% having a university degree, 19.1% having a degree from an applied university, 22.1% having a certificate that would allow them to attend university, and 3.5% having completed vocational training.

Instruments

As in Study 1, the AHI (α = 0.93) and the CES-D (α = 0.88) were used.

The *Sense of Humor Scale* (SHS: by McGhee, 2010a; used in the German version by Proyer et al., 2010) assesses *playfulness* vs. *serious attitude*, *positive* vs. *negative mood* and *sense of humor* with its six sub-facets (enjoyment of humor, laughter, verbal humor, humor in everyday life, laughing at yourself, and humor under stress), as well as a total score for a more global assessment of sense of humor (see Müller and Ruch, 2011; Ruch and Heintz, 2018). The internal consistency at pre-test was α = 0.92 for the SHS Total Score, α = 0.71 for the playfulness dimension, α = 0.85 for the mood dimension, and α = 0.85 for sense of humor (for its sub-facets it ranged from α = 0.51 for the enjoyment of humor to α = 0.84 for the humor under stress sub-facet; median = 0.69). The SHS consists of 40 items (e.g., "I often find humor in things that happen at work") on a 7-point answer-scale.

Procedure

The procedure is comparable to Study 1 using the same recruitment strategy, but data were collected independently in the two studies. Participants were randomly assigned to one of the five humor-based PPIs (short descriptions are given in the introduction of the present article) or the placebo control-condition (i.e., writing about early childhood experiences). The dropout rate in the intervention groups varied between 55.3%

and 58.3%, and was 56.8% for the placebo control condition. Happiness and depressive symptoms were also assessed at pre- and posttest as well as at follow-up after 1-, 3-, and 6-months. Participants completed the SHS at pretest and at the 1-month follow-up.

Results

Preliminary Analyses

Descriptive statistics and the relations between the SHS scales and the AHI and CES-D at pretest are presented in **Table 2**.

The table shows that the means are in the expected range. Correlations with happiness and depressive symptoms were comparable with those reported by Proyer et al. (2010) for personal well-being. The dependent variables were robustly negatively correlated at pretest ($r = -0.58$, $p < 0.01$) without indicating redundancy.

Moderating Effects of Sense of Humor

To examine the moderating role of the sense of humor as measured with the SHS (McGhee, 2010a) on the effectiveness of humor-based PPIs, we computed the interaction-effects between the conditions (i.e., the humor-based PPIs vs. the placebo control condition) and the SHS Total Score on happiness and depressive symptoms, averaged over the four follow-ups, while controlling for pretest scores in happiness and depressive symptoms, and the SHS Total Score. As in Study 1, the same macro by Hayes (2013) was used for the analyses (**Table 3**).

Table 3 shows that *none* of the interaction-effects were significant.

While **Table 3** shows the analyses for the total score of the SHS only, we also computed the respective analyses for the playfulness scale, the positive vs. negative mood scale, the sense of humor scale, and the six humor skills. However, none of these analyses showed significant interaction effects (findings are not shown in detail, but are available upon request from the authors). In these analyses, the *t*-values for happiness ranged between 0.00 and 0.79 (median = 0.02) and between 0.02 and 1.40 (median = 0.15) for depression (all *n.s.*).

For a more in-depth analysis, initial changes in the SHS scales (changes from baseline to 1 month after completion of the intervention) were used for the prediction of changes in happiness and depressive symptoms (=criteria). Hierarchical regression analyses were conducted. In Step 1 age and sex were entered as predictors (yielding no incremental contribution in the prediction of happiness or depression; ≤0.01%). In Step 2, the initial changes in the SHS scales (changes from pretest to the 1-month follow-up) were entered as predictors of changes in happiness and depressive symptoms. The analyses were conducted for a total score of changes (i.e., an average score for the 1-, 3-, and 6-months follow-ups), but also separately for changes from the pretest to the 1 month follow-up, the 3 months follow-up, and the 6 months follow-up. The results for Step 2 are displayed in **Table 4**.

The table shows that, as expected, early changes in humor predicted changes in happiness and in depressive symptoms at most of the time points. The multiple squared correlation coefficients for Step 2 for the averaged follow-ups ranged between 0.03 (enjoyment of humor) and 0.18 (total score of the SHS; median = 0.05) for happiness and between 0.00 (enjoyment of humor) and 0.11 (positive mood; median = 0.02) for depression. On average, these coefficients were larger for the 1-month follow-up than for the later follow-ups, but the trends were more or less comparable in all cases.

TABLE 2 | Bivariate correlations between the AHI, the CES-D, and the components of the sense of humor scale controlled for age and sex for Study 2.

	M	*SD*	r_{AHI}	r_{CES-D}
AHI at pretest	3.16	0.48		
CES-D at pretest	10.28	5.70		
SHS tot	4.47	0.71	0.51	−0.38
Playful	4.86	0.80	0.45	−0.33
Mood	4.81	0.99	0.69	−0.56
SoH	4.23	0.80	0.32	−0.23
Enjoy	3.79	0.98	0.11	−0.05
Laughter	3.75	1.04	0.31	−0.24
Verbal	3.99	1.05	0.24	−0.19
Eday	4.92	0.94	0.33	−0.25
YSelf	4.66	1.10	0.22	−0.15
Stress	4.28	1.21	0.25	−0.15

N = 628. AHI, Authentic Happiness Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; SHS tot, total score in the Sense of Humor Scale; Playful, playful vs. serious attitude; Mood, positive vs. negative mood; SoH, sense of humor; Enjoy, enjoyment of humor; Verbal, verbal humor; Eday, humor in everyday life; YSelf, laughing at yourself; Stress, humor under stress. All correlations are significant at the 0.1%-level (two-tailed) except for "enjoy humor" at 1% for the AHI and non-significant for the CES-D.

DISCUSSION

This study provides first data on moderating effects of three basic personality traits on a *humor*-based PPI; namely, the *three funny things*-intervention. Those higher in *extraversion* demonstrated greater benefit from the intervention. This finding is in line with data on positive associations of extraversion and well-being (e.g., Pavot et al., 1990; Oerlemans and Bakker, 2014). We did not find effects for psychoticism and neuroticism; also the tendency toward socially desirable answering behavior was not related to the interventions' effectiveness. For psychoticism, the coefficients might have been slightly affected by the comparatively low reliability of this scale. The findings for extraversion are in line with Senf and Liau's (2013) work, who found similar results for a signature strengths and gratitude intervention (see Seligman et al., 2005). Similarly, Schueller (2012) found, when varying the gratitude visit-intervention with different degrees of social interactions needed, that delivering a gratitude letter in person also yielded greater benefits for those higher in extraversion, than without any personal contact. One might argue that the *three funny things*-intervention (at least implicitly) also addresses social interaction situations—as funny things might be more likely to be experienced in the company of others or that people actively engaged in more contact with others for experiencing more funny things. The latter would be in line with findings that only

TABLE 3 | Moderating effects of the sense of humor total score at baseline on happiness and depressive symptoms in five different humor-based interventions compared to the placebo control-condition (n = 105) for Study 2.

	n	Happiness					Depression				
		df	*B*	*SE B*	*t*	*p*	*df*	*B*	*SE B*	*t*	*p*
Three funny things	101	201	−0.04	0.06	−0.70	0.48	201	0.05	0.05	0.96	0.34
Collecting funny things	105	205	−0.00	0.05	−0.02	0.98	205	0.02	0.05	0.32	0.74
Counting funny things	108	208	0.01	0.06	0.11	0.92	208	0.03	0.05	0.62	0.54
Applying humor	104	204	−0.02	0.06	−0.32	0.74	204	−0.00	0.05	−0.02	0.98
Solving stressful situations	109	209	−0.03	0.06	−0.49	0.62	209	0.07	0.05	1.40	0.16

Happiness/Depression = Sense of humor × condition interaction (0 = Placebo control condition, 1 = Humor-based intervention) as predictor of the happiness/depression scores after the intervention (all follow-ups averaged), when controlling for pretest scores in happiness/depression and sense of humor. Solving stressful situations = Solving stressful situations in a humorous way. p (two-tailed).

behaving more extravert could already contribute to a persons' well-being (see Fleeson et al., 2002).

It might be advisable to include variations of the standard instructions in future studies to make the activity more accessible to introverts. Otherwise, a different humor-based intervention (see McGhee, 1999, 2010a; Wellenzohn et al., 2016b; Ruch and Hofmann, 2017) may be more suitable for those low in extraversion. One might speculate that presenting ideas on situations or experience that provide humorous incidents without other people being present might make this intervention equally effective for extraverts and introverts. Hence, one aim for future application might be to develop interventions that are equally suitable for individuals with different levels of extraversion, or change the instructions in a way that all can work well with the included activities (e.g., introverts might find additional examples of observing humor in situations with people they know well rather than with strangers or persons that are less well-known to them, easier to work with).

Findings of Study 2 show that the sense of humor (as conceptualized by McGhee, 1999, 2010a) had *no* moderating effects on the effectiveness of five humor-based interventions. From a practical point of view this can be seen as "good news" since participants with varying levels of sense of humor (not only those with greater inclinations) seem to benefit from these interventions. It seems as if the interventions are accessible to participants similarly irrespective of self-reported sense of humor. Although there were some trends in the conducted analyses, they seem to be negligible from a practical point of view.

Although, McGhee's (2010a) *Sense of Humor Scale* is only one way of assessing sense of humor, and the coefficients might have been slightly affected by the rather low reliability of the enjoyment of humor subscale, one might argue that a measurement which is closer to the interventions and more sensitive for (upward) changes, would be able to detect moderating qualities of sense of humor; in this case we would argue similarly to what Seligman et al. (2005) have put forward when introducing the Authentic Happiness Inventory for the assessment of happiness in PPI studies (Proyer et al., in press). However, our findings show that changes in sense of humor are associated with success in the interventions. The changes in sense of humor from pretest to 1 month after the intervention predicted the changes in happiness and depressive symptoms for up to 6 months. Thus, sense of humor might be a working mechanism for humor-based PPIs. Additionally, other models have recently been put forward, which might also be used for developing interventions, and/or assessing the moderating role of humor-related variables (see Ruch, 2012; Ruch and Heintz, 2016; Ruch et al., 2018a).

It is argued that humor-based interventions have a great potential for improving well-being. Given that there is large variety in how humorous behavior is expressed in daily life (Craik et al., 1996; Heintz, 2017) it would be interesting to study (a) whether certain of these behaviors are more strongly related to changes in the desired direction than others and (b) whether personality and/or sense of humor moderate effects of interventions that are based on different humorous behaviors (e.g., those pursued alone vs. in groups). For the latter, a new theoretical framework is needed that enables differentiating among types of humor. One such framework could be the study of the shared and distinct effects of interventions based on preferences and usages of comic styles (i.e., fun, humor, nonsense, wit, irony, satire, sarcasm, and cynicism; Ruch et al., 2018a). Such an approach will help developing humor-based interventions and the study of potentially moderating effects further in a structured way.

Limitations

We do not know what exactly the participants in our study wrote down and what they experienced as being funny. The latter is a limitation of online studies in general as it is more difficult to control whether and to which degree (as instructed, more or less) participants completed the assigned interventions at home (or in other environments) on their own (or with help of others). This is of particular importance as it was shown that such factors (i.e., continued practice of an intervention, effort invested in the activity, the preference for an activity, or early reactivity in the desired direction) are potent predictors of the effectiveness of an intervention (Schueller, 2010; Proyer et al., 2015; see also Lyubomirsky and Layous, 2013).

Another limitation of Study 1 is, that the sample consisted solely of women. This was due to the opportunity to advertise the study through an article in a women's magazine. Thus, we do not know, if extraversion would also moderate the effectiveness of humor-based PPIs in men, or if other basic personality traits would play a role in a more diverse sample.

TABLE 4 | Hierarchical regression analyses (step 2) of initial changes in sense of humor and its components on changes in happiness and depressive symptoms in the humor-based PPIs controlled for age and sex for Study 2.

		Changes in all follow-ups		Changes after 1 month		Changes after 3 months		Changes after 6 months	
Initial changes		ΔF	ΔR^2	ΔF	ΔR^2	ΔF	ΔR^2	ΔF	ΔR^2
SHS tot	AHI	113.81***	0.18	194.05***	0.27	60.11***	0.10	36.74***	0.07
	CES-D	35.08**	0.06	104.79***	0.17	9.67**	0.02	1.76	0.00
Playful	AHI	41.73***	0.07	53.27***	0.09	27.54***	0.05	16.59***	0.03
	CES-D	7.13**	0.01	22.02***	0.04	2.64	0.01	0.04	0.00
Mood	AHI	116.31***	0.18	243.69***	0.32	55.46***	0.10	31.31***	0.07
	CES-D	66.52***	0.11	211.03***	0.30	21.05***	0.04	2.77†	0.01
SoH	AHI	57.56***	0.10	86.57***	0.14	32.15***	0.06	21.52***	0.04
	CES-D	13.07***	0.02	36.74***	0.07	2.64	0.01	1.12	0.00
Enjoy	AHI	18.54***	0.03	23.97***	0.04	10.75**	0.02	8.58**	0.02
	CES-D	1.40	0.00	6.49*	0.01	0.03	0.00	0.22	0.00
Laughter	AHI	24.33***	0.04	39.58***	0.07	13.15***	0.02	8.44**	0.02
	CES-D	6.01*	0.01	19.53***	0.04	1.49	0.00	0.11	0.00
Verbal	AHI	27.40***	0.05	27.97***	0.05	23.98***	0.04	10.57**	0.02
	CES-D	3.45†	0.01	8.74**	0.02	1.20	0.00	0.17	0.00
Eday	AHI	25.40***	0.05	40.46***	0.07	14.55***	0.03	8.53**	0.02
	CES-D	9.23**	0.02	23.26***	0.04	2.86†	0.01	0.63	0.00
YSelf	AHI	21.44***	0.04	36.69***	0.07	7.59**	0.01	10.39**	0.02
	CES-D	8.87**	0.02	13.33***	0.03	0.62	0.00	5.97*	0.01
Stress	AHI	19.11***	0.04	30.56***	0.06	10.84**	0.02	6.46*	0.01
	CES-D	5.68*	0.01	20.83***	0.04	2.59	0.01	0.07	0.00

*N = 527. PPIs, positive psychology interventions ; Initial changes, changes in sense of humor and its components from pretest to the 1-month follow-up; AHI, Authentic Happiness Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; SHS tot, total score in the Sense of Humor Scale; Playful, playful vs. serious attitude; Mood, positive vs. negative mood; SoH, sense of humor; Enjoy, enjoyment of humor; Verbal, verbal humor; Eday, humor in everyday life; YSelf, laughing at yourself; Stress, humor under stress; Changes in happiness = changes in happiness from pretest to the averaged follow-ups; Changes in depression = changes in depressive symptoms from pretest to the averaged follow-ups. †$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (two-tailed).*

Hence, a replication with a more diverse (also in terms of ages represented) and representative sample will be needed for strengthening the findings and exploring potential further moderation effects.

A limitation of both studies is that the dependent variables and also the potential moderators were assessed via self-reports only. Thus, it would be helpful to have more objective indicators of these variables (e.g., including peer-ratings from knowledgeable others). To the best of our knowledge there is only one study that has also considered peer-reports in a humor-based intervention study (or intervention study in general; Ruch et al., 2018b). One might argue that sense of humor is a highly observable trait and, thus, people might also react differently if a person behaves more humorously after an intervention. Their feedback (e.g., eliciting positive emotions due to joint laughter; verbal and facial reactions; etc.) may encourage future humorous behavior, which may have a further positive effect. Hence, it could be tested whether perceived changes in sense of humor by others are associated with changes in the dependent variables. It would also be interesting to see whether an inept use of humor would leads to more negative feedbacks and may even have detrimental effects. One might think of, for example, gelotophobes that have difficulties seeing positive effects in humor and may experience or only anticipate laughter in others as being negative or feel uncomfortable when trying to engage more actively with humor (for an overview see Ruch et al., 2014). Additionally, we did not have data available for sense of humor (in McGhee's conceptualization) *and* the basic personality traits simultaneously for a joint analysis. Thus, the present findings warrant more investigations of potential moderators of humor-based PPIs, for example to examine their relative importance.

ETHICS STATEMENT

The federal ethics committee of the canton of Zurich, Switzerland provided approval. These studies were carried out in accordance with the recommendations of the Ethical Principles of Psychologists and Code of Conduct (APA) and the Ethical Guidelines for Psychologists of the Swiss Psychological Society (SGP), as outlined by the ethics committee of the Faculty of Arts at the University of Zurich, with online informed consent from all subjects. All subjects gave online informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

SW, RP, and WR: conception or design of the work, data collection, data analysis and interpretation, and final approval of the published version. SW: drafted the article. RP and WR: critical revision of the article.

FUNDING

This study was supported by research grants from the Swiss National Science Foundation (SNSF; Grants Nos. 100014_132512 and 100014_149772) awarded to RP and WR.

ACKNOWLEDGMENTS

The authors are grateful to Fabian Gander for his help with the data collection and to Marisa De Lannay for proofreading the manuscript.

REFERENCES

Auerbach, S., Ruch, W., and Fehling, A. (2016). Positive emotions elicited by clowns and nurses: an experimental study in a hospital setting. *Transl. Issues Psychol. Sci.* 2, 14–24. doi: 10.1037/tps0000055

Berger, P., Bitsch, F., Böhl, H., and Falkenberg, I. (2017). Play and playfulness in psychiatry: a selective review. *Int. J. Play* 6, 1–16. doi: 10.1080/21594937.2017.1383341

Bolier, L., Haverman, M., Westerhof, G., Riper, H., Smit, F., and Bohlmeijer, E. (2013). Positive psychology interventions: a meta-analysis of randomized controlled studies. *BMC Public Health* 13:119. doi: 10.1186/1471-2458-13-119

Craik, K. H., Lampert, M. D., and Nelson, A. J. (1996). Sense of humor and styles of everyday humorous conduct. *Humor Int. J. Humor Res.* 9, 273–302. doi: 10.1515/humr.1996.9.3-4.273

Crawford, S. A., and Caltabiano, N. J. (2011). Promoting emotional well-being through the use of Humour. *J. Posit. Psychol.* 6, 237–252. doi: 10.1080/17439760.2011.577087

Eysenck, H. J., and Eysenck, M. W. (1985). *Personality and Individual Differences: A Natural Science Approach*. New York, NY: Plenum Press. doi: 10.1007/978-1-4613-2413-3

Falkenberg, I., Buchkremer, G., Bartels, M., and Wild, B. (2011). Implementation of a manual-based training of humor abilities in patients with depression: a pilot study. *Psychiatry Res.* 186, 454–457. doi: 10.1016/j.psychres.2010.10.009

Fleeson, W., Malanos, A. B., and Achille, N. M. (2002). An intraindividual process approach to the relationship between extraversion and positive affect: is acting extraverted as "good" as being extraverted? *J. Pers. Soc. Psychol.* 83, 1409–1422. doi: 10.1037//0022-3514.83.6.1409

Gander, F., Proyer, R. T., Ruch, W., and Wyss, T. (2013). Strength-based positive interventions: further evidence for their potential in enhancing well-being and alleviating depression. *J. Happiness Stud.* 14, 1241–1259. doi: 10.1007/s10902-012-9380-0

Hautzinger, M., and Bailer, M. (1993). *Allgemeine Depressionsskala (ADS) [General Depression Scale]*. Göttingen: Hogrefe.

Hayes, A. F. (2013). "Introduction to mediation, moderation, and conditional process analysis. A regression-based approach," in *Methodology in the Social Sciences*, ed. T. D. Little (New York, NY: Guilford).

Heintz, S. (2017). Putting a spotlight on daily humor behaviors: dimensionality and relationships with personality, subjective well-being, and humor styles. *Pers. Individ. Dif.* 104, 407–412. doi: 10.1016/j.paid.2016.08.042

Hirsch, R. D., Junglas, K., Konradt, B., and Jonitz, M. F. (2010). Humortherapie bei Alten Menschen mit einer Depression: ergebnisse einer empirischen Untersuchung. *Z. Gerontol. Geriatr.* 43, 42–52. doi: 10.1007/s00391-

Köhler, G., and Ruch, W. (1996). Sources of variance in current sense of humor inventories: how much substance, how much method variance? *Humor Int. J. Humor Res.* 9, 363–398. doi: 10.1515/humr.1996.9.3-4.363

Konradt, B., Hirsch, R. D., Jonitz, M. F., and Junglas, K. (2013). Evaluation of a standardized humor group in a clinical setting: a feasibility study for older patients with depression: humor therapy for older patients with depression. *Int. J. Geriatr. Psychiatry* 28, 850–857. doi: 10.1002/gps.3893

Lyubomirsky, S., and Layous, K. (2013). How do simple positive activities increase well-being? *Curr. Dir. Psychol. Sci.* 22, 57–62. doi: 10.1177/0963721412469809

McGhee, P. (2010a). *Humor as Survival Training for a Stressed-Out World: The 7 Humor Habits Program*. Bloomington, IN: AuthorHouse.

McGhee, P. (2010b). *Humor: The Lighter Path to Resilience and Health*. Bloomington, IN: AuthorHouse.

McGhee, P. E. (1999). *Health Healing and Amuse System: Humor As Survival Training*, 3rdEdn. Dubuque, IA: Kendall Hunt Pub Co.

Müller, L., and Ruch, W. (2011). Humor and strengths of character. *J. Posit. Psychol.* 6, 368–376. doi: 10.1080/17439760.2011.592508

Ng, W. (2015). Use of positive interventions: does neuroticism moderate the sustainability of their effects on happiness? *J. Posit. Psychol.* 11, 51–61. doi: 10.1080/17439760.2015.1025419

Oerlemans, W. G. M., and Bakker, A. B. (2014). Why extraverts are happier: a day reconstruction study. *J. Res. Pers.* 50, 11–22. doi: 10.1016/j.jrp.2014.02.001

Otake, K., Shimai, S., Tanaka-Matsumi, J., Otsui, K., and Fredrickson, B. L. (2006). Happy people become happier through kindness: a counting kindnesses intervention. *J. Happiness Stud.* 7, 361–375. doi: 10.1007/s10902-005-3650-z

Papousek, I., and Schulter, G. (2008). Effects of a mood-enhancing intervention on subjective well-being and cardiovascular parameters. *Int. J. Behav. Med.* 15, 293–302. doi: 10.1080/10705500802365508

Pavot, W., Diener, E., and Fujita, F. (1990). Extraversion and happiness. *Pers. Individ. Dif.* 11, 1299–1306. doi: 10.1016/0191-8869(90)90157-M

Platt, T., Hofmann, J., Ruch, W., and Proyer, R. T. (2013). Duchenne display responses towards sixteen enjoyable emotions: individual differences between no and fear of being laughed at. *Motiv. Emot.* 37, 776–786. doi: 10.1007/s11031-013-9342-9

Proyer, R. T., Gander, F., Wellenzohn, S., and Ruch, W. (2014). Positive psychology interventions in people aged 50-79 years: long-term effects of placebo-controlled online interventions on well-being and depression. *Aging Ment. Health* 8, 997–1005. doi: 10.1080/13607863.2014.899978

Proyer, R. T., Gander, F., Wellenzohn, S., and Ruch, W. (in press). The authentic happiness inventory revisited: addressing its psychometric properties, validity, and role in intervention studies. *J. Well-Being Asses.* doi: 10.1007/s41543-018-0006-0

Proyer, R. T., Gander, F., Wellenzohn, S., and Ruch, W. (2016). Addressing the role of personality, ability, and positive and negative affect in positive psychology interventions: findings from a randomized intervention based on the authentic happiness theory and extensions. *J. Posit. Psychol.* 11, 609–621. doi: 10.1080/17439760.2015.1137622

Proyer, R. T., Ruch, W., and Müller, L. (2010). Sense of humor among the elderly: findings using the German version of the SHS. *Z. Gerontol. Geriatr.* 43, 19–24. doi: 10.1007/s00391-009-0082-0

Proyer, R. T., Wellenzohn, S., Gander, F., and Ruch, W. (2015). Toward a better understanding of what makes positive psychology interventions work: predicting happiness and depression from the person × intervention fit in a follow-up after 3.5 years. *Appl. Psychol. Health Well Being* 7, 108–128. doi: 10.1111/aphw.12039

Radloff, L. S. (1977). The CES-D scale. A self-report depression scale for research in the general population. *Appl. Psychol. Meas.* 1, 385–401. doi: 10.1177/014662167700100306

Rashid, T., and Anjum, A. (2008). "Positive psychotherapy for young adults and children," in *Handbook of Depression in Children and Adolescents*, eds J. R. Z. Abela and B. L. Hankin (New York, NY: Guilford), 250–287.

Ruch, W. (1999). Die revidierte fassung des eysenck personality questionnaire und die konstruktion des deutschen EPQ-R bzw. *EPQ-RK. Z. Differ. Diagnostische Psychol.* 20, 1–24. doi: 10.1024//0170-1789.20.1.1

Ruch, W. (2001). "The perception of humor," in *Emotion, Qualia, and Consciousness*, ed. A. W. Kaszniak (Tokyo: Word Scientific Publisher), 410–425. doi: 10.5167/uzh-77826

Ruch, W. (2007). *The Sense of Humor: Explorations of a Personality Characteristic*, 2nd Edn. Berlin: de Gruyter. doi: 10.1515/9783110804607

Ruch, W. (2008). "Psychology of humor," in *The Primer of Humor Research*, ed. V. Raskin (Berlin: de Gruyter), 17–100. doi: 10.1515/9783110198492.17

Ruch, W. (2009). "Amusement," in *The Oxford Companion to the Affective Sciences*, eds D. Sander and K. Scherer (New York, NY: Oxford University Press), 27–28.

Ruch, W. (2012). *Towards a New Structural Model of the Sense of Humor: Preliminary Findings*. Available at: http://www.aaai.org/ocs/index.php/FSS/FSS12/paper/viewFile/.(5652)/5809

Ruch, W., and Heintz, S. (2016). The virtue gap in humor: exploring benevolent and corrective humor. *Transl. Issues Psychol. Sci.* 2, 35–45. doi: 10.1037/tps0000063

Ruch, W., and Heintz, S. (2018). Psychometric evaluation of the revised Sense of Humor Scale and the construction of a parallel form. *Humor Int. J. Humor Res.* 31, 235–257. doi: 10.1515/humor-2016-0085

Ruch, W., Heintz, S., Platt, T., Wagner, L., and Proyer, R. T. (2018a). Broadening humor: comic styles differentially tap into temperament, character, and ability. *Front. Psychol.* 9:6. doi: 10.3389/fpsyg.2018.00006

Ruch, W., Hofmann, J., Stolz, H., and Rusch, S. (2018b). The impact of the 7 humor habits Program on the sense of humor and satisfaction with life. *Humor Int. J. Humor Res.* doi: 10.1515/humor-2017-0099 [Epub ahead of print].

Ruch, W., and Hofmann, J. (2017). "Fostering humor," in *Positive Psychology Interventions in Practice*, ed. C. Proctor (New York, NY: Springer), 65–80. doi: 10.1007/978-3-319-51787-2_5

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. T. (2014). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor Int. J. Humor Res.* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., and McGhee, P. (2014). "Humor intervention programs," in *Handbook of Positive Psychological Interventions*, eds A. C. Parks and S. M. Schueller (Oxford: Wiley-Blackwell), 193–197.

Ruch, W., Proyer, R. T., Harzer, C., Park, N., Peterson, C., and Seligman, M. E. P. (2010). Values in action inventory of strengths (VIA-IS): adaptation and validation of the German version and the development of a peer-rating form. *J. Individ. Dif.* 31, 138–149. doi: 10.1027/1614-0001/a000022

Schueller, S. M. (2010). Preferences for positive psychology exercises. *J. Posit. Psychol.* 5, 192–203. doi: 10.1080/17439761003790948

Schueller, S. M. (2012). Personality fit and positive interventions: extraverted and introverted individuals benefit from different happiness increasing strategies. *Psychology* 3, 1166–1173. doi: 10.4236/psych.2012.312A172

Schueller, S. M. (2014). "Person–activity fit in positive psychological interventions," in *The Wiley Blackwell Handbook of Positive Psychological Interventions*, eds A. C. Parks and S. M. Schueller (Hoboken, NJ: John Wiley & Sons, Ltd), 385–402. doi: 10.1002/9781118315927.ch22

Seligman, M. E. P. (2002). *Authentic Happiness*. NewYork, NY: Simon and Schuster.

Seligman, M. E. P., and Csikszentmihalyi, M. (2000). Positive psychology: an introduction. *Am. Psychol.* 55, 5–14. doi: 10.1037/0003-066X.55.1.5

Seligman, M. E. P., Steen, T. A., Park, N., and Peterson, C. (2005). Positive psychology progress: empirical validation of interventions. *Am. Psychol.* 60, 410–421. doi: 10.1037/0003-066X.60.5.410

Senf, K., and Liau, A. (2013). The effects of positive interventions on happiness and depressive symptoms, with an examination of personality as a moderator. *J. Happiness Stud.* 14, 591–612. doi: 10.1007/s10902-012-9344-4

Sin, N. L., and Lyubomirsky, S. (2009). Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. *J. Clin. Psychol.* 65, 467–487. doi: 10.1002/jclp.20593

Tagalidou, N., Faschingbauer, M., Mussuros, T., Ruch, W., and Laireiter, A.-R. (in press). Training for promoting humor and cheerfulness in people with mental disorders of a community mental health center. *Humor Int. J. Humor Res*.

Tagalidou, N., Loderer, V., Distlberger, E., and Laireiter, A.-R. (2018). Feasibility of a humor training to promote humor and decrease stress in a subclinical sample: a single-arm pilot study. *Front. Psychol.* 9:577. doi: 10.3389/fpsyg.2018.00577

Wang, R. A. H., Nelson-Coffey, S. K., Layous, K., Bao, K. J., Davis, O. S. P., and Haworth, C. M. A. (2017). Moderators of wellbeing interventions: why do some people respond more positively than others? *PLoS One* 12:e0187601. doi: 10.1371/journal.pone.0187601

Wellenzohn, S., Proyer, R. T., and Ruch, W. (2016a). How do positive psychology interventions work? A short-term placebo-controlled humor-based study on the role of the time focus. *Pers. Individ. Dif.* 96, 1–6. doi: 10.1016/j.paid.2016.02.056

Wellenzohn, S., Proyer, R. T., and Ruch, W. (2016b). Humor-based online positive psychology interventions: a randomized placebo-controlled long-term trial. *J. Posit. Psychol.* 11, 584–594. doi: 10.1080/17439760.2015.113762

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Wellenzohn, Proyer and Ruch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 16 November 2017
doi: 10.3389/fpsyg.2017.02006

Telling Friend from Foe: Listeners Are Unable to Identify In-Group and Out-Group Members from Heard Laughter

*Marie Ritter and Disa A. Sauter**

Department of Social Psychology, University of Amsterdam, Amsterdam, Netherlands

Group membership is important for how we perceive others, but although perceivers can accurately infer group membership from facial expressions and spoken language, it is not clear whether listeners can identify in- and out-group members from non-verbal vocalizations. In the current study, we examined perceivers' ability to identify group membership from non-verbal vocalizations of laughter, testing the following predictions: (1) listeners can distinguish between laughter from different nationalities and (2) between laughter from their in-group, a close out-group, and a distant out-group, and (3) greater exposure to laughter from members of other cultural groups is associated with better performance. Listeners ($n = 814$) took part in an online forced-choice classification task in which they were asked to judge the origin of 24 laughter segments. The responses were analyzed using frequentist and Bayesian statistical analyses. Both kinds of analyses showed that listeners were unable to accurately identify group identity from laughter. Furthermore, exposure did not affect performance. These results provide a strong and clear demonstration that group identity cannot be inferred from laughter.

Keywords: laughter, groups, emotion, in-group advantage, motivation

OPEN ACCESS

Edited by:
Willibald Ruch,
University of Zurich, Switzerland

Reviewed by:
Will Curran,
Queen's University Belfast, United Kingdom
Kai Alter,
Newcastle University, United Kingdom
Ursula Beermann,
University of Innsbruck, Austria

***Correspondence:**
Disa A. Sauter
d.a.sauter@uva.nl

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *03 September 2017*
Accepted: *02 November 2017*
Published: *16 November 2017*

Citation:
Ritter M and Sauter DA (2017) Telling Friend from Foe: Listeners Are Unable to Identify In-Group and Out-Group Members from Heard Laughter. Front. Psychol. 8:2006. doi: 10.3389/fpsyg.2017.02006

INTRODUCTION

Group membership is important for how we perceive others: Across a range of domains, people perform better when processing information from in-group members. For example, we attend more closely to faces from our own group (Byatt and Rhodes, 2004), we are better at recognizing the identity of in-group members (Hehman et al., 2010), and we are more accurate in identifying emotions from non-verbal expressions produced by members of our own group (Elfenbein and Ambady, 2002). In some cases, the *belief* that another is a member of the perceiver's own group is sufficient to confer these advantages. In a study by Thibault et al. (2006), participants were asked to identify the emotion on faces that participants were told belonged either to their own or to another, group. When participants thought that they were making judgments about an in-group member, they were better at recognizing the expressed emotion, regardless of the actual group membership of the expresser. This lends support to the motivational account, which explains the performance advantage for in-group members as the result of greater motivation to process information from in-group members more deeply (Thibault et al., 2006). If we think that someone is a member of our own group, we are thus more motivated to, for example, find out what they are feeling. In order for this motivational mechanism to operate, the perceiver first has to be able to accurately judge

whether the other person is a member of their own group. In the current study, we aimed to test whether listeners can discern group membership from hearing non-verbal expressions, specifically laughter.

Most research to date that has examined group membership has studied visual, rather than auditory perception. Visual experimental stimuli often contain clear features that distinguish groups, such as skin color (Cassidy et al., 2011). However, even for visual perception, determining group membership is not always entirely straightforward. In one study, Marsh et al. (2003) presented American participants with pictures of American–Japanese (American citizens with Japanese heritage) and Japanese (Japanese citizens with Japanese heritage) people, who posed with either neutral or emotional expressions. Participants were asked to categorize the pictures according to whether they thought the person was American–Japanese or Japanese. Participants performed better when judging the emotional expressions, as compared to the neutral expressions, suggesting that the emotional expressions may contain information about group membership akin to an accent in speech (e.g., Clopper and Pisoni, 2004b). Indeed, studies suggest that while observers agree on prototypical expressions of specific emotions (e.g., Ekman and Friesen, 1978), they also show culture-specific differences in how they express emotions, which has been dubbed emotion dialects (Elfenbein et al., 2007). These emotion dialects may be what perceivers use to infer group membership (Marsh et al., 2003), which then affects emotion recognition accuracy. However, less work has examined group membership inferences from vocal expressions beyond language.

For language-like vocal expressions, even brief vocal segments can convey group membership, as shown by Walton and Orlikoff (1994). They found that people could identify the ethnicity of a speaker 60% of the time from *a* sounds alone. More recently, Bryant et al. (2016) found that listeners could infer information about social relationships from human laughter. Specifically, listeners could identify whether people laughing together were friends or strangers. This suggests that human non-verbal vocalizations convey some information about social relationships, and perhaps might also carry group information. This would also be in line with research on chimpanzee calls, which has found that chimpanzees adjust their calls to distinguish themselves from close living groups (Crockford et al., 2004), and that these differences are meaningful to listeners (Herbinger et al., 2009).

Laughter is arguably the most extensively researched human non-verbal vocalization (Owren and Amoss, 2014). It occurs frequently, typically in social situations (Provine, 2004; Scott et al., 2014). Although different forms of laughter can communicate a range of social messages (Szameitat et al., 2010; Wildgruber et al., 2013), laughter is recognized across cultures as indicating amusement (Sauter et al., 2010). There are many different types of laughter, such as joyous, taunting, or tickling laughter, that seem to play distinct roles in social cognition (Szameitat et al., 2010; Wildgruber et al., 2013). Laughter can function as a signal of affiliation (Bryant et al., 2016), and may even constitute an extended form of grooming, through which social bonds are maintained and strengthened (Dezecache and Dunbar, 2004). Laughter thus presents a good candidate for examining group membership identification, given its ubiquity, sociality, and occurrence across cultures.

Only a single study to date has examined whether listeners can infer group membership from human non-verbal vocalizations. Sauter (2013, Experiment 1) tested Dutch participants' perception of vocalizations expressing amusement, relief, triumph, and sensual pleasure. The stimuli were from three different countries: the Netherlands (in-group), England (close out-group), and Namibia (distant out-group). Participants were first asked to classify the expressed emotion, and then to identify whether the person was from the Netherlands, another European country, or a country outside Europe. In the emotion recognition task, an in-group advantage was found, meaning that participants were more accurate in judging emotional expressions from members of their own cultural group. In contrast, participants were no better than chance at identifying group membership.

This result casts doubt on whether non-verbal vocalizations of emotion provide reliable group membership information. However, it is worth noting some limitations of Sauter's (2013) study: Firstly, it included vocalizations of multiple emotions. While this was necessary to test the in-group advantage for emotion recognition, it may have increased task difficulty in the group classification task. Secondly, the study by Sauter only included one nationality per group. This could have resulted in participants performing poorly due to the fact that they were unable to, for example, distinguish in-group from close out-group, even though they may have been able to accurately differentiate, for example, in-group from distant out-group. Thirdly, the study by Sauter employed only frequentist statistical analyses, which cannot provide support for a null hypothesis. The current study sought to remedy those limitations in order to provide a tougher test of the question of whether listeners can judge group membership from non-verbal vocalizations of emotion. We further sought to examine a potential role for familiarity in group identification judgments.

Although there is little evidence on the impact of familiarity on group identification in the context of non-verbal emotional expressions, studies of language perception point to a link between familiarity and accuracy for group identification (see Elfenbein and Ambady, 2002 for a similar result for emotion recognition). In one study, participants who had lived in many different US states were better at telling from which state a speaker came, compared to participants who had lived in one state for most of their lives (Clopper and Pisoni, 2004a). Baker et al. (2009) found a similar pattern in a study of the perceptions of an accent from the American state Utah. They found that participants who were from a state close to Utah (i.e., a close out-group), were nearly as good as the Utahans (i.e., members of the in-group), at identifying a Utahan accent. In contrast, participants from more distant states (i.e., the distant out-group), performed considerably worse, which was explained as being due to low familiarity with the Utahan accent. These results point to familiarity as a possible factor in group identification from vocal cues, and we therefore included a measure of exposure to other cultures in the current study, in order to test this possibility directly.

The Current Study

The current study sought to examine whether listeners could identify in- and out-group members from laughter segments. Following Sauter (2013), we employed nationality as a proxy for group membership, as national identity is a salient and reliable group dimension (Smith, 1991). In addition, we distinguished between in-group, close out-group, and distant out-group (Sauter, 2013).

In examining the question of whether listeners would be able to identify group membership from laughter, we made the following predictions, based on the literature reviewed above: We hypothesized that listeners would be able to distinguish between laughter from different nationalities (Specific Group Identification Hypothesis). We further predicted that listeners would be able to accurately judge whether a laughing person belonged to the listener's own in-group, a close out-group, or a distant out-group (Broad Group Identification Hypothesis). Finally, we predicted that greater exposure to laughter from members of other cultural groups would be associated with better performance (Familiarity Hypothesis).

METHODS

Design and Procedure

Before the experimental trials, participants were asked to report their age, sex, and level of education. They were also asked how many foreign countries they had traveled to, taken as a proxy for familiarity with laughter from other cultures. Participants were not asked to list the specific countries they had visited as it was assumed that participants would most likely have traveled primarily to countries geographically close to the Netherlands (e.g., France, England). Finally, as an exploratory measure, participants were asked how well they expected to perform in the experimental trials. As participants' expectations of their performance were not found to be related to their actual performance, this measure is not discussed further.

The experimental study had a within-participant design with six conditions, reflecting the six nationalities of the laughter stimuli: Dutch, English, French, US American, Japanese, and Namibian. Each stimulus was presented once in a random order that was fixed across participants. On each trial, participants listened to a laugh, and were asked in a six-way forced choice task from which nationality they thought the laughing person came. Participants were free to do the study with headphones or speakers and to set the sound level themselves. The study did not have a time limit. Upon completion of the study, participants were given feedback on how well they had done in the form of a total score of correct answers.

Stimuli

The study included a total of 24 stimuli, comprising four amused laughs per nationality. The Dutch, English, and Namibian laughter were taken from Sauter (2013); the US American laughter stimuli were taken from Simon-Thomas et al. (2009); the Japanese laughter stimuli were taken from Sauter et al. (in preparation). The French laughter stimuli were recorded in an equivalent way to those of Sauter (2013). All laughs were part of larger sets of recordings of emotional vocalizations. During the recordings, individuals posed laughs, but also laughed spontaneously. Consequently, there was some variability in spontaneity within each set.

The stimuli from each culture were randomly selected from each set of laughs, with the constraints that there is an equal number of male and female tokens of each nationality and that minimally two different speakers were included for each gender for each culture. The stimuli were recorded individually in a soundproof environment and were on average 2.37 (1.16) s long (see Table 1 in the Supplementary Material for average duration per condition).

Participants

The study was run online on the website of a Dutch popular science magazine (quest.nl) from June 12th to 26th, 2014, and was publicly accessible. Given that the Quest website in general, and the current study in particular, were in Dutch, participants are assumed to have been either Dutch or Belgian (or sufficiently acculturated to regard the Dutch as their in-group).

The study used an opportunistic sample, collecting as many responses as possible in the available time. Participants were asked whether they consented for their anonymous answers to be analyzed for scientific purposes, but were also given the option to participate without allowing scientific analysis of their data. The study was approved by the University of Amsterdam Department of Psychology ethics committee (reference code: 2014-SP-3736). All participants whose data are included in this manuscript provided written informed consent in accordance with the Declaration of Helsinki.

A total of 1,500 participants took part in the online study. Participants were excluded because (a) they did not consent for their test data to be used for scientific purposes (264 participants), (b) errors in the data log (5 participants), (c) they were less than 18 years old (75 participants), or (d) they did not complete the study (342 participants). The remaining 814 participants (527 women, 287 men) had a mean age of 30.87 years (range: 18–75 years).

RESULTS

Data Processing

To examine performance accuracy, H_u scores were calculated (Wagner, 1993). H_u scores are unbiased hit rates that correct for response biases, such as disproportionate use of one response alternative. Moreover, H_u scores correct for disproportionate presentation of one stimulus type (e.g., presentation of 12 close out-group stimuli vs. 4 in-group stimuli). Raw H_u scores range from 0 to 1, with 0 indicating only incorrect classifications, and 1 indicating perfect accuracy. The H_u scores for each condition are shown in **Figure 1**. The H_u scores were averaged across all conditions to provide a general measure of performance for each participant. This is referred to as the Mean H_u score. For ease of interpretation, the classifications are also provided in **Table 1** in percent.

Because H_u scores are proportional measures, the scores were arcsine transformed prior to further analysis to stabilize

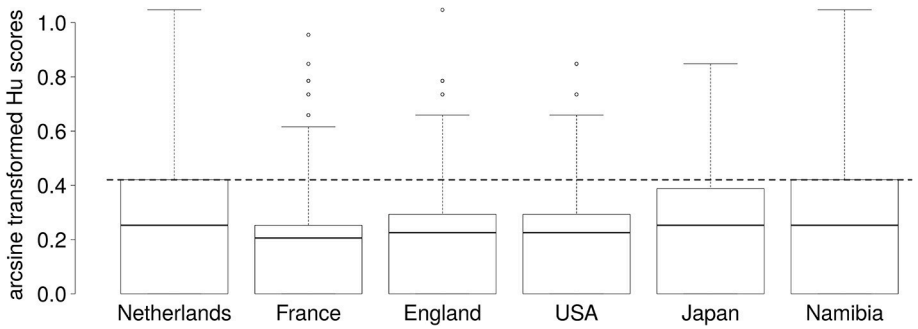


FIGURE 1 | Boxplots of arcsine transformed H_u scores for laughter stimuli from each of six countries. The dashed line indicates the chance level. Each box represents the interquartile range, the thick line in each box represents the median score, and the whiskers represent the maximum scores, excluding outliers. No lower whiskers are shown as the minimum scores fall within the interquartile range. Outliers (represented as circles) are scores that were higher or lower than the median by 1.5 times the interquartile range. Outliers were not excluded from any analyses.

TABLE 1 | Confusion matrix of answer proportions in percent.

Stimulus	Netherlands	France	England	USA	Japan	Namibia
JUDGMENT						
Netherlands	**26.32**	20.76	17.29	13.76	13.73	8.14
France	19.44	**18.52**	14.47	8.51	23.80	15.26
England	18.46	20.64	**18.86**	14.96	8.94	18.15
USA	7.63	17.20	13.45	**17.60**	30.28	13.85
Japan	12.75	21.93	15.14	10.84	**25.80**	13.54
Namibia	15.14	10.29	13.88	27.95	6.70	**26.04**

Classifications across the diagonal are correct classifications, shown in bold.

variance and normalize the data (see Wagner, 1993). Following this transformation, all variables were checked for normality with Shapiro–Wilk tests, which indicated that they were not normally distributed ($ps < 0.001$). We therefore employed a non-parametric equivalent of the *t*-test, the Wilcoxon Signed-Rank test for all comparisons between two conditions. For ANOVAs and regression analyses, parametric tests were used, as they are known to be robust against normality violations (Norman, 2010). ANOVAs were employed in all comparisons across three conditions and regressions were used in cases in which the independent variable was not nominal.

In order to allow us to accept or reject the null hypothesis with known certainty, all of the described tests were run with H_u scores using both frequentist analyses and the Bayesian equivalents. Frequentist analyses test the probability of the null hypothesis, given the data. Bayesian analyses test the probability of both the alternative and the null hypothesis, given the data. Consequently, conducting Bayesian analyses can yield evidence for either the null or the alternative hypothesis. Bayesian analyses calculate the probability distribution of a parameter (e.g., a difference score) by using the data to update the prior distribution, a parameter distribution based on what is known about the parameter from previous research or theoretical considerations (for an introduction to Bayesian analysis and modeling see Lee and Wagenmakers, 2013). The frequentist analyses were conducted with R (R Core Team, 2013). The Bayesian parametric analyses were run in JASP (The JASP Team, 2017). The non-parametric Bayesian one-sample *t*-tests were run using a computer program by van Doorn et al. (in preparation) The test estimates the effect size δ which is the difference between scores and chance level. The test uses a prior of $\delta \sim Cauchy\,(0, 1)$, a t-distribution with a single degree of freedom (Rouder et al., 2009). The Cauchy distribution offers a useful prior because it puts less weight on unrealistic values of δ, and it assumes that small effects occur with greater frequency. Bayes factors were computed with the Savage-Dickey density ratio. If the Bayes factor is greater than 1 then the analysis shows evidence for the alternative hypothesis. If the Bayes factor is lower than 1 then the analysis shows evidence for the null hypothesis. Bayes factors above 100 are considered "extreme evidence for the alternative hypothesis" (Jeffreys, 1961; for more information see Wetzels et al., 2010).

The Specific Group Identification Hypothesis

The Specific Group Identification Hypothesis predicted that participants can accurately infer group membership from laughter, when groups are operationalized as countries. The mean overall H_u scores were therefore compared to the chance level (i.e., 1/6). The frequentist test in the form of a Wilcoxon-Signed Rank tests showed that participants performed significantly worse than chance (Median of mean H_u score: 0.218, $p < 0.001$, $r = -0.85$). The Bayesian test also showed overwhelming evidence for the alternative hypothesis of participants performing significantly worse than chance. The effect size was estimated to have a median of −1.151 with a Bayesian 95% confidence interval of [−1.246, −1.058]. The prior and posterior distributions can be seen in **Figure 2**. These tests thus provided no support for the Specific Group Identification Hypothesis.

Although the overall scores clearly showed that performance was below chance levels, participants may have been able to detect laughter from individual countries at better-than-chance levels. Therefore, country-specific H_u scores were computed

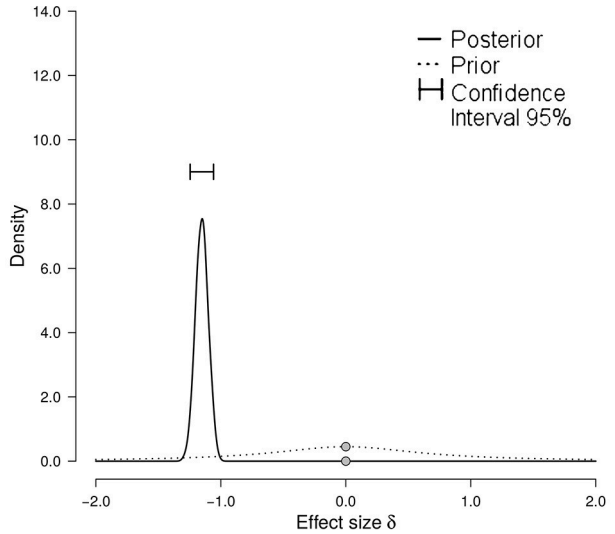


FIGURE 2 | Prior and posterior distribution with Bayesian confidence interval of the effect size δ. The prior distribution (dashed line) shows the distribution expected under the null hypothesis with no data (i.e., performance at chance level). The posterior distribution (solid line) shows the distribution that is expected given the data. The point of interest (zero) is marked with gray dots on both distributions. A score of zero on the x-axis represents performance at chance level.

TABLE 2 | Comparisons of group scores with chance level for Wilcoxon Signed-Rank Test and Bayesian equivalents using arcsine transformed H_u scores of laughter from individual countries (above) and grouped countries (below).

Stimulus origin	Median (H_u score)	Chance level (H_u score)	Effect size[a]
Netherlands	0.253	0.421	−0.60
France	0.206	0.421	−0.79
England	0.226	0.421	−0.75
USA	0.226	0.421	−0.77
Japan	0.253	0.421	−0.66
Namibia	0.253	0.421	−0.58
In-Group	0.252	0.421	−0.60
Close Out-Group	0.502	0.784	−0.86
Distant Out-Group	0.361	0.615	−0.85

All tests were significant at an α-level of 0.001, Bonferroni corrected for multiple comparisons.

[a]*Effect sizes are applicable to the frequentist analyses only.*

(see **Table 2**). These were individually compared to chance level using multiple Wilcoxon-Signed Rank tests, Bonferroni corrected for multiple comparisons, and the Bayesian equivalent test. All comparisons showed that the H_u scores were significantly below chance and Bayes factors showed that the alternative hypothesis with scores lower than chance was over 1,000 times more likely given the data. These results indicate that participants were not able to accurately infer group identity at the country-level for any of the countries.

Broad Group Identification Hypothesis

Next, we sought to test the Broad Group Identification Hypothesis, which predicted that participants can accurately infer group membership, when operationalized as in-group, close out-group, and distant out-group. H_u scores do not control for differing chance levels across conditions. Therefore, in order to test the Broad Group Identification Hypothesis, the difference between H_u score and chance level was calculated for each condition. When Dutch laughter was presented, there was only one correct answer out of the six response alternatives, and consequently, the chance level for the in-group was 1/6. For trials in the close out-group condition, there were three correct answers (French, English, US American) out of the six response alternatives. In that condition, the chance level was thus 3/6 (i.e., 1/2). When participants heard laughter from the distant out-group, there were two correct answers (Japanese, Namibian) out of the six response options. Therefore, the chance level was 2/6 (i.e., 1/3). In each condition, chance was subtracted from the H_u scores, resulting in difference scores.

A one-way repeated-measures ANOVA was run with the difference scores, comparing performance for in-group (the Netherlands), close out-group (England, France, and USA), and distant out-group (Japan, Namibia). As Mauchly's test indicated violation of the sphericity assumption ($W = 0.96$, $p = 0.002$, $\eta = 0.98$), Greenhouse-Geisser corrected scores are reported[1]. Performance differed significantly across the three conditions: $F_{GG(4.9,\ 3983.7)} = 43.79$, $p < 0.001$. In the Bayesian analyses, the alternative model which allowed differences between conditions was tested against a null model which did not allow for differences. As in the *t*-test, the prior was specified as a Cauchy distribution. There was a significant difference; $BF_{10} > 1{,}000$.

As can be seen in **Figure 3**, participants performed worse in the close out-group condition compared to the in-group ($V = 22{,}482$, $p < 0.001$; $BF_{10} > 1{,}000$) and distant out-group conditions ($V = 294{,}730$, $p < 0.001$; $BF_{10} > 1{,}000$). Moreover, participants performed better in the in-group condition compared to the distant out-group condition; $V = 90{,}108$, $p < 0.001$; $BF_{10} > 1{,}000$. Yet, in none of the conditions did participants perform better than chance (see **Table 2**).

The Familiarity Hypothesis

We predicted that greater exposure to laughter from members of other cultural groups would be associated with better performance (the Familiarity Hypothesis). There was considerable variability in how many countries participants had visited, with 20.1% having visited 1–5 countries, 39.3% having visited 6–10 countries, 33.5% having been to 11–20 countries, and 7.0% reporting having traveled to 21 or more countries.

A linear model was estimated to check whether the number of countries that participants had visited would predict group identification performance. In the Bayesian analysis, the JASP program uses multivariate generalizations of Cauchy priors on

[1]As Greenhouse-Geisser corrected ANOVAs can suffer from lower power, the analysis was rerun using a multilevel approach that is not affected by sphericity violations. The pattern of results was identical to those reported in the main text.

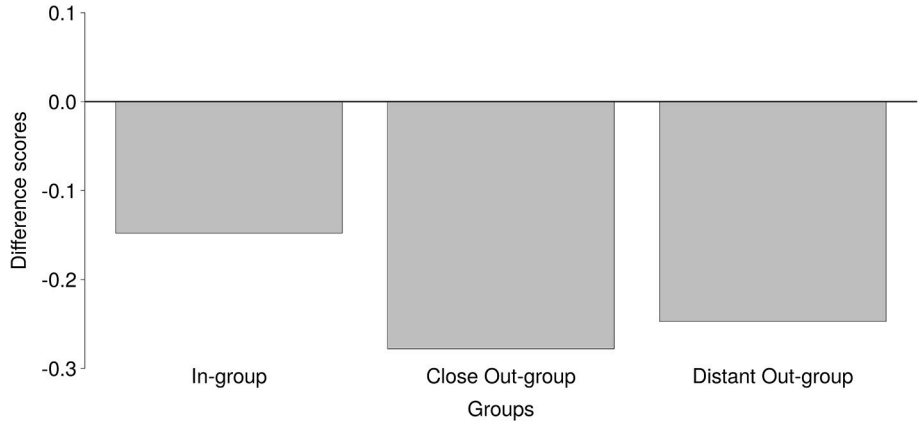


FIGURE 3 | Difference scores (H_u scores—chance level) for performance in the separate conditions In-group, Close Out-group, and Distant Out-group. Higher scores represent better performance.

standardized effects with a prior width of 0.5 (see Rouder et al., 2012). The results of both the frequentist and the Bayesian analysis showed that familiarity was not associated with performance [$F_{(3, 810)} = 1.066$, $p = 0.36$; $BF_{01} = 7.288$]. Note that this Bayes factor denotes the factor in favor of the null hypothesis. The Bayes factor in favor of the alternative hypothesis was $BF_{10} = 0.137$.

A further exploratory analysis was conducted because we considered it likely that Dutch participants would have mainly traveled to foreign countries that are in the close out-group, such as France or England, compared to countries that are less popular travel destinations from the Netherlands, such as Namibia or Japan. Therefore, we speculated that familiarity may be relevant mainly for the close out-group. We therefore tested whether performance in the close out-group condition was higher for participants with greater exposure to foreign cultures. However, there was no significant association [$F_{(3, 810)} = 1.93$, $p = 0.12$; $BF_{01} = 9.430$]. The Bayes factor in favor of the alternative hypothesis BF_{10} was 0.11.

DISCUSSION

This study investigated whether listeners can identify group membership from individual laughter segments. Neither frequentist nor Bayesian analyses yielded any support for participants being able to reliably perform group identification based on laughter sounds: Participants consistently performed below chance levels. Participants performed especially poorly with close out-group laughs (from England, France, and USA), compared to in-group laughs (from the Netherlands) and distant out-group laughs (from Japan and Namibia), but in no case did performance exceed chance. The current study also asked whether variability in participants' exposure to other cultures would be linked to their performance. However, neither frequentist nor Bayesian analyses yielded support for this prediction either: no association was found between familiarity and group identification performance. It is worth acknowledging, however, that our measure of familiarity was indirect (number of foreign countries visited) and thus did not directly probe whether participants had visited the countries included in the current study.

These results support the findings of Sauter (2013), which showed that listeners were unable to judge group membership from non-verbal vocalizations, including laughter. However, previous research has found that perceivers can accurately judge group membership from facial expressions (Marsh et al., 2003) and language dialects (e.g., Kerswill and Williams, 2002). It is worth noting that task complexity may have played a role. In the study by Marsh and colleagues, in which participants differentiated Japanese-American and Japanese faces, participants performed a two-way forced choice (Marsh et al., 2003). In the current study, participants performed a six-way forced choice. The current set of results does not rule out the possibility that the accents in emotional expressions are sufficient to communicate whether a signal is from one's own, as opposed to another, group, but little beyond that.

Another possibility is that facial, but not vocal cues, provide group identity cues. This seems unlikely, given that spoken language is strongly connected to social identity (Giles and Viladot, 1994), and accents differ sufficiently between groups for others to use it for accurate group classification (Kerswill and Williams, 2002). Observers even preferentially rely on a speaker's linguistic dialect compared to their visual appearance (Rakić et al., 2011). However, this clear encoding of identity and group cues may be limited to volitionally produced vocalizations. Volitionally produced vocalizations involve more articulation and more complex coordination than the production of spontaneous laughter (Ruch and Ekman, 2001). A recent study found that speaker identity recognition was impaired for authentic, as compared to volitional, laughter (Lavan et al., 2016), which may reflect differences in vocal production between signals produced under reduced volitional control, such as spontaneous laughter, and volitional vocalizations, such as speech and volitional laughter. Future research could compare spontaneous and posed laughter directly to shed more light on this issue. It may thus be that the cues that listeners use to judge group identity

and individual identity are reduced in spontaneous non-verbal emotional vocalizations, including laughter.

The current results point to a potential boundary condition for motivational mechanisms of emotion perception. If perceivers cannot reliably judge group membership from non-verbal emotional vocalizations, this suggests that motivational mechanisms likely do not operate on these kinds of cues. As already shown by Sauter (2013), emotion recognition is superior for in-group non-verbal expressions. This indicates that vocalizations from different groups are not identical, and that these dialects in expressions are sufficient for the in-group advantage to occur in the absence of motivational factors. This does not mean that motivational mechanisms do not operate in cases where a perceiver is able to infer the group membership of the expresser, such as for example, for facial expressions.

AUTHOR CONTRIBUTIONS

MR analyzed the data. DS designed the study and supervised the analysis. Both authors interpreted the results, wrote the manuscript, and approved prior to submission. The authors agree to be accountable for the content of this work.

ACKNOWLEDGMENTS

The authors would like to thank Quest magazine (https://www.quest.nl) for allowing us to collect data through their website, Dora Matzke and Johnny van Doorn for their help with aspects of the Bayesian analysis, and all of the participants for taking part. The writing of this article was supported by grant 275-70-033 to DS from the Netherlands Organization for Scientific Research. This manuscript is based on the MSc internship of the first author.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02006/full#supplementary-material

REFERENCES

Baker, W., Eddington, D., and Nay, L. (2009). Dialect recognition: the effects of region of origin and amount of experience. *Am. Speech* 84, 48–71. doi: 10.1215/00031283-2009-004

Bryant, G. A., Fessler, D. M. T., Fusaroli, R., Clint, E., Aarøe, L., and Apicella, C. L. (2016). Detecting affiliation in colaughter across 24 societies. *Proc. Natl. Acad. Sci. U.S.A.* 113, 4682–4687. doi: 10.1073/pnas.1524993113

Byatt, G., and Rhodes, G. (2004). Identification of own-race and other-race faces: implications for the representation of race in face space. *Psychon. Bull. Rev.* 11, 735–741. doi: 10.3758/BF03196628

Cassidy, K. D., Quinn, K. A., and Humphreys, G. W. (2011). The influence of ingroup/outgroup categorization on same- and other-race face processing: the moderating role of inter- versus intra-racial context. *J. Exp. Soc. Psychol.* 47, 811–817. doi: 10.1016/j.jesp.2011.02.017

Clopper, C. G., and Pisoni, D. B. (2004a). Homebodies and army brats: some effects of early linguistic experience and residential history on dialect categorization. *Lang. Var. Change* 16, 31–48. doi: 10.1017/S0954394504161036

Clopper, C. G., and Pisoni, D. B. (2004b). Some acoustic cues for the perceptual categorization of American english regional dialects. *J. Phon.* 32, 111–140. doi: 10.1016/S0095-4470(03)00009-3

Crockford, C., Herbinger, I., Vigilant, L., and Boesch, C. (2004). Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology* 110, 1221–1243. doi: 10.1111/j.1439-0310.2004.00968.x

Dezecache, G., and Dunbar, R. (2004). Sharing the joke: the size of natural laughter groups. *Evol. Hum. Behav.* 33, 775–779. doi: 10.1016/j.evolhumbehav.2012.07.002

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press.

Elfenbein, H. A., and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychol. Bull.* 128, 203–235. doi: 10.1037/0033-2909.128.2.203

Elfenbein, H. A., Beaupré, M., Lévesque, M., and Hess, U. (2007). Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. *Emotion* 7, 131–146. doi: 10.1037/1528-3542.7.1.131

Giles, H., and Viladot, A. (1994). Ethnolinguistic differentiation in Catalonia. *Multilingua* 13, 301–312. doi: 10.1515/mult.1994.13.3.301

Hehman, E., Mania, E. W., and Gaertner, S. L. (2010). Where the division lies: common ingroup identity moderates the cross-race facial-recognition effect. *J. Exp. Soc. Psychol.* 46, 445–448. doi: 10.1016/j.jesp.2009.11.008

Herbinger, I., Papworth, S., Boesch, C., and Zuberbühler, K. (2009). Vocal, gestural and locomotor responses of wild chimpanzees to familiar and unfamiliar intruders: a playback study. *Anim. Behav.* 78, 1389–1396. doi: 10.1016/j.anbehav.2009.09.010

The JASP Team (2017). *JASP (Version 0.8.1.1) [Computer Software]*. Available online at: https://jasp-stats.org/

Jeffreys, H. (1961). *Theory of Probability, 3rd Edn*. Oxford: Oxford University Press.

Kerswill, P., and Williams, A. (2002). "Dialect recognition and speech community focusing in new and old towns in England: the effects of dialect levelling, demography and social networks," in *Handbook of Perceptual Dialectology, Vol. 2*, eds D. Long and D. R. Preston (Amsterdam: Benjamin), 173–205.

Lavan, N., Scott, S. K., and McGettigan, C. (2016). Impaired generalization of speaker identity in the perception of familiar and unfamiliar voices. *J. Exp. Psychol. Gen.* 145, 1604–1614. doi: 10.1037/xge0000223

Lee, M. D., and Wagenmakers, E.-J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. New York, NY: Cambridge University Press.

Marsh, A. A., Elfenbein, H. A., and Ambady, N. (2003). Nonverbal "Accents": cultural differences in facial expressions of emotion. *Psychol. Sci.* 14, 373–376. doi: 10.1111/1467-9280.24461

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ.* 15, 625–632. doi: 10.1007/s10459-010-9222-y

Owren, M. J., and Amoss, R. T. (2014). "Spontaneous human laughter," in *Handbook of Positive Emotions,* eds M. M. Tugade, M. N. Shiota, and L. D. Kirby (New York, NY: Guilford), 159–178.

Provine, R. R. (2004). Laughing, tickling, and the evolution of speech and self. *Curr. Dir. Psychol. Sci.* 13, 2015–2018. doi: 10.1111/j.0963-7214.2004.00311.x,

Rakić, T., Steffens, M. C., and Mummendy, A. (2011). Blinded by the Accent! The minor role of looks in ethnic categorization. *J. Pers. Soc. Psychol.* 100, 16–29. doi: 10.1037/a0021522

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rouder, J. N., Morey, R. D., Speckmann, P. L., and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *J. Math. Psychol.* 56, 356–374. doi: 10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckmann, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* 16, 225–237. doi: 10.3758/PBR.16.2.225

Ruch, W., and Ekman, P. (2001). "The expressive pattern of laughter," in *Emotion, Qualia, and Consciousness*, ed A. Kasniak (Tokyo: World Scientific Publisher), 426–443.

Sauter, D. A. (2013). The role of motivation and cultural dialects in the in-group advantage for emotional vocalizations. *Front. Psychol.* 4:814. doi: 10.3389/fpsyg.2013.00814

Sauter, D. A., Eisner, F., Ekman, P., and Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2408–2412. doi: 10.1073/pnas.0908239106

Scott, S. K., Lavan, N., Chen, S., and McGettigan, C. (2014). The social life of laughter. *Trends Cogn. Sci.* 18, 618–620. doi: 10.1016/j.tics.2014.09.002

Simon-Thomas, E. R., Keltner, D. J., Sauter, D. A., Sinicropi-Yao, L., and Abramso, A. (2009). The role of motivation and cultural dialects in the in-group advantage for emotional vocalizations. *Emotion* 9, 838–346. doi: 10.1037/a0017810

Smith, A. D. (1991). *National Identity*. London: Penguin Books.

Szameitat, D. P., Kreifelts, B., Alter, K., Szameitat, A. J., Sterr, A., Grodd, W., et al. (2010). It is not always tickling: distinct cerebral responses during perception of different laughter types. *Neuroimage* 53, 1264–1271. doi: 10.1016/j.neuroimage.2010.06.028

Thibault, P., Bourgeois, P., and Hess, U. (2006). The effect of group-identification on emotion recognition: the case of cats and basketball players. *J. Exp. Soc. Psychol.* 42, 676–683. doi: 10.1016/j.jesp.2005.10.006

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *J. Nonverbal Behav.* 17, 3–28. doi: 10.1007/BF00987006

Walton, J. H., and Orlikoff, R. F. (1994). Speaker race identification from acoustic cues in the vocal signal. *J. Speech Lang. Hear.* 37, 738–745. doi: 10.1044/jshr.3704.738

Wetzels, R., Grasman, R. P. P. P., and Wagenmakers, E.-J. (2010). An encompassing prior generalization of the Savage-Dickey density ratio test. *Comput. Stat. Data Anal.* 54, 2094–2102. doi: 10.1016/j.csda.2010.03.016

Wildgruber, D., Szameitat, D. P., Ethofer, T., Brück, C., Alter, K., Grodd, W., et al. (2013). Different types of laughter modulate connectivity within distinct parts of the laughter perception network. *PLoS ONE* 8:e63441. doi: 10.1371/journal.pone.0063441

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Ritter and Sauter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 January 2018
doi: 10.3389/fpsyg.2017.02342

Social Context Disambiguates the Interpretation of Laughter

William Curran[1], Gary J. McKeown[1], Magdalena Rychlowska[1], Elisabeth André[2], Johannes Wagner[2] and Florian Lingenfelser[2]*

[1] School of Psychology, Queen's University Belfast, Belfast, United Kingdom, [2] Human-Centered Multimedia, Institut für Informatik Universität Augsburg, Augsburg, Germany

Despite being a pan-cultural phenomenon, laughter is arguably the least understood behaviour deployed in social interaction. As well as being a response to humour, it has other important functions including promoting social affiliation, developing cooperation and regulating competitive behaviours. This multi-functional feature of laughter marks it as an adaptive behaviour central to facilitating social cohesion. However, it is not clear how laughter achieves this social cohesion. We consider two approaches to understanding how laughter facilitates social cohesion – the 'representational' approach and the 'affect-induction' approach. The representational approach suggests that laughter conveys information about the expresser's emotional state, and the listener decodes this information to gain knowledge about the laugher's felt state. The affect-induction approach views laughter as a tool to influence the affective state of listeners. We describe a modified version of the affect-induction approach, in which laughter is combined with additional factors – including social context, verbal information, other social signals and knowledge of the listener's emotional state – to influence an interaction partner. This view asserts that laughter by itself is ambiguous: the same laughter may induce positive or negative affect in a listener, with the outcome determined by the combination of these additional factors. Here we describe two experiments exploring which of these approaches accurately describes laughter. Participants judged the genuineness of audio–video recordings of social interactions containing laughter. Unknown to the participants the recordings contained either the original laughter or replacement laughter from a different part of the interaction. When replacement laughter was matched for intensity, genuineness judgements were similar to judgements of the original unmodified recordings. When replacement laughter was not matched for intensity, genuineness judgements were generally significantly lower. These results support the affect-induction view of laughter by suggesting that laughter is inherently underdetermined and ambiguous, and that its interpretation is determined by the context in which it occurs.

Keywords: laughter, social interaction, social context, laughter interpretation, non-verbal communication

OPEN ACCESS

Edited by:
Willibald Ruch,
University of Zurich, Switzerland

Reviewed by:
Jill Ann Jacobson,
Queen's University, Canada
Jo-Anne Bachorowski,
Vanderbilt University, United States

****Correspondence:***
William Curran
w.curran@qub.ac.uk

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *30 August 2017*
Accepted: *22 December 2017*
Published: *12 January 2018*

Citation:
Curran W, McKeown GJ, Rychlowska M, André E, Wagner J and Lingenfelser F (2018) Social Context Disambiguates the Interpretation of Laughter. Front. Psychol. 8:2342. doi: 10.3389/fpsyg.2017.02342

INTRODUCTION

Because of its ubiquitous nature laughter has become a recent focus of research across a range of scientific disciplines. The thesis that it is an evolutionarily ancient behaviour preceding spoken language is supported by reports of laughter-like behaviour in non-human primates (Davila Ross et al., 2009, Davila-Ross et al., 2011) and of similarities in acoustic elements of laughter in humans

and play vocalisations of other primates (Vettin and Todt, 2005). As in human infants (Rothbart, 1973) laughter-like behaviour in chimpanzees prolongs play actions (Matsusaka, 2004), suggesting that laughter is an important tool for promoting social affiliation and developing cooperative and competitive behaviours (Davila-Ross et al., 2011). Thus from an evolutionary perspective laughter can be viewed as a key adaptive behaviour because of its facilitative effect on social cohesion.

Just as speech follows a set of rules so too, it seems, does laughter, with conversation analysis uncovering a number of rules of laughter behaviour (Glenn, 2003; Holt, 2010, 2011). For example in dyadic conversations the speaker is more likely than the listener to laugh first, while in group conversations the listeners are more likely to laugh first (Glenn, 2003). Laughter also has a regulatory function in conversations by serving as a turn-taking cue or signalling that the speaker may be approaching a transition point in his/her utterance (O'Donnell Trujillo and Adams, 1983; Holt, 2010; Bonin et al., 2012).

Neuropsychological research has revealed the existence of two partially dissociable neural pathways underlying two different types of laughter – spontaneous and volitional. One pathway is emotionally driven and involuntary, arising in subcortical, limbic, and brainstem areas and culminating in a "laughter-coordinating" centre in the dorsal upper pons; the other is a voluntary motor pattern that originates in frontal premotor areas and directly influences the motor cortex. Damage to the former pathway inhibits the production of spontaneous (but not volitional) laughter production, and damage to the latter pathway inhibits volitional (but not spontaneous) laughter production (Wild et al., 2003). Spontaneous laughter and volitional laughter do not only arise from separate neural systems but also appear to be processed in distinct brain regions, as evidenced by increased amPFC and anterior cingulate cortex activity when participants listen to volitional laughter as opposed to spontaneous laughter (McGettigan et al., 2013). McGettigan et al. (2013) suggest that volitional laughter induces a stronger engagement of mentalising processes, and postulate that this is indicative of attempts to assess the emotional state and intentions of the laugher.

The conclusions drawn by McGettigan et al. (2013) speak to a long-standing debate (Fridlund, 1994; Parkinson, 2005; Feldman Barrett, 2006, 2017) on the function of social signals and emotions – do they merely indicate the expresser's felt state or can they function as active socio-communicative instruments? Laughter, as one of our most common non-verbal social signals, is well placed to illuminate this debate. According to the mainstream view, laughter conveys information about the laugher's underlying emotional state and listeners can decode this information to gain knowledge about other people's feelings and motivations. This 'representational' approach, where signals are referential and "about" something, often relies on the "conduit metaphor" (Reddy, 1979; Rendall et al., 2009), according to which signals transmit information from and about their sender and convey it to a receiver who then decodes the message. The 'representational' view of laughter fits with our everyday use of various types of laughter, such as joyful laughter, schadenfreude laughter, taunting laughter, and embarrassed laughter: the laughter indicates the expresser's internal state and serves as a "readout" of the laugher's current emotions. If the function of social signals is to represent an internal felt state then discrete representational coding provides an efficient indication of felt state. These internal states should be relatively easy to distinguish and there would be distinct laughter signals that distinguish each discrete "natural kind" of felt state (Feldman Barrett, 2006). Coming back to the examples above, distinct kinds of laughter should accompany feelings of joy, embarrassment, or schadenfreude. However, the lack of evidence for affect-specific laughter types challenges the notion that laughter is used to communicate underlying emotional states. It may be that in these "joint state" cases natural kinds are not represented, that is there is no discrete or specific "nervous laughter" or "schadenfreude laughter," but there is an additive combination of social signals. For example, some signals may be to do with nervousness and the laughter signal may have a different communicative function. An alternative theoretical view is that, rather than passively transmitting information about the laugher's emotional state, laughter is used to influence the affective states of listeners (Owren and Bachorowski, 2003). To differentiate between the representational and affect-induction views of laughter, Owren and Bachorowski use the analogy of a crying baby. According to the representational account, the crying merely delivers appropriate encoded information about the unpleasant feelings that the baby is experiencing and should stop when the information is provided. An affect-induction view, on the other hand, holds that the acoustic and visual qualities of crying induce negative affect in a listener that persists until the problem is resolved. Similarly, in the case of primate vocalisations, "the primary function of calling is to influence listener attention, arousal, and emotion rather than to transmit information." Adult human communication, however, is more complex than a baby's cry or primate calls and is often accompanied by verbal messages.

According to Owren and Bachorowski (2003), laughter is also a form of affect-induction communication. As evidence against the representational view of this behaviour, they point to its generalizability and lack of specificity. Namely, the same laugh may be used to induce positive or negative affect in a listener, and the interpretation of the same "laugh episode" will be determined by a number of factors, including the behaviours the laughter accompanies, the relationship between the laugher and listener, or the listener's emotional state when hearing the laughter. Thus, while in one context a listener may experience a laugh as derisory, the same laugh may become a positive experience in a different context–as Papousek et al. (2014) have shown.

The key contrasts between these two approaches are: (1) the nature of the motivation for the behaviour – in the representational case the laugher passively indicates their felt state, in the affect-induction case laughter is used to influence the receiver's affect; (2) the level of ambiguity – in the representational case different laughter types would facilitate distinguishing between felt states, in the affect-induction case laughter would be used in conjunction with other social signals and contextual factors to induce a desired affect. Furthermore, the laughter would be ambiguous in the absence of social context.

Here, we propose a modified version of the affect-induction approach. From the affect-induction perspective affect is induced through the combination of laughter with additional factors that dynamically unfold throughout the course of a social interaction – such as verbal information, social context, or knowledge of the listener's emotional state. These factors influence the receiver and their accurate understanding ensures that the laughter is interpreted in accordance with the expresser's socio-communicative goals. In this view, a laugh by itself is an underdetermined and ambiguous social signal. In other words, hearing or viewing somebody laughing without additional information does not provide enough information to be sure of their emotional state. Our modified version of the affect-induction approach takes into account the role of intensity in laughter communication.

Intensity is an important communicative component of many social signals and emotional expressions (Banse and Scherer, 1996; Bänziger and Scherer, 2005; Biele and Grabowska, 2006), and also appears to play an important role in laughter (Darwin, 1872; van Hooff, 1972; Preuschoft and van Hooff, 1997, McKeown et al., in preparation). It adds a level of complexity to our distinctions between representational views and affect-induction views. Given that voiced laughter produces stronger affect-related responses in a listener relative to unvoiced laughter (Bachorowski and Owren, 2001), it follows from the perspective of the affect-induction approach that different laughter intensities should vary in the extent to which they induce affect in the listener; the representational model, on the other hand, would predict that differing laughter intensities would indicate different levels of a given felt emotion. It is probable that laughter intensity reflects a complex interplay between felt emotion and contextual influences on affect induction, in the same way that emotional facial expressions are influenced by both these factors (Fridlund, 1994). However, apart from components of the laughter signal that make a laugh more or less intense, we argue (as others have argued; e.g., Russell et al., 2003) that there are no morphological or acoustic markers of laughter that contain meaning in a representational sense. In other words, there is not a one-to-one relationship between a felt emotion and laughter produced while experiencing that emotion. Take the following scenario as an example. Two people, James and Robin, are having an intense argument while being observed by a neutral group. James makes a witty comment that highlights a central flaw in Robin's argument. The people observing the argument laugh in response to the witty comment. Both participants in the argument will hear the same laughter, but are likely to interpret it differently; James will interpret the laughter as humorous and an appreciation of his wit while Robin is likely to interpret it as derisory laughter. In this scenario identical laughter is taken to signal two very different emotions – humour and derision. This would not occur if different emotions were represented by laughter with distinct acoustic markers. Rather, we propose that laughter is intrinsically underdetermined and ambiguous, and a listener's interpretation of laughter is determined by factors such as the context in which it occurs. The same view has been taken by Owren and Bachorowski's affect-induction approach. However, the modified affect-induction model differs from the original model in one key respect. While both models would agree that its inherent ambiguity allows laughter from different interactions to be wholly interchangeable without affecting apparent genuineness of an interaction, the modified version proposes that this interchangeability is restricted to laughter of similar intensity.

Using a novel experimental approach to test our hypothesis, we reasoned that participants' judgements of social interactions should be unaffected if the laugh response in an original dyadic interaction is replaced with a similar intensity laugh response from a different part of the same interaction. In contrast, if a high intensity laugh response is replaced with a low intensity laugh response, or vice versa, there should be a measurable reduction in 'genuineness' ratings. Interchangeability of laughter would be taken as evidence supportive of the affect-induction model's central tenet that laughter is an inherently ambiguous signal; and evidence of intensity-specific interchangeability would support the modified affection-induction model's incorporation of intensity as an important factor in interpreting laughter. If, however, exchanging laughter always results in an interaction seeming less genuine, this will be taken as support for the representational model.

MATERIALS AND METHODS

Overview

Here we report on two experiments, in which participants viewed video sequences displaying two persons. Each sequence involved a 'listener' laughing in response to something a 'story-teller' said. Participants then judged how real or genuine the interaction was, that is how confident they were that the interaction actually took place. The recorded interactions contained either high intensity or low intensity laughter. To differentiate the story-teller from the laugher we refer to the story-teller as producing a high or low intensity "laughable context," and the laugher produces either a high or low intensity "laugh response."

Stimuli

Stimuli were generated from interactions created as part of the ILHAIRE laughter database (McKeown et al., 2015) using a naturalistic story telling task. The Social Signal Interpretation framework (Wagner et al., 2013) was used to capture video and audio information of groups of two or more people. The task was designed to exert minimal influence on the behaviours of the interlocutors as they conversed with one another, while allowing the synchronised capture of high quality audio and video material.

Intensity Selection

Laugh stimuli were extracted from the original interactions and rated by participants along a number of dimensions including laugh intensity and humour, with participants recruited using Amazon's Mechanical Turk. An informed consent form explaining the study's procedure and the experimenter's contact details was placed at the very beginning of the ratings form. MTurk participants were informed that their identity would remain confidential and that they could withdraw from the

experiment at any time by simply logging out before completing the rating exercise. An incomplete data set from a participant was interpreted as the participant deciding to withdraw from the experiment, and the relevant data were destroyed. Analysis of 9421 ratings of 870 laugh episodes revealed a strong relationship between the intensity of a laugh and how much people judged it to be related to humour (see McKeown and Curran, 2015, for more detail).

The stimuli used in the following experiment are taken from two conversational partners participating in an interaction between three people that lasted for 70 min. Since both social context and intensity affect emotion perception (Fridlund, 1994), for this initial experiment we only used the recordings of two conversational partners who were both male, who shared the same cultural background, and who were friends. According to Owren and Bachorowski (2003) male friend pairs should produce high rates of laughter that are acoustically extreme in both pitch and duration. The naturalistic form of data gathering has the effect of producing natural laughter with greater ecological validity; however, it also means that many of the laugh instances must be excluded from the experiment as they include other verbal cues (e.g., speech) and non-verbal cues (e.g., not looking at the speaker, covering the face with a hand) not related to the effect of interest.

The laughter selected for intensity rating was laughter in which there was an unobstructed frontal view of the laugher's face, the laugher was looking at the speaker, and there was no speech during laughter. Laughter was rated for intensity using the question "Can you rate the intensity of the laugh on a 10 point scale, from 1 no intensity to 10 maximum intensity?" This rating strategy assumes the laugh rater has a degree of expertise in laughter through being a lifetime observer of laughter and consequently minimal instructions are provided to avoid leading the rater into a particular interpretation of the concept of laugh intensity. Laughter stimuli assigned to the lowest quartile of intensity ratings were designated as 'low intensity' laughter, and laughs assigned to the highest quartile were designated as 'high intensity' laughter. The corresponding laughable contexts were not independently rated, but are termed high and low intensity laughable contexts by virtue of the fact that they resulted in high or low laugh responses. After the exclusion of laughter instances that contained verbal and non-verbal cues not related to the effect of interest, 8 laugh responses combined with their "laughable" context remained – 4 high intensity and 4 low intensity. These were used to generate the experimental stimuli.

Stimulus Generation

Each laugh and laughable context was placed alongside each other on a computer screen to produce a reconstruction of the interaction (see **Figure 1**). We generated stimuli for six conditions: two control conditions containing both the high and low intensity original interactions; two same intensity conditions, one containing high intensity laughable contexts with swapped high intensity laugh responses and another containing low intensity laughable contexts with swapped low intensity laugh responses; and two opposite intensity conditions, one containing high intensity laughable contexts with swapped low intensity laugh responses and another containing low intensity laughable contexts with swapped high intensity laugh responses. The listener's video stream was frozen at a frame containing a neutral facial expression up until the point where the laugh response began; this was to avoid unwanted social cues interfering with the participants' perception of the laugh.

FIGURE 1 | Screenshot of the interaction stimuli. The story teller and listener were positioned on the right and left, respectively. Each interaction lasted approximately ten seconds, during which time the unfolding story led up to the laughter event. The listener's audio–visual stream was frozen at a frame containing a neutral facial expression up until the point where the laugh response began. Faces were not blurred during the actual experiment.

In each experiment we adopted a 2 (laugh response intensity) × 3 (laughable context) design. Laugh responses had two levels, high or low intensity. Laughable context had three levels, one control condition (original video) and two experimental conditions (same intensity, and opposite intensity). In the 'same intensity' condition the listener's audio–visual stream was replaced with a recording of the same listener producing a laugh response with the same intensity, but taken from a different point in the conversation. In the 'opposite intensity' condition the listener's laugh response was replaced with a laugh response by the same listener but with the opposite intensity; that is, low and high intensity laugh responses were replaced with high and low intensity laugh responses, respectively. In the control condition participants viewed the original story-teller/listener interaction.

The dependent variable is the level of confidence that the interaction is genuine, i.e., that the interaction actually took place. The exact question is "Can you provide a rating of your confidence level between 0 (no confidence at all) and 10 (highly confident) that this is a genuine interaction?

We adopt the statistical recommendations of Cumming (2012, 2014) and Cumming and Calin-Jageman (2017), using point estimates with confidence intervals and effect sizes to convey precision and the magnitude of the experimental effects. This approach does not change the fundamental frequentist philosophy in the statistics but alters the emphasis toward presenting effect sizes and away from point estimates of *p*-values through the use of confidence intervals. In addition, we have created estimated values for our hypotheses, these are arbitrary estimates in absolute terms but the pattern of results is based on the reasoning we have outlined. Our scale for the assessment

of genuine interactions runs from 0 to 10. Aware of the central tendency and range restriction errors outlined by Saal et al. (1980) we assume that even when participants strongly believe that an interaction is genuine they will be reluctant to suggest a rating of 10; we therefore place our estimate of belief in a genuine interaction at the upper quartile, 7.5, and our estimate of a not genuine interaction at the lower quartile of 2.5. For the two control conditions, which are genuine interactions, we estimate they will both be viewed as genuine: thus, we predict that participants will give a maximum genuineness score of 7.5 for original recordings regardless of intensity of the laughable context/laugh response. Similarly, we predict that maximum genuineness scores will be given in the interchanged conditions in which a laugh is replaced with a laugh of the same intensity. However, where the interchange involves swapping laughs of different intensity (i.e., the replacement laugh does not match the laughable context), we predict that such interactions will be seen as not genuine and will be assigned the lowest genuineness score of 2.5. **Figure 2** displays these estimates in a graphical form.

EXPERIMENT 1

Method

Participants

One hundred and one participants (40 women, 61 men, mean age = 33.16 years, age range = 20–68 years) were recruited using Amazon's Mechanical Turk, a crowdsourcing website which produces high quality data that are at least as reliable as those obtained through traditional methods (Casler et al., 2013; Paolacci and Chandler, 2014).

Materials

The general stimuli generation has already been outlined. In this experiment, we generated stimuli from two of the four high intensity laughable contexts and two of the four low intensity laughable contexts. These laughable contexts were paired with their original laugh responses to create two stimuli for the high intensity control condition and two stimuli for the low intensity control condition. Two different high intensity laugh responses were randomly selected and paired with the two high intensity laughable contexts to create two stimuli for the interchanged same-high-intensity condition. Two different low intensity laugh responses were randomly selected and paired with the two low intensity laughable contexts to create two stimuli for the interchanged same-low-intensity condition. Two different high intensity laugh responses were randomly selected and paired with the two low intensity laughable contexts to create two stimuli for the interchanged opposite-high-intensity condition. Finally, two different low intensity laugh responses were randomly selected and paired with the two high intensity laughable contexts to create two stimuli for the interchanged opposite-low-intensity condition. This gave a total of 12 stimulus clips in 6 conditions, 2 in each condition.

Procedure

All participants viewed all 12 clips, and provided ratings of level of confidence that the interaction was genuine for each stimulus.

Results

The general pattern of the results (**Figure 3**) show that participants' genuineness ratings were unaffected when the listener's laugh was replaced with a same-intensity laugh from

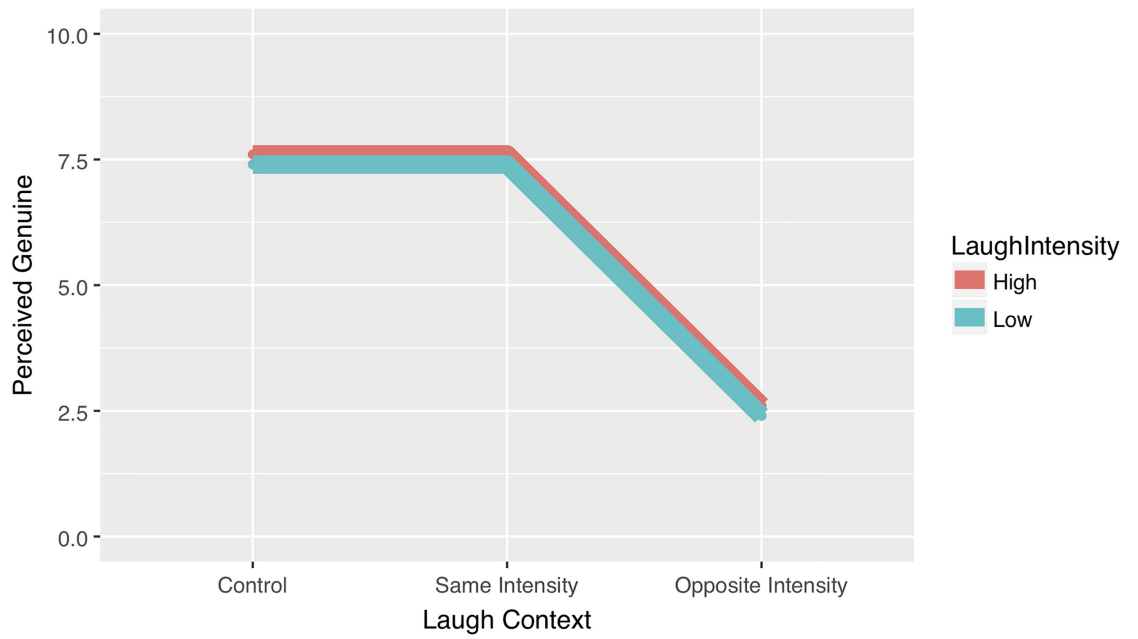


FIGURE 2 | Hypothesised pattern of results for Experiment 1: Interchanged laughter leads to perceptions of genuine interactions when intensity is the same but not when intensity is different.

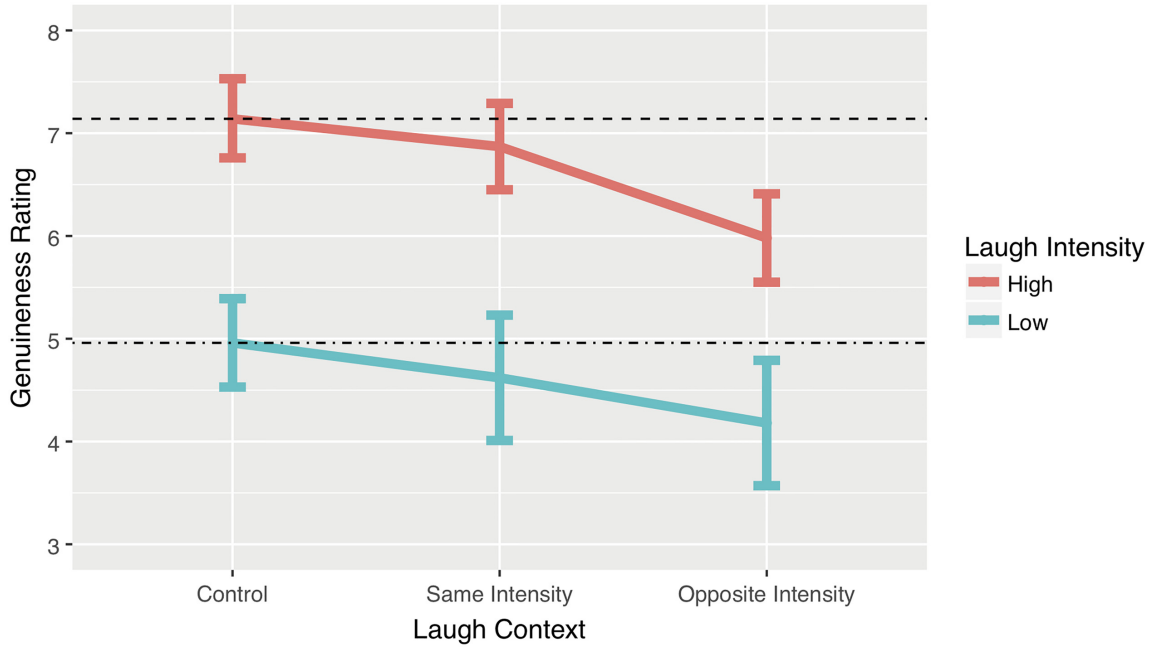


FIGURE 3 | Point estimates derived from the model for each of the conditions. Red and blue lines plot genuineness ratings of interactions containing high and low intensity laughter, respectively. Although interactions containing low intensity laughter were consistently judged less genuine than those containing high intensity laughter, both sets of data reveal a similar trend. In both cases there was no significant change in genuineness ratings when laughter in an original interaction was replaced with similar intensity laughter. However, replacements with laughter of the opposite intensity result in a measurable reduction in genuineness ratings. Error bars represent 95% Confidence Intervals.

TABLE 1 | Table of the fixed effects factors for the multi-level model of laugh responses for Experiment 1.

	Condition	*b*	*SE*	*t*	95% CI	Point estimate	*M*	*SD*
Intercept	Control High intensity	7.14	0.20	36.10	6.76, 7.53	7.14	7.14	2.17
High context	High context Same intensity	−0.28	0.22	−1.26	−0.70, 0.15	6.87	6.87	2.22
Low context	Low context Opposite intensity	−1.16	0.22	−5.29	−1.59, −0.73	5.98	5.98	2.36
Low intensity	Control Low intensity	−2.18	0.22	−9.93	−2.61, −1.75	4.96	4.96	2.68
LowC*LowI	Low context Same intensity	0.83	0.31	2.66	0.22, 1.44	4.62	4.62	2.71
HighC*LowI	High context Opposite intensity	−0.5	0.31	−1.61	−1.11, 0.11	4.18	4.18	2.3

a different point in the conversation. However, replacing a laugh with an opposite-intensity laugh resulted in a measurable reduction in genuineness ratings. An unexpected finding was that real interactions containing low intensity laughter (**Figure 3**, lower line) were consistently judged as less genuine than real interactions containing high intensity laughter (**Figure 3**, upper line). We address the implications of this finding in the discussion.

The analyses were performed using multi-level models to generate point estimates of the mean with confidence intervals. The multi-level approach accounts for the dependency in the data due to using the same participants to rate more than one video clip, and avoids underestimation of the standard errors (Quené and van den Bergh, 2004; McKeown and Sneddon, 2014). Point estimates are labelled M_{est}, and arithmetic means and standard deviations are provided in **Table 1**. The R (R Core Team, 2017) package lme4 (Bates et al., 2015) was used for the multilevel models and generation of profile confidence intervals.

Genuineness scores in the high intensity control condition were very similar to our hypothesised estimates (M_{est} = 7.14). The lowest level of reported genuineness (M_{est} = 4.18, for high intensity laughter inserted into a low intensity context) was considerably more than the hypothesised 2.5, the lowest estimate of genuineness predicted. The first important difference with our hypothesised estimates is that the low intensity control condition was judged as less genuine (M_{est} = 4.96, 95% CI [4.53, 5.39]) than the high intensity control condition (M_{est} = 7.14, 95% CI [6.76, 7.53]), even though both conditions involved genuine interactions. This suggests that something in the nature of the low intensity laugh responses and laughable contexts results in the overall interaction being judged as less genuine than interactions that contain high intensity laugh responses and laughable contexts. As a result, we will treat the low intensity laugh results and high intensity laugh results independently. We, therefore, use the ratings for the control interactions as our reference point estimate in the models against which the

experimental conditions can be compared. Participant variance was modelled as a random parameter using a random intercept multilevel model (participant variance = 1.51, *SD* = 1.23; residual variance = 4.89, *SD* = 2.21). Fixed effect statistics are provided in **Table 1**.

High Intensity Laughter

Same context

When laughter from high intensity laughable contexts were swapped for laughter taken from other high intensity laughable contexts to produce stimuli of interactions that never occurred, we found that genuineness ratings were similar to those in the control condition (M_{est} = 6.87, 95% CI [6.45, 7.29]). The model *b* coefficient provides the best effect size in the units of the study between the high intensity control condition and the same intensity condition, representing a reduction of perceived genuineness of the interaction by 0.28 on the 0–10 scale. Given the difficulty of choosing a standardised effect size measure for local effects within mixed-effects regression models (Selya et al., 2012), we adopt a technique used by Friedmann et al. (2008) where mean difference scores are calculated from the model generated point estimates and the control group standard deviation is used to provide a measure of Cohen's *d*; here *d* = 0.13. The Common Language Effect Size (CLES) (McGraw and Wong, 1992; Lakens, 2013) indicates that, after controlling for individual differences, the likelihood that a person rates the control interaction stimuli as more genuine than the swapped laugh stimuli is 54% (50% corresponds to no difference). Thus, interchanging high intensity laughs has little or no effect on ratings of the genuineness of an interaction.

Different context

In contrast to the same-context condition, when high intensity laugh responses were inserted into low intensity laughable contexts, the mean genuineness ratings were considerably lower (M_{est} = 5.98, 95% CI [5.55, 6.41]). A *b* coefficient of −1.16 provides a study unit effect size estimate of the difference between the high-intensity control condition and the opposite-intensity condition–in this case a low-intensity laughable context with a high-intensity laugh. It represents a reduction of perceived genuineness of the interaction by 1.16 on the 0–10 scale. Cohen's *d* (0.54) and the CLES indicate that, after controlling for individual differences, the likelihood that a person rates the swapped laugh stimuli as less genuine than the control interaction stimuli is 65%. In terms of Cohen's *d* rule of thumb this would be a medium effect size (Cohen, 1988).

Low Intensity Laughter

Same context

When laughter in low intensity laughable contexts is replaced with laughter taken from other low intensity laughable contexts we find that genuineness ratings are once again similar to those in the control condition (M_{est} = 4.62, 95% CI [4.01, 5.23]). A *b* coefficient of 0.83 corresponds to a reduction of perceived genuineness of the interaction by 0.34 on the 0–10 scale. Cohen's *d* (0.13) and the CLES indicate that, after controlling for individual differences, the likelihood that a person rates the control interaction stimuli as more genuine than the swapped laugh stimuli is 54%. Thus, interchanging a low intensity laugh has little or no effect on ratings of the genuineness of an interaction.

Different context

When low intensity laugh responses were inserted into high intensity laughable contexts the mean genuineness ratings were the lowest observed in this experiment (M_{est} = 4.18, 95% CI [3.57, 4.79]). A *b* coefficient of −0.5 corresponds to a reduction of the perceived genuineness of the interaction by 0.78 on the 0–10 scale. Cohen's *d* (0.29) and the CLES indicate that the likelihood a person rates the control interaction stimuli as more genuine than the swapped laugh stimuli is 58%. In terms of Cohen's *d* rule of thumb this would be a small effect.

Acknowledging the historical context of the discipline and the role of null hypothesis significance testing within this context, and due to the importance of the issues raised by Gelman and Stern (2006) we also present the results of this analysis using a 2 × 3 ANOVA. We present the main effects and interaction effect of the ANOVA but encourage researchers to give more prominent attention toward the simple main effects using the multi-level model generated point estimates and confidence interval approach for detailed analysis with respect to theoretical concerns. There is a significant main effect of laughable context $F(2,1206) = 9.38$, $p < 0.001$, $\eta^2 = 0.01$. There is also a significant main effect of laugh intensity $F(1,1206) = 204.04$, $p < 0.001$, $\eta^2 = 0.14$. Finally, there is a significant interaction of laughable context and laugh intensity $F(2,1206) = 7.1$, $p < 0.001$, $\eta^2 = 0.01$.

Discussion

The results of Experiment 1 provide support for our hypothesis that laughter of the same intensity can be interchanged with other laughs of similar intensity without affecting the apparent genuineness of the interaction. Where the level of laughter intensity does not match the context into which it is inserted, effects are observed – a medium effect size in the case of high intensity laugh response and a small effect size in the case of low intensity laugh responses. The finding that laughs of similar intensity are wholly interchangeable without affecting an interaction's perceived genuineness provides a proof of concept for our hypothesis that laughter is inherently ambiguous; as such this finding poses a challenge for the representational model of laughter and is consistent with the affect-induction model.

An additional important finding is the overall reduction in genuineness associated with low intensity laughs. The genuine low-intensity situation was judged to be less genuine than the worst case condition that contained a high intensity laugh response. It appears that even when strong laughter occurs with no expectation cues, these interactions are deemed to be more genuine than real interactions that contain low intensity laughter.

There are some limitations to this study. We only used four of the eight actual laugh contexts selected, and we cannot rule out the possibility that the present findings were due to specific features of the contexts displayed in the stimuli. Another caveat is that all the stimuli were judged by each participant in this experiment, allowing for the possibility that judgements

of a laugh were made relative to responses to other laughs paired with the same context. These limitations are addressed in Experiment 2.

EXPERIMENT 2

Experiment 2 was a direct replication of Experiment 1, but with additional manipulations that address the above limitations. The number of stimuli used was increased to the maximum possible given the selection of eight usable contexts. In addition, we also wished to remove the possibility that responses to previous stimuli combinations were interfering with judgements being made about the genuineness of a given interaction. We excluded this possibility by ensuring that each participant saw each laugh context only once. We hypothesised that the results would follow the same pattern observed in Experiment 1.

Method

Participants

As we were not getting more than one rating per context from the participants in this experiment we increased the sample size and recruited 404 participants (153 women, 251 men, mean age = 33.16 years, age range = 20–68 years) via Amazon's Mechanical Turk.

Stimulus Generation

The stimuli were created using the same eight laughs selected for use in Experiment 1. The main difference was that on this occasion we created all possible stimulus combinations with the eight laugh contexts. This gave a total of 64 stimulus clips. The same 2 (laugh response intensity) × 3 (laughable context) design was used in the presentation of the stimuli, giving six conditions. There were four original high intensity laughter clips in the high intensity control condition; 4 original low intensity clips in the low intensity control condition; 12 high intensity clips in the high-same-intensity condition; 12 low intensity clips in the low-same-intensity condition; 16 low intensity laughter clips in the high-opposite-intensity condition; 16 high intensity clips in the low-opposite-intensity condition.

Procedure

The procedure was the same as in Experiment 1 except on this occasion we showed eight videos to each participant and ensured that no laughter-inducing context was repeated in a given experimental run. We randomly selected one of the laughter stimuli for each laugh context for presentation to the participants. The condition cells are necessarily imbalanced due to the nature of the stimulus generation; there are only eight genuine interactions to create the control conditions, only 24 stimuli can be used in the same intensity conditions as the control conditions cannot be used, and there are 32 possible combinations for the opposite intensity conditions. These differences in number of stimuli and number of participants across the cell sizes are largely accommodated by the use of multi-level models, which are more robust to unequal cell sizes than repeated measure ANOVA models.

Results

The absolute numerical estimates for the conditions were slightly different from Experiment 1, but the pattern was the same (see **Figure 4**): in the control condition using low intensity laughter genuineness was again rated lower (M_{est} = 5.08, 95% CI [4.59, 5.57]) than in the control condition using high intensity

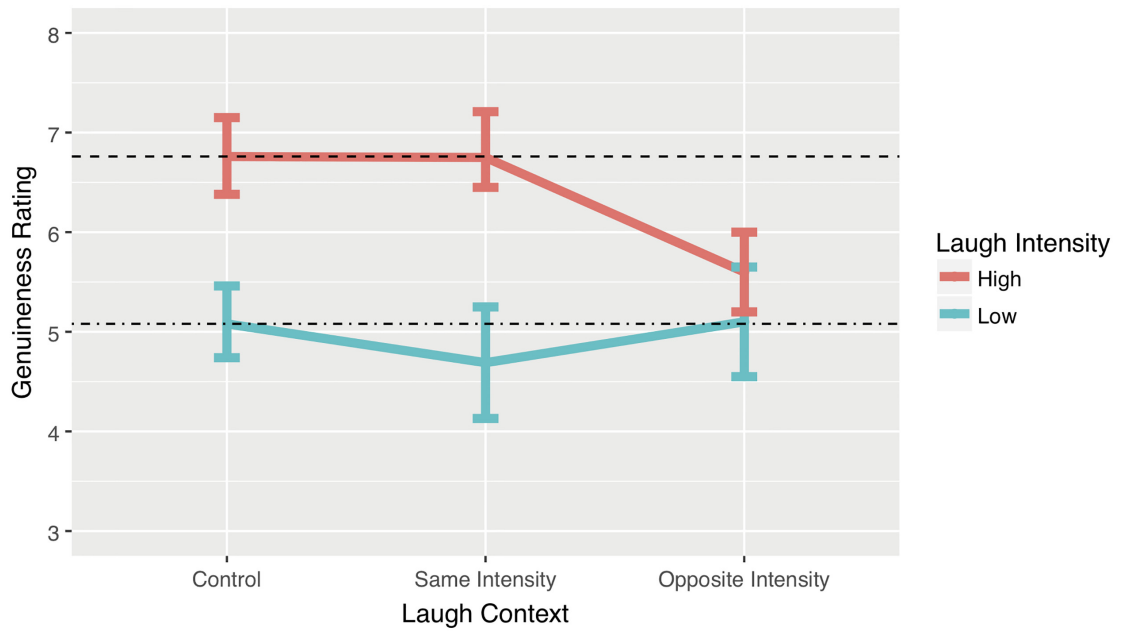


FIGURE 4 | Point estimates derived from the model for each of the conditions. Error bars represent 95% Confidence Intervals.

TABLE 2 | Table of the fixed effects factors for the multi-level model of laugh responses for Experiment 2.

	Condition	*b*	*SE*	*t*	95% CI	Point estimate	*M*	*SD*
Intercept	Control High intensity	6.76	0.20	34.67	6.38, 7.15	6.76	6.76	2.17
High context	High context Same intensity	−0.02	0.21	−0.13	−0.44, 0.38	6.75	6.7	2.39
Low context	Low context Opposite intensity	−1.16	0.20	−5.69	−1.56, −0.76	5.60	5.57	2.94
Low intensity	Control Low intensity	−1.69	0.25	−6.74	−2.18, −1.20	5.08	5.1	2.8
LowC*LowI	Low context Same intensity	0.78	0.29	2.74	0.22, 1.34	4.69	4.72	2.81
HighC*LowI	High context Opposite intensity	0.05	0.28	0.19	−0.50, 0.61	5.1	5.13	2.86

laughter (M_{est} = 6.76, 95% CI [6.38, 7.15]). Once again, we used a multi-level model to generate point estimates of the mean with profile confidence intervals. Participant variance was modelled as a random parameter using a random intercept multilevel model (participant variance = 1.69, *SD* = 1.3; residual variance = 5.93, *SD* = 2.43). Fixed effect statistics are provided in **Table 2**.

High Intensity Laughter

Same context

When laugh responses from high intensity laughable contexts are swapped for laugh responses taken from other high intensity laughable contexts we find that genuineness ratings are almost identical to those in the control condition (M_{est} = 6.75, 95% CI [6.45, 7.21]). A *b* coefficient of −0.02 represents a reduction of perceived genuineness of the interaction relative to the control condition by 0.02 on the 0–10 scale. Cohen's *d* (0.01) and the CLES indicate that the likelihood that a person rates the control interaction stimuli as more genuine than the swapped laugh stimuli is 50%; this suggests that interchanging high intensity laughs has no effect on ratings of the genuineness of an interaction.

Different context

When high intensity laughter was inserted into low intensity laughable contexts the mean genuineness ratings were considerably lower (M_{est} = 5.60, 95% CI [5.2, 6]). A *b* coefficient of −1.16 represents a reduction of perceived genuineness of the interaction by 1.16 on the 0–10 scale. Cohen's *d* (0.47) and the CLES indicate that the likelihood a person rates the swapped laugh stimuli as less genuine than the control interaction stimuli is 63%. In terms of Cohen's *d* rule of thumb this would be a medium effect size.

Low Intensity Laughter

Same context

When laughter from low intensity laughable contexts is replaced with laughter from other low intensity laughable contexts we obtain a pattern of ratings (M_{est} = 4.69, 95% CI [4.13, 5.25]) similar to experiment 1. A *b* coefficient of 0.78 represents a difference in perceived genuineness by 0.78 between the two conditions. Cohen's *d* (0.14) and the CLES indicate that the likelihood that a person rates the swapped laugh stimuli as less genuine than the control interaction stimuli is 54%. Thus, interchanging low intensity laughs has little or no effect on ratings of the genuineness of an interaction.

Different context

On this occasion when low intensity laughter was inserted into high intensity laughable contexts the mean genuineness ratings (M_{est} = 5.1, 95% CI [4.55, 5.65]) were at similar levels to the control reference condition. A *b* coefficient of 0.05 indicates an increase in the perceived genuineness of the interaction by 0.05 on the 0–10 scale. Cohen's *d* (0.02) and the CLES indicate that the likelihood a person rates the swapped laugh as less genuine than the control interaction stimuli is 51%. Interchanging a low intensity laugh into a high intensity laughable context has no effect on ratings of the genuineness of an interaction. Interchanging a low intensity laugh into a high intensity laughable context has no effect on ratings of the genuineness of an interaction.

Once again, we present the main effects and interaction effect of the ANOVA but encourage researcher to give more prominent attention toward the simple main effects using the multi-level model generated point estimates and confidence interval approach for detailed analysis with respect to theoretical concerns. There is a significant main effect of laughable context $F(2,3226) = 31.15$, $p < 0.001$, $\eta^2 = 0.01$. There is also a significant main effect of laugh intensity $F(1,3226) = 138.75$, $p < 0.001$, $\eta^2 = 0.04$. Finally, there is a significant interaction of laughable context and laugh $F(2,3226) = 7.01$, $p < 0.001$, $\eta^2 = 0.004$.

Discussion

The pattern of results for high intensity laughter in Experiment 2 is even more similar to the original hypothesis than those observed in Experiment 1. Genuineness ratings are almost identical for the control and 'same intensity' conditions with a drop off in genuineness ratings for the 'opposite intensity' condition. This provides strong support for the original hypothesis in the case of high intensity laughter. Thus, interchanging high intensity laughter seems to have little or no effect on the perceived level of genuineness of the interaction.

The pattern of results for low intensity laughter is somewhat different to those from Experiment 1, the key difference being that genuineness ratings do not drop off in the 'opposite intensity' condition. This is the only condition that failed to replicate results of Experiment 1. It may be that low intensity laughter is inherently more ambiguous than high intensity laughter; McKeown et al. (in preparation) argue that low intensity laughter has many more functions than high intensity laughter, with the latter more closely related to the assessment of humour production. Although low intensity laughter is seen as being part of less genuine interactions, these interactions are not rated as

'not genuine'; rather they occupy the midpoint on the scale. It may be that high intensity laughter may more unequivocally indicate genuineness whereas low intensity laughter can be interpreted in many different ways–thus increasing the likelihood it will be viewed as being consistent with the context it is inserted into.

Another way of understanding the results may be that the nature of these laughter effects in the two opposite scenario interactions leads to quite different interpretations of the social interaction. In one a high intensity laugh response occurs despite the story-teller not providing contextual cues that a high intensity laugh was expected. Laughter in such a scenario might reasonably be interpreted as an over effusive laugh response, and consequently the interaction deemed to be less genuine. In the other case the contextual cues did indicate that a high intensity laugh response was expected but was greeted with a low intensity laugh–an interactional situation that may be observed to be an insult or social rejection–these situations were rated as the least genuine of all.

GENERAL DISCUSSION

The experiments reported here address an important question regarding the function of laughter; namely, is it used to signal the laugher's underlying emotional state or is its function to influence the listener's emotional state? This question addresses a current debate on the function of human laughter. While the representational model of laughter proposes that laughter encodes information about the laugher's emotion(s), which is then decoded by the listener, the affect-induction model depicts laughter as a communicative tool with which to influence the emotions of the listener. The representational model views laughter as an unambiguous signal of the laugher's emotional state; the affect-induction model, on the other hand, proposes that laughter interpretation is determined by situational factors. Thus, unlike the representational model, the affect-induction model would predict that laughter is an ambiguous signal subject to contextual influences. Our experiments sought to pit the two models against each other by switching laughter in an original interaction with different laughter from a different part of the interaction. The representational model would predict that participants should be able to differentiate between original interactions and interactions in which laughter has been switched; the affect-induction model, on the other hand, would predict that participants would not be able to make this distinction. Furthermore, we argued that any inability to differentiate between original interactions and those in which laughter has been switched should be restricted to original and replacement laughs of similar intensity.

The results of Experiment 1 show that participants' ratings of an interaction's genuineness were unaffected when the listener's laughter was swapped for another instance of laughter of similar intensity and from the same listener, but from a different point of the interaction. However, replacing listener laughter with laughter of a different intensity resulted in participants rating the story telling interaction as less genuine. This demonstration of the ambiguous nature of same intensity laughter is consistent with the affect-induction model, which would argue that laughter is necessarily ambiguous as regards the laugher's underlying emotional state.

Experiment 1 had a number of limitations. For example, not all possible context – laughter combinations were used to generate the stimuli. Furthermore, participants would have viewed each context several times, with each containing a different laugh stimulus; thus it is feasible that responses to a given context would have been tempered by previous responses to the same context (but paired with different laughter). Experiment 2 overcame these limitations by using the full range of context-laughter combinations and ensuring that participants were presented with each context only once. The results of Experiment 2 replicated those of Experiment 1 in all but one condition. When laughter (high or low intensity) was switched for laughter of a similar intensity, participants' genuineness judgements were unaffected. When high intensity laughter was inserted into a low intensity context, participants judged the interaction as less genuine. However, this was not the case when low intensity laughter was inserted into a high intensity context; rather, the interaction was judged to be as genuine as the control and same-intensity conditions. It has been proposed that low intensity laughter is functionally more complex and inherently more ambiguous than high intensity laughter (McKeown et al., in preparation), and this may explain why inserting a low intensity laugh into a high intensity context did not result in the interaction appearing less genuine.

An interesting finding was that interactions retaining their original low intensity laughter were consistently judged less genuine than those retaining their original high intensity laughter. Previous research highlighting physiological differences between spontaneous and volitional laughter production (Bryant and Aktipis, 2014) might offer an explanation for this finding. It would be reasonable to assume that differences in the sounds of spontaneous and volitional laughter are likely to be magnified with increasing laughter intensity, and that it should be more difficult to differentiate between low intensity spontaneous and volitional laughter. The low genuineness ratings of the low intensity laughs might, therefore, be a consequence of the relative difficulty in correctly identifying spontaneous low intensity laughter. In other words, if participants are uncertain about a laugh's spontaneity they will be more likely to identify the interaction as being less genuine. This may explain why the observed reduction in genuineness scores in Experiment 1 when low intensity laughter was combined with a different intensity context did not generalise to Experiment 2. Recall that Experiment 2 was motivated by a desire to use more laughter stimuli than in Experiment 1. A consequence of this was that a higher proportion (75%) of low intensity laughs in Experiment 2 were voiced compared to Experiment 1 (50%). If voiced laughter is judged as more genuine, then the higher proportion of voiced laughter in Experiment 2 might explain the different results in this condition across experiments.

The results of our experiments provide compelling evidence that laughter is an inherently ambiguous stimulus, and that its

interpretation is largely determined by the context in which it occurs. As such these results support the affect-induction model. While it is the case that we often talk of laughter 'types,' such as sardonic laughter, joyful laughter, taunting laughter, and schadenfreude laughter, our results suggest that different instances of same-intensity laughter are largely interchangeable and that their specific meaning may be largely determined by the context in which they occur. This flexibility is reminiscent of recent, similar findings relating to the facial expression of emotions. Historically, it has been widely accepted that the facial expression of emotion is underpinned by emotion-specific facial musculature activation patterns (Ekman and Friesen, 1971); from this perspective a sad facial expression and a fearful facial expression are associated with distinct combinations of facial movements. However, recent developments suggest that, despite all their information value, facial expressions can be ambiguous and that their meaning is largely dependent on contextual information beyond the face (Aviezer et al., 2011; Barrett et al., 2011; Hassin et al., 2013). Our results suggest that the important role of context in the perception of facial expressions also applies to the interpretation of laughter.

CONCLUSION

We tested which of two models, representational or affect-induction, best describes the function of laughter. We devised novel experiments such that the two models made opposite predictions, and found that the results are consistent with the affect-induction model's prediction.

ETHICS STATEMENT

The experiments were carried out in accordance with the recommendations of the School of Psychology Ethics Committee, Queen's University Belfast, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the School of Psychology Ethics Committee, Queen's University Belfast.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

The research leading to these results has received funding from the European Union Seventh Framework Program under grant agreement No. 270780 (ILHAIRE), The Leverhulme Trust under grant agreement No. RPG-2016-326, and the European Union Horizon 2020 Research and Innovation Program under grant agreement No. 645378 (ARIA-VALUSPA).

REFERENCES

Aviezer, H., Bentin, S., Dudarev, V., and Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion* 11, 1406–1414. doi: 10.1037/a0023578

Bachorowski, J. A., and Owren, M. J. (2001). Not all laughs are alike: voiced but not unvoiced laughter readily elicits positive affect. *Psychol. Sci.* 12, 252–257. doi: 10.1111/1467-9280.00346

Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614–636. doi: 10.1037/0022-3514.70.3.614

Bänziger, T., and Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Commun.* 46, 252–267. doi: 10.1016/j.specom.2005.02.016

Barrett, L. F., Mesquita, B., and Gendron, M. (2011). Context in emotion perception. *Psychol. Sci.* 20, 286–290. doi: 10.1177/0963721411422522

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Biele, C., and Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Exp. Brain Res.* 171, 1–6. doi: 10.1007/s00221-005-0254-0

Bonin, F., Campbell, N., and Vogel, C. (2012). Laughter and topic changes: temporal distribution and information flow. *Paper Presented at the CogInfoCom 2012 – 3rd IEEE International Conference on Cognitive Infocommunications*, Kosice. doi: 10.1109/CogInfoCom.2012.6422056

Bryant, G. A., and Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. *Evol. Hum. Behav.* 35, 327–335. doi: 10.1016/j.evolhumbehav.2014.03.003

Casler, K., Bickel, L., and Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Comput. Hum. Behav.* 29, 2156–2160. doi: 10.1016/j.chb.2013.05.009

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: L. Erlbaum Associates.

Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals and Meta-Analysis*. New York, NY: Routledge.

Cumming, G. (2014). The new statistics: why and how. *Psychol. Sci.* 25, 7–29. doi: 10.1177/0956797613504966

Cumming, G., and Calin-Jageman, R. (2017). *Introduction to the New Statistics: Estimation, Open Science, and Beyond*. New York, NY: Routledge.

Darwin, C. (1872). *The Expressions of the Emotions in Man and Animal*. London: John Murray. doi: 10.1037/10001-000

Davila-Ross, M., Allcock, B., Thomas, C., and Bard, K. A. (2011). Aping expressions? Chimpanzees produce distinct laugh types when responding to laughter of others. *Emotion* 11, 1013–1020. doi: 10.1037/a0022594

Davila Ross, M., Owren, M. J., and Zimmermann, E. (2009). Reconstructing the evolution of laughter in great apes and humans. *Curr. Biol.* 19, 1106–1111. doi: 10.1016/j.cub.2009.05.028

Ekman, P., and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.* 17, 124–129. doi: 10.1037/h0030377

Feldman Barrett, L. (2006). Are emotions natural kinds? *Perspect. Psychol. Sci.* 1, 28–58. doi: 10.1111/j.1745-6916.2006.00003.x

Feldman Barrett, L. (2017). *How Emotions Are Made*. Basingstoke: Macmillan.

Fridlund, A. J. (1994). *Human facial expression: An Evolutionary View*. San Diego, CA: Academic Press.

Friedmann, P. D., Rose, J. S., Swift, R., Stout, R. L., Millman, R. P., and Stein, M. D. (2008). Trazodone for sleep disturbance after alcohol detoxification: a double-blind, placebo-controlled trial. *Alcohol. Clin. Exp. Res.* 32, 1652–1660. doi: 10.1111/j.1530-0277.2008.00742.x

Gelman, A., and Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *Am. Stat.* 60, 328–331. doi: 10.1198/000313006X152649

Glenn, P. J. (2003). *Laughter in Interaction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511519888

Hassin, R. R., Aviezer, H., and Bentin, S. (2013). Inherently ambiguous: facial expressions of emotions, in context. *Emot. Rev.* 5, 60–65. doi: 10.1177/1754073912451331

Holt, E. (2010). The last laugh: shared laughter and topic termination. *J. Pragmat.* 42, 1513–1525. doi: 10.1016/j.pragma.2010.01.011

Holt, E. (2011). On the nature of "laughables": laughter as a response to overdone figurative phrases. *Pragmatics* 21, 393–410. doi: 10.1075/prag.21.3.05hol

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science, a practical primer for *t*-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863

Matsusaka, T. (2004). When does play panting occur during social play in wild chimpanzees? *Primates* 45, 221–229.

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., et al. (2013). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cereb. Cortex* 25, 246–257. doi: 10.1093/cercor/bht227

McGraw, K. O., and Wong, S. P. (1992). A common language effect size statistic. *Psychol. Bull.* 111, 361–365. doi: 10.1037/0033-2909.111.2.361

McKeown, G., and Curran, W. (2015). "The relationship between laughter intensity and perceived humour," in *Proceedings of the 4th International Workshop on Laughter and Other Non-verbal Vocalisations in Speech*, eds K. Truong, D. Heylen, J. Trouvain, and N. Campbell (Dublin: Queen's University Belfast), 27–29.

McKeown, G., Curran, W., Wagner, J., Lingenfelser, F., and André, E. (2015). "The Belfast storytelling database: a spontaneous social interaction database with laughter focused annotation," in *Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (Xi'an: IEEE) 166–172. doi: 10.1109/ACII.2015.7344567

McKeown, G. J., and Sneddon, I. (2014). Modeling continuous self-report measures of perceived emotion using generalized additive mixed models. *Psychol. Methods* 19, 155–174. doi: 10.1037/a0034282

O'Donnell Trujillo, N., and Adams, K. (1983). Heheh in conversation: some coordinating accomplishments of laughter. *West. J. Commun.* 47, 175–191.

Owren, M., and Bachorowski, J. A. (2003). Reconsidering the evolution of nonlinguistic communication: the case of laughter. *J. Nonverbal Behav.* 27, 183–200. doi: 10.1023/A:1025394015198

Paolacci, G., and Chandler, J. (2014). Inside the turk: understanding mechanical turk as a participant pool. *Curr. Dir. Psychol. Sci.* 23, 184–188. doi: 10.1177/0963721414531598

Papousek, I., Aydin, N., Lackner, H. K., Weiss, E. M., Bühner, M., Schulter, G., et al. (2014). Laughter as a social rejection cue: gelotophobia and transient cardiac responses to other persons' laughter and insult. *Psychophysiology* 51, 1112–1121. doi: 10.1111/psyp.12259

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Pers. Soc. Psychol. Bull.* 9, 278–311. doi: 10.1207/s15327957pspr0904_1

Preuschoft, S., and van Hooff, J. A. R. A. M. (1997). "The social function of "Smile" and "Laughter": variations across primate species and societies," in *Nonverbal Communication: Where Nature Meets Culture*, eds U. Segerstråle and P. Molnár (Mahweh, NJ: Lawrence Erlbaum Associates), 171–189.

Quené, H., and van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Commun.* 43, 103–121. doi: 10.1016/j.specom.2004.02.004

R Core Team (2017). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Available at: https://www.R-project.org/

Reddy, M. J. (1979). The conduit metaphor: a case of frame conflict in our language about language. *Metaphor Thought* 2, 164–201.

Rendall, D., Owren, M. J., and Ryan, M. J. (2009). What do animal signals mean? *Anim. Behav.* 78, 233–240. doi: 10.1016/j.anbehav.2009.06.007

Rothbart, M. K. (1973). Laughter in young children. *Psychol. Bull.* 80, 247–256. doi: 10.1037/h0034846

Russell, J. A., Bachorowski, J. A., and Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annu. Rev. Psychol.* 54, 329–349. doi: 10.1146/annurev.psych.54.101601.145102

Saal, F. E., Downey, R. G., and Lahey, M. A. (1980). Rating the ratings: assessing the psychometric quality of rating data. *Psychol. Bull.* 88, 413–428. doi: 10.1037/0033-2909.88.2.413

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., and Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f^2, a measure of local effect size, from PROC MIXED. *Front. Psychol.* 3:111. doi: 10.3389/fpsyg.2012.00111

van Hooff, J. A. R. A. M. (1972). "A comparative approach to the phylogeny of laughter and smiling," in *Non-Verbal Communication,* ed. R. A. Hinde (Cambridge: Cambridge University Press), 209–241.

Vettin, J., and Todt, D. (2005). Human laughter, social play, and play vocalizations of non-human primates: an evolutionary approach. *Behaviour* 142, 217–240. doi: 10.1163/1568539053627640

Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., and André, E. (2013). "The social signal interpretation (SSI) framework - multimodal signal processing and recognition in real-time," in *Proceedings of the 2013 ACM Multimedia*, Barcelona.

Wild, B., Rodden, F. A., Grodd, W., and Ruch, W. (2003). Neural correlates of laughter and humour. *Brain* 126(Pt 10), 2121–2138. doi: 10.1093/brain/awg226

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Curran, McKeown, Rychlowska, André, Wagner and Lingenfelser. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 July 2018
doi: 10.3389/fpsyg.2018.01182

Candidate Performance and Observable Audience Response: Laughter and Applause–Cheering During the First 2016 Clinton–Trump Presidential Debate

Patrick A. Stewart[1], Austin D. Eubanks[2], Reagan G. Dye[1], Zijian H. Gong[3], Erik P. Bucy[3], Robert H. Wicks[4] and Scott Eidelman[2]*

[1] Department of Political Science, University of Arkansas, Fayetteville, AR, United States, [2] Department of Psychological Science, University of Arkansas, Fayetteville, AR, United States, [3] College of Media and Communication, Texas Tech University, Lubbock, TX, United States, [4] Department of Communication, University of Arkansas, Fayetteville, AR, United States

OPEN ACCESS

Edited by:
Tracey Platt, University of Wolverhampton, United Kingdom

Reviewed by:
Ulrich Von Hecker, Cardiff University, United Kingdom
Martina Raue, Massachusetts Institute of Technology, United States

***Correspondence:**
Patrick A. Stewart
pastewar@uark.edu

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *30 September 2017*
Accepted: *19 June 2018*
Published: *20 July 2018*

Citation:
Stewart PA, Eubanks AD, Dye RG, Gong ZH, Bucy EP, Wicks RH and Eidelman S (2018) Candidate Performance and Observable Audience Response: Laughter and Applause–Cheering During the First 2016 Clinton–Trump Presidential Debate. Front. Psychol. 9:1182. doi: 10.3389/fpsyg.2018.01182

Raucous audience applause–cheering, laughter, and even booing by a passionately involved electorate marked the 2016 presidential debates from the start of the primary season. While the presence and intensity of these observable audience responses (OARs) can be expected from partisan primary debates, the amount of not just laughter, but also applause–cheering and booing during the first general election debate between Hillary Clinton and Donald Trump was unprecedented. Such norm-violating audience behavior raises questions concerning not just the presence, strength, and timing of these OAR, but also their influence on those watching on television, streaming video, or listening to radio. This report presents findings from three interconnected studies. Study 1 provides a baseline for analysis by systematically coding the studio audience response in terms of utterance type (laughter, applause–cheering, booing, and mixtures), when and how intensely it occurred, and in response to which candidate. Study 2 uses observational analysis of 362 undergraduate students at a large state university in the southern United States who watched the debate on seven different news networks in separate rooms and evaluated the candidates' performance. Study 2 considered co-occurrence of OAR in the studio audience and in the field study rooms, finding laughter predominated and was more likely to co-occur than other OAR types. When standardized cumulative strength of room OAR was compared, findings suggest co-occurring OAR was stronger than that occurring solely in the field study rooms. Analysis of truncated data allowing for consideration of studio audience OAR intensity found that OAR intensity was not related to OAR type occurring in the field study rooms, but had a small effect on standardized cumulative strength. Study 3 considers the results of a continuous response measure (CRM) dial study in which 34 West Texas community members watched and rated the candidates during the first debate. Findings suggest

that applause–cheering significantly influenced liking of the speaking candidate, whereas laughter did not. Further, response to applause–cheering was mediated by party identity, although not for laughter. Conclusions from these studies suggest laughter as being more stereotypic and likely to be mimicked whereas applause–cheering may be more socially contagious.

Keywords: defining moments, observed audience response (OAR), laughter, applause–cheering, booing, presidential debates, intra-audience effects

INTRODUCTION

The 2016 election can be seen as one in which a passionately involved electorate was key for its unexpected outcome as novice political outsider Donald Trump became president of the United States. Trump's success defied early predictions, with few political experts anticipating the intensity from his base of support when compared to more traditional candidates during both the Republican primaries and general election. Despite dispensing with traditional expectations and violating presidential debate norms, Trump's performance and the associated audience response of raucous applause–cheering, laughter, and even booing during the initial 2016 primary debates (Stewart et al., 2016) and the general election debates can be seen as providing insights concerning his populist appeal. Beyond their populist overtones, these observable audience responses (OARs) can thus be seen as valid and reliable audible indicators of the intensity of shared individual and emergent group attitudes toward political candidates more generally (Stewart, 2012, 2015; Stewart et al., 2016).

Existing debate-focused research has documented the role of these salient media events in reinforcing existing preferences, producing issue knowledge, and influencing perceptions of candidate character, thus affecting undecided voter choices (McKinney and Warner, 2013). Debate viewing may also reorder the relative importance of issues in viewers' minds and shift leadership potential to the foreground as a salient consideration (Benoit et al., 2001; Schrott and Lanoue, 2008; Schroeder, 2016). However, most existing research treats debates as monolithic events and examines overall debate effects rather than communication dynamics occurring during the debates themselves.

While providing useful insights concerning the impact of mediated events on electoral dynamics, these approaches do not take into consideration the unpredictable events that occur during debates and how they affect perceptions. Even after accounting for how campaigns pitch-and-spin their candidates' performance (Norton and Goethals, 2004; Schroeder, 2016) multiple, relatively unexplored factors occurring during the debate affects candidate evaluations. Candidate rhetorical approach (Benoit, 2013), non-verbal behavior (Bucy and Stewart, Forthcoming), and media presentation style (Cavari et al., 2017) influence debate viewer perceptions. In other words, most research does not consider the process of change in debate viewer perceptions or those critical defining moments, which are often met with audience laughter, applause–cheering, and/or booing (Clayman, 1995).

Recent research addresses this oversight through continuous response measures (CRMs) and dial testing of debates, eye tracking of candidate exchanges, and focus group analysis of memorable debate moments (Gong and Bucy, 2015). Analysis of social media such as Twitter also suggests that candidate non-verbal behavior, even more so than their verbal acclaims, attacks and defenses (Shah et al., 2015), influence audience response. Still, these approaches may not capture the contemporaneous, in-person emotional response of viewers, instead representing more considered appraisals (Nagel et al., 2012) prone to social conformity pressures (Fein et al., 2007; Davis et al., 2011). Furthermore, by focusing on individual response such measures might be missing the highly important attribute of implicit sociality embedded with audible responses by the audience, especially during emotionally charged political events.

Research considering OAR to political candidates at events such as debates tends not to focus on the audience itself, and its social influence on other audience members. The existing research that does consider OAR on participant evaluation, including those considering political figures, are experimental and do not disambiguate positive response such as laughter, applause, and/or visually oriented non-verbal signals (Hylton, 1971; Duck and Baggaley, 1975; Cummins and Gong, 2017). Specifically, the studies by Wiegman (1987) and Fein et al. (2007), while providing insight into the social influence of OAR on participant evaluation of the candidates and policy issues, tend to include both audible reactions and OAR. For instance, Wiegman (1987) carried out a field experiment with a well-known Dutch political figure that involved a studio audience either reacting positively, negatively, or neutrally through a range of audible utterances and variety of gestures and facial displays. Fein et al. (2007) found that "... absent the applause, laughter, and general approval of [United States President Ronald] Reagan's one-liners, these responses were not seen as particularly noteworthy by the participants." (p. 178) However, they did not differentiate between applause and laughter nor the moderator's verbal and non-verbal response.

While not dealing with political figures, Axsom et al. (1987) auditory-based lab experiment comparing "enthusiastic applause–cheering" to unenthusiastic and polite applause with occasional derisive cries, found that OAR influenced response to a specific policy issue (imprisonment vs. probation). They noted that "the persuasive impact of audience cues may reflect subjects' tendencies to use a simple consensus heuristic such as "if other people think the message is correct, then it is probably valid." (p. 39) In summary, while previous research provides useful insights, a gap in the literature exists by the authors not

differentiating between OAR types or systematically considering OAR intensity.

The existing research that does differentiate between the types of OAR tends to consider these utterances as a means by which large groups of followers provide feedback to their leaders. Specifically, audience responses such as applause–cheering, laughter, and booing provide audible signals indicating the type of response while indexing level of follower support or opposition. Furthermore, the timing of OAR indicates their level of synchrony with the speaker, as well as that with fellow audience members (Bull and Wells, 2002). Thus, the type and magnitude of the OAR supplies audible information indicating coalition size and strength (Dunbar, 1993) providing the speaker with immediate and unobtrusive feedback that may be continuously monitored and allow for enhanced speechmaking (West, 1984).

At the same time, media audiences, whether streaming the debates, watching on television, or listening through other broadcast media, as well as journalists reporting on the event, may be affected by this information. Indeed, OAR can lead to change regarding how the speaker is evaluated, indeed, even more so than the eliciting comments themselves (Fein et al., 2007). In other words, social influence asserted through OAR affects resultant viewer and listener perceptions, attitudes, and behavior; however, the specific influence of different OAR such as applause–cheering, laughter, and booing remains to be studied in depth.

Observable Audience Response (OAR) Reliability

Observable audience response such as applause–cheering, laughter, and booing may be seen as belonging to a class of behavior that is almost automatic and highly contagious, which in turn might lead to affective, cognitive, and behavioral response with political implications (Fein et al., 2007). In other words, there likely is a high level of behavioral mimicry by audience members as they match each other's audible response (Sachisthal et al., 2016; Moody et al., 2017). This audible response may in turn influence individual emotional response, and with it the evaluation of the candidate eliciting the response (as well as those sharing in the response) through emotional contagion (Hatfield et al., 1994, 2014; Lakin et al., 2003). These group vocalizations can thus provide evidence of the type and intensity of connection the audience members have with the candidates, and perhaps as important, the members have with each other in the room.

The overarching issue regarding OAR concerns their reliability in differentially reflecting the audience's putative emotional and behavioral intent. Here, reliable indicators of emotion may be defined as being first, an accurate recognition of the emotional state of the communicator, and their resultant behavioral intent, and second, the signal being an index of the sender's underlying state by being costly to produce (Mehu et al., 2011). Because of the social nature of group vocalizations, these utterances should be stereotyped and contagious; in other words, such behaviors as laughing and yawning have coherent and identifiable vocalic, facial and even postural display behavior associated with them. As defined by Hatfield et al. (2014) this primitive social contagion is "(T)he tendency to automatically mimic and synchronize facial expressions, vocalizations, postures, and movements with those of another person's and, consequently, to converge emotionally" (p. 169).

Despite the rather sparse nature of existing research on location of debate viewing and audience composition, we expect differences in the in-person studio audiences and those having a mediated experience. In other words, the studio audience likely reacts differently from those watching a video of the event. This may be due in part to a location's acoustic qualities that may enhance or diminish the subjective emotional and physiological response of audience members (Stewart, 2012; Pätynen and Lokki, 2016) as well as the physical presence of contending candidates. Differences in response may further be affected by whether individuals are watching independently or amongst other individuals, whether known acquaintances or strangers, with increased laughter, if not the other OAR types, occurring with greater sociality. Furthermore, social norms likewise play a role in what is acceptable behavior or not, although this may be determined by audience member assumptions and relationships with each other (Devereux and Ginsburg, 2001; Platow et al., 2005; Fridlund, 2017).

Thus, in addition to the type of OAR (e.g., applause–cheering, laughter, and booing) identified and potential mixtures that might occur, the intensity of studio audience response may be characterized by its length in time combined with its perceived audible strength. This intensity may in turn affect onlookers, whether in the studio audience – yet not affiliated with any social group or faction – or watching on television, live streaming over the internet, or listening on the radio and thus experiencing intra-audience mediated effects from the OAR (Cummins and Gong, 2017).

Research Question 1: Is there a relationship between the presence of television studio audience OAR and the field study audience OAR?

Research Question 2: Is there a relationship between the intensity of television studio audience OAR and the strength of the field study audience OAR?

Observable Audience Response (OAR) Types: Laughter, Applause–Cheering, and Booing

Generally speaking, one can identify three general types of audible OAR as applause–cheering[1], laughter, and booing each serve to signal shared audience response to political candidates (Atkinson, 1984; West, 1984; Heritage and Greatbatch, 1986; Clayman, 1992, 1993; Bull and Miskinis, 2014). These OAR types, in addition to their effects being characterized by length and strength, may be accentuated or attenuated depending on audience member characteristics and the intensity of

[1]These two forms of audience audible utterances are combined for the sake of this analysis; we do appreciate that they reflect different kinds of communication using different non-verbal channels (manipulation of hands and vocalizations) (Schweingruber and McPhail, 1999).

their response. Each OAR type serves distinct communicative ends allowing for audiences to communicate their support or disapproval for statements by leaders and putative leaders, with concomitant intensity and mixtures providing insight concerning passion and unanimity regarding these positions.

Laughter is the most studied of all vocalizations discussed here; however, the focus tends not to be on the group. Individual laughter is focused on due to it serving as a pervasive social signal in interpersonal interactions by punctuating speech and indicating speaking turn taking and transition (Provine, 1993; Gilmartin et al., 2013; Bonin et al., 2014; Scott et al., 2014). Individual laughter can indicate social intent through it being voiced and unvoiced (Bachorowski and Owren, 2001; Owren and Bachorowski, 2003) as well as communicating the different emotions of amusement, contempt, schadenfreude, and tickle (Szameitat et al., 2009, 2011).

As a result, laughter may be seen as a costly signal by virtue of it either being evoked in a manner that is difficult to control whereas even emitted laughter that is initially faked leads to physiological change (Provine, 1992; Bachorowski and Owren, 2001; Devereux and Ginsburg, 2001; Ruch and Ekman, 2001; McGettigan et al., 2015). Individual laughter likewise serves as a social lubricant by affecting subject mood states by decreasing negative affect, increasing positive affect and enhancing pain tolerance while increasing social cooperation and group identity (Van Vugt et al., 2014). It thus serves as a highly reliable social signal regarding behavioral intent (van Hooff and Preushoft, 2003; Panksepp, 2007; Pellis et al., 2014).

When considering group level behavior, research regarding laughter tends to focus on the target and intent of the verbal utterances leading to this type of response (Wells and Bull, 2007; Stewart, 2012; Choi et al., 2016). Thus, research concerning group laughter tends to reflect findings regarding response to individual speakers. The group vocalic utterances of laughter is limited in length of time to a much greater extent than those created through rhythmic mechanical noisemaking such as applause tending to last from 1 to 3 s in comparison with 2 to 8 s for applause–cheering (Stewart, 2015), as well as likely booing (although these types of rare OAR makes strong assertions untenable). Furthermore, when an audience shows their appreciation for a humorous comment, applause–cheering prolongs the laughing utterance (Stewart, 2012; Stewart et al., 2016). This suggests high levels of social mimicry in the immediate OAR and then likely social contagiousness through its continuation.

Of all the forms of OAR, applause–cheering is perhaps most likely to be observed in group settings such as political speeches and intra-party debates. This is likely due to the ease with which candidates are able to evoke it among supporters in partisan settings. As a result, applause–cheering has been appreciated for the role it plays in providing an important barometer of a politicians' individual appeal during speeches (Atkinson, 1984; West, 1984; Heritage and Greatbatch, 1986; Bull, 2003) or when in direct competition with other candidates during debates (Stewart, 2015; Stewart et al., 2016).

On the other hand, due to applause–cheering likely not being as costly to produce physiologically and easier for audience members to inhibit than laughter (Stewart, 2015), it might not be as reliable a social signal. That does not mean that this activity is not stereotyped and thus easy to identify while also being contagious. Research concerning applause bouts in small groups (13–20) found that most involve only 9–15 claps per person, although some last over 30 claps (Mann et al., 2013). A study considering applause in larger groups suggests this activity typically begins with an uncoordinated loud burst of high frequency clapping that then synchronizes through a form of social contagion and coordination (Néda et al., 2000a,b). Thus, while the initial applause is louder, the synchronicity of OAR afterward suggests social contagion between audience members.

Much rarer than supportive in-person audience response through laughter and applause–cheering at political events are boos and jeers (Clayman, 1992, 1993; Bull and Miskinis, 2014). However, besides research regarding individuals jeering/heckling carried out over 40 years ago (Sloan et al., 1974; Silverthorne and Mazmanian, 1975), little research on the nature of booing, especially regarding physiological characteristics and group-level attributes, has been carried out. Existing research on booing finds it rarely occurring. Even in the highly divisive 2016 presidential primary debates, booing, both alone and mixed with applause–cheering and laughter, occurred only in 5% of OAR observed (Stewart et al., 2016). Beyond audibly signaling negative response, the intent and target matters; disaffiliative booing by the "right" crowd can enhance electoral status by emphasizing willingness to take an unpopular stand whereas affiliative booing may be used to attack on out-group leaders and policy positions (Bull and Miskinis, 2014). However, the key factor is that the booing occurred during speeches in front of relatively coherent partisan audiences.

In summary, laughter, applause–cheering, and booing provide means by which the audience physically present with a politician can communicate as a group in distinctive and easily identifiable ways. While pre-verbal, these OAR can successfully be used to strategically communicate factional preferences to not just the speaker, but also to other potential group members. As a result, there are social benefits and costs from participating or not participating in OAR; audience members must consider if engaging in different OAR types will be socially costly to them or if joining in with other audience members when candidates break norms of politeness and civility will pay off socially (Dailey et al., 2005). To the point, the social norms of politeness by audience members instructed to not influence the proceedings through their laughter, applause–cheering, and booing can be contravened if their preferred candidate welcomes, even incites it, and there is no effective sanction laid upon them. While we expect the candidates to successfully evoke OAR through punchlines, claptrap, and all manner of rhetorical tools at their disposal, the type of OAR will likely vary systematically. Because laughter is difficult to control, we do not expect that the candidate evoking it will influence either its occurrence or the strength of the field study audiences' response. On the other hand, with applause–cheering we do expect that both the candidate making the comment inciting this response and the intensity of the studio audience's response will influence the strength of the response.

Research Question 3: Is there a relationship between studio audience OAR type and field study audience OAR type?

Research Question 4: Is there a relationship between studio audience OAR type and field study audience OAR intensity?

Present Research

This report presents the findings of distinct, yet interconnected studies to explore the nature of OAR and their potential influence on evaluation of presidential candidates during a general election debate. We take a bottom-up/reverse engineering approach to study behavior as it occurs in a naturalistic environment (de Gelder, 2017); essentially we use the observational methods used in human ethology (Schubert, 1988; Eibl-Eibesfeldt, 1989; Masters, 1989; Weisfeld, 1993; Salter, 2007) and apply them to a political event of great importance as it occurs. As a result, this report is by necessity correlational and exploratory.

We focus on the first general election debate between Donald Trump and Hillary Clinton in a multipart approach. Study 1 uses ANVIL content coding software to characterize and analyze studio audience response in terms of when the OAR occurred, what type they were (laughter, applause–cheering, booing, and mixtures), their duration and perceived strength, and in response to which candidate. Study 2 builds off of Study 1 by collecting and analyzing a unique dataset in which 362 undergraduate students took part in a field experiment watching or listening to the first presidential general election debate in seven different rooms. We use ethological analysis of the field study participants' OAR by considering when different types occurred and how strong they were perceived to be by observers. This allows us to compare relatively unfettered field study audiences to the studio audience, where moderator instructions and politeness expectations presumably played a role in constraining an elite partisan audience, to the less inhibited university student-occupied rooms. We draw conclusions regarding both laughter and applause–cheering by considering four research questions concerning the co-occurrence of the OAR of laughter and applause–cheering (i.e., simultaneously occurring in both the studio audience and in the field study rooms). With Study 3, we evaluate the effect of studio audience laughter and applause–cheering on mediated viewer moment-to-moment (MTM) response of liking the speaking candidate. We finish this report by discussing the implications of our findings for future research.

STUDY 1

Materials and Methods

The first of three general election debates between Democratic Party presidential nominee Hillary Clinton and Republican Party nominee Donald Trump occurred the evening of Monday September 26, 2016 and was hosted outside New York City by Hofstra University. Sponsored by the *Commission on Presidential Debates* and moderated by NBC News anchor Lester Holt, the 90-min debate focused on the topics *achieving prosperity*, *America's direction*, and *securing America* with specific questions regarding jobs, race relations, taxes, and the prospect of cyberattacks. With an estimated 84 million viewers, the highly anticipated first head-to-head confrontation between Trump and Clinton became the most watched debate in United States history (Cavari et al., 2017).

Speaking time and studio audience OAR used ANVIL content analysis software, which allows for frame-by-frame coding (Kipp, 2012). The inter-coder reliability (ICR) between two coders considering speaking time and OAR assessed approximately 30 min randomly chosen from video clips coded surpassed acceptable levels ($\kappa > 0.92$).

Findings

Trump had nearly 5 min more speaking time at 47 min (2,795 s) when compared to Clinton's 42 min (2,492 s). This was likely due to his interruptions, as Trump had nearly twice as many speaking turns ($n = 80$) as Clinton ($n = 43$). With moderator Lester Holt's speaking time of 10 min (597 s) and 91 speaking turns, the total floor time of the three debate participants was 98 min over 214 total speaking turns, suggesting a high level of overlap.

A total of 34 OAR were identified during the debate proper (we did not code for the welcoming or concluding applause). These 34 studio audience OAR to the candidates' statements/retorts – or in one case response to the moderator – lasted a total of 102.72 s and averaged just over 3 s ($M = 3.02$; $SD = 1.96$). When considering types of OAR, 21 laughter ($M = 2.09$; $SD = 1.20$; Min = 0.4, Max = 4.17), nine applause–cheering ($M = 5.48$; $SD = 1.48$; Min = 3.4, Max = 7.97), two booing ($M = 1.52$; $SD = 0.26$; Min = 1.33, Max = 1.7), and two mixed vocalizations [applause and laughter (4.3 s); applause and booing (2.17 s)] were identified. Due to the lack of variance, the two booing and two mixed responses are omitted from statistical analyses, but considered in the descriptive analysis.

In addition to evaluating length of the audience's utterances, we coded for the subjective strength of these responses on a 1- to 5-point scale ranging from "barely audible" to "extremely audible" (Ekman and Friesen, 2003) using three coders ($\alpha = 0.76$). The mean of the three was computed to form our *strength* variable ($M = 2.78$; $SD = 1.26$). Due to the high level of correlation between these two measures of OAR length and strength (Pearson's $r = 0.81$), we created an additive studio audience intensity index ($M = 5.08$; $SD = 3.06$). Throughout this manuscript we report *t*- and *p*-values to allow for standard statistical consideration, but note that analysis of the population of studio audience and field study OAR means that such statistical standards are not strictly appropriate.

Discussion

In comparison with previous general election debates (Rhea, 2012; Stewart, 2012) the first 2016 meeting between Hillary Clinton and Donald Trump was a raucous affair just in terms of the 21 laughter events. This finding aligns with expectations and findings suggesting that while both OAR types involve levels of social contagion, laughter likely is more reliable due to the relative absence of control over it (Stewart, 2015). However, it is the amount of voluntary audience involvement that sets this debate apart. Nine (26.5%) OAR involved applause and cheering, one involved laughter mixed with applause, and two involved

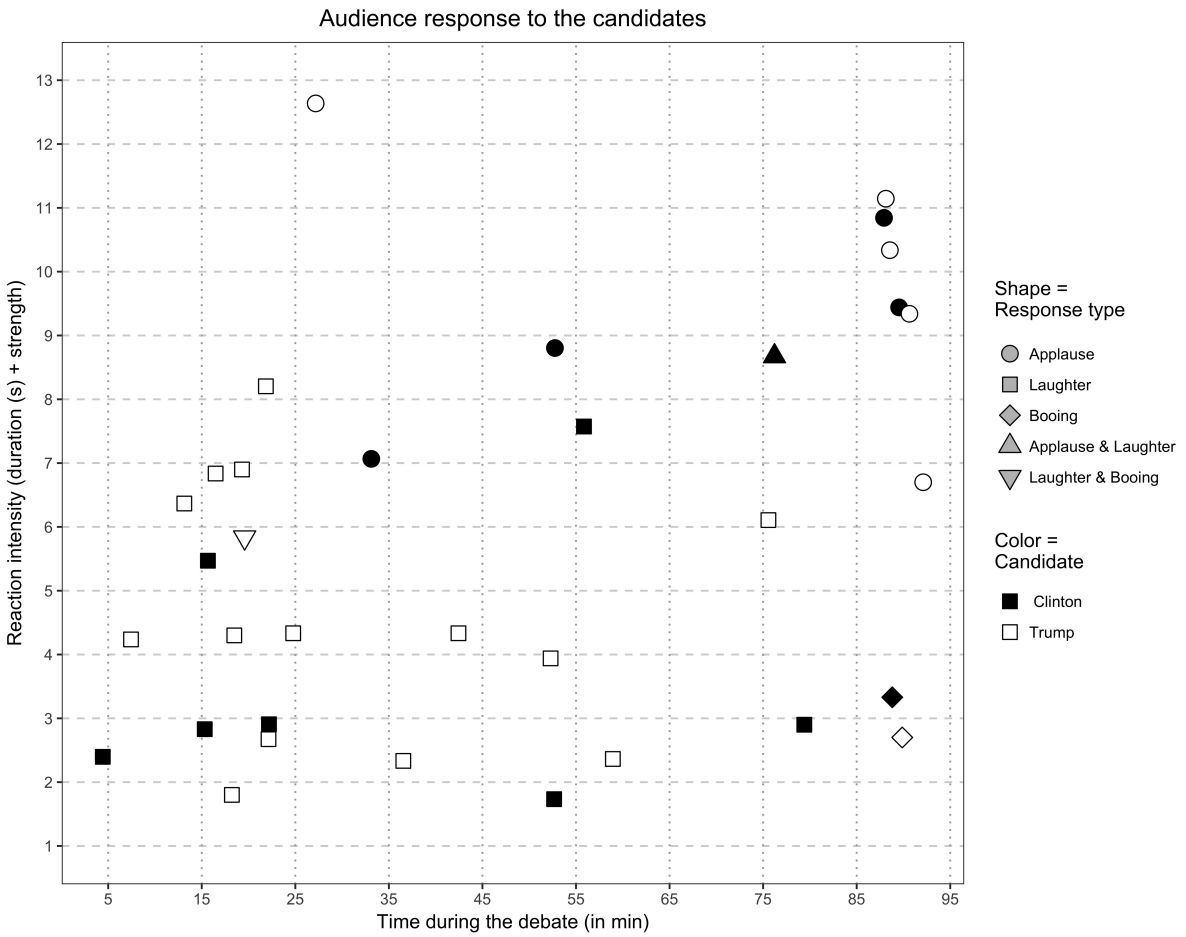


FIGURE 1 | Studio audience OAR to Hillary Clinton and Donald Trump during first 2016 presidential debate.

booing, with a third occurrence of booing in conjunction with laughter.

The norms of civility respected in previous presidential debates by the audience through their OAR were not followed in the first 2016 general election presidential debate. Arguably, the norm-bending behavior of Trump through his many interruptions and perhaps more importantly, his use of laughter-inducing rhetoric led to the studio audience departing from customary expectations concerning their collective behavior (Dailey et al., 2005). This is not to diminish Clinton's or moderator Lester Holt's role in audience actions. Clinton's attacks on Trump likely stirred a defensive group response from his supporters. When **Figure 1** is considered, Holt's lack of control over the audience can be seen with escalating incidence and intensity of laughter. This likely enabled the more consciously controlled studio audience applause–cheers in response to Trump's attack on Clinton's email controversy to occur.

The ability of both candidates to instigate OAR suggests similarities; however, there are revealing differences. Specifically, while both Trump and Clinton invited equal numbers of studio audience applause–cheering with four apiece, Trump was able to elicit five more studio audience OAR than Clinton. This was mainly through his laughter-eliciting attacks; he was also arguably more polarizing by eliciting boos-jeering in one case and a combination of laughter and boos in another instance. For her part, Clinton produced laughter followed by cheers in two cases, suggesting unconstrained support by her followers, especially in response to her attacks on Trump.

STUDY 2

Materials and Methods

Questions remain concerning the nature of the relationship between OAR by those in the studio audience and those watching the presidential debate on television, streaming on the internet, or listening on radio. Individuals hearing studio audience OAR in response to candidates utterances may potentially have also have experience intra-audience mediated effects through the OAR (Cummins and Gong, 2017) and been affected not just by the candidate statements (Fein et al., 2007). This intra-audience effect had the potential to affect millions, especially undecided voters, and more explicitly sets the stage for testing Research Questions 1–4.

Participants and Method

To understand the potential influence of both the candidate utterances and studio audience response on network viewers observing the televised debate, this study built from a field experiment being conducted at large university in the southern United States (Cavari et al., 2017). Participants were recruited from approximately 2,000 undergraduate students in more than 100 communication, political science, and psychology course sections from the researchers' home departments, received extra credit for taking part, and were not informed as to the study's purpose.

A total of 610 participants filled out an online omnibus survey prior to the debate (between August 29 and September 26, 2016) and were randomly assigned to one of seven rooms after their identity was verified. Each room, which was built to hold from 46 to 138 individuals, presented a different network (*ABC, FOX News, MSNBC/NBC, CNN, NPR, CBS, C-SPAN*) to 42–57 participants in classrooms. Post-test survey data was collected immediately after the debate, but due to the unanticipated amount of OAR, was not usable due the ceiling effect on pertinent measures.

The debate was viewed by 362 participants who took part as specified by university IRB protocols. Usable post-debate data from the 341 participants who filled out and returned the post-debate survey showed the sample was composed of 64% females, had a mean age of 19.53 (SD = 2.71) and was predominantly Caucasian (83%; African American [6%], Hispanic [4%], Asian [3%], Native American [1%], the remainder self-identified as "other").

Politically approximately 77% reported being registered voters, half (50.1%) self-identified as Republicans, just over a quarter (27.6%) as Democrats, and the remainder as independent/non-affiliated (22%). Political ideology as measured on a 7-point Likert-type scale (1 = *very liberal;* 7 = *very conservative*) was normally distributed and slightly right of center (M = 4.30). Chi-square (p_{gender} = 0.59; p_{party} = 0.29; p_{race} = 0.85) and ANOVA ($p_{ideology}$ = 0.15; p_{age} = 0.16) analyses show no significant room differences suggesting successful random assignment.

The participants were observed in seven different on-campus classrooms by study volunteers drawn from University Honors students and graduate students in Communication, Political Science, and Psychological Science programs. All rooms had three observers positioned at both front and one back room corners except the room watching ABC, which had six observers due to the additional observers mistakenly reporting to the incorrect room. Additionally, one observer was removed from analysis for coding only two OAR, when the average was 28.13 (SD = 8.65). In addition to checking in the students and keeping order, these observers were instructed to identify and code the field study room OAR in terms of type (Applause/Clapping, Laughter, Booing/Jeers, Other response), time that it occurred (to the minute), the individual (Clinton, Trump, or moderator) eliciting the OAR, the perceived strength of the OAR (see Study 1), and a brief description of the evoking comment or action.

To analyze the co-occurrence of field study room OAR with studio audience OAR, in other words the intra audience media effects, data for each of the field study rooms were first considered in terms of what was being observed and measured before being aggregated for analysis. Thus, we initially consider how OAR is not necessarily experienced and coded in the same manner. First, an OAR may be experienced and coded as having greater strength due to the observers' proximity to the individual(s)' utterance, and not necessarily due to the entire room vocalizing at higher levels. Second, identification of OAR type, whether laughter, applause–cheering, booing, or combinations of these, may be influenced by the strength of the OAR itself. In either case, greater involvement from greater numbers of audience members might lead to either enhanced clarity of signal, or greater ambiguity.

Findings

Based upon the time of the occurrence and the comments, we were able to identify 113 unique OAR across the seven field study rooms with all showing a similar pattern (**Figure 2**), including which responses co-occurred with the studio audience (as noted in Study 1). While each of the field study rooms had multiple OAR that did not co-occur with those by the studio audience, we focused on those that represent a co-occurrence of audience response potentially signifying either shared response to candidate utterances or social contagion.

From this data, a clear pattern of agreement emerges: of the 321 verified field study room OAR correlating with candidate or moderator utterances, nearly four-fifths (n = 255; 79.4%) involved exclusively laughter. Of the other OAR, only 10 (3.1%) were distinguishable as solely applause–cheering (n = 8) or booing (n = 2). The remaining room responses were either identified as a mixture of applause–cheering and laughter (n = 1), laughter and booing (n = 3), an unidentified mixture (n = 41) or as no selection/other (n = 11). Thus we aggregated these responses into an "other" category.

While it is apparent that laughter predominated and was the most easily identified of OAR, with from one-to-three coders (or in the case of the ABC room, one-to-six coders) in each room, the level of agreement does not necessarily reflect ICR so much as the location of the coder and the individual(s) audibly responding to the debate and the strength of the OAR itself. For instance, while when laughter occurred there was strong inter-observer agreement, the other types of OAR rarely resulted in agreement. There may be a notable relationship between the observers distinctively hearing laughter, applause, booing, or mixtures of these responses due to position in the room. Some coders may perceive one type of OAR as more prominent due to proximity to the audible response within the field study room. As such, inter-coder approaches typically used with content analysis (e.g., Cronbach's alpha, Krippendorff's alpha) are not appropriate; instead, we develop a variable of cumulative strength. Cumulative strength thus considers the OAR occurring in each room and creates an index where each of the observers, using the 1–5 strength scale used in Studies 1 and 2, add their scores together. Next, due to the disparity in the number of coders across all rooms, cumulative

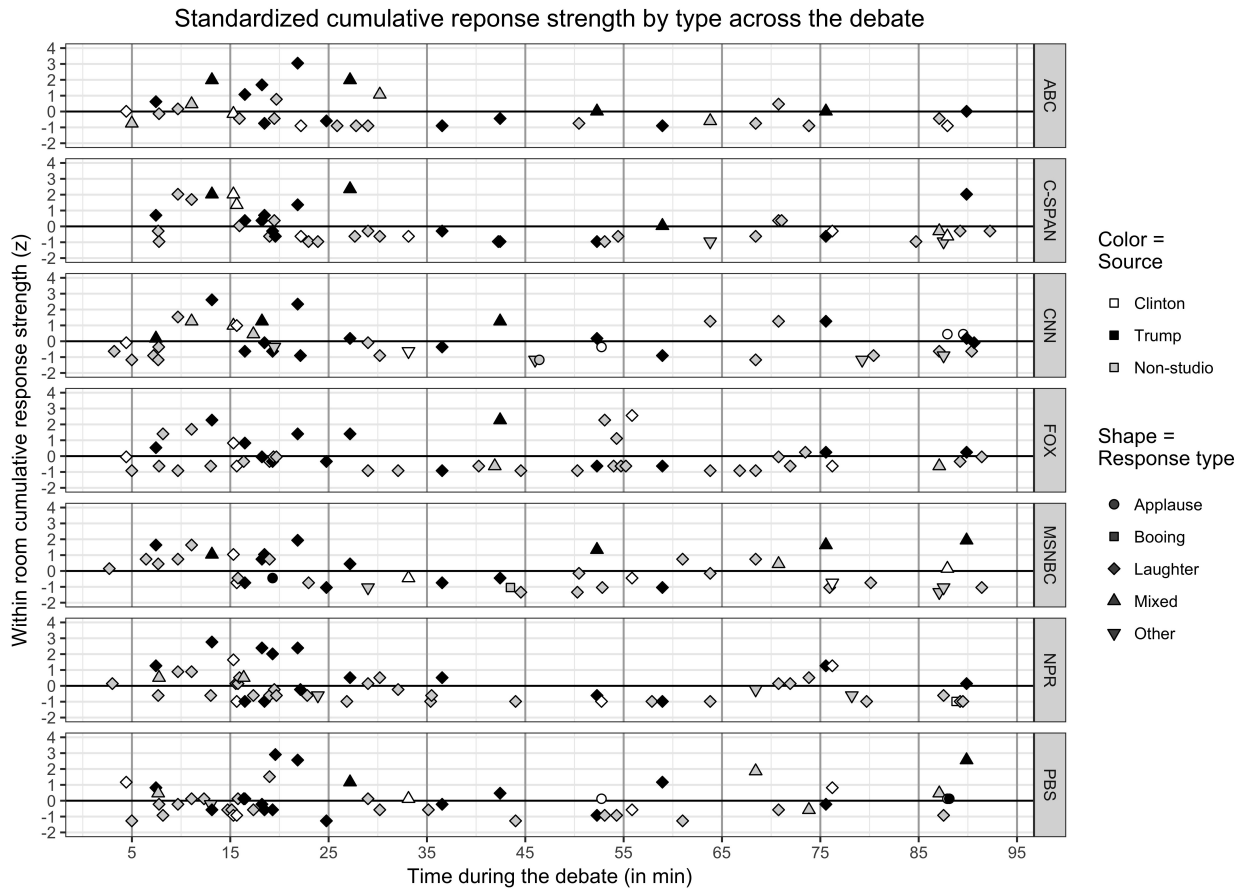


FIGURE 2 | Field study OAR to Hillary Clinton and Donald Trump during first 2016 presidential debate.

strength was standardized within each respective room by creating z-scores allowing us to compare across treatment rooms.

Co-occurrence of Studio and Field Study OAR: Full Sample

In keeping with the previous studies, and due to statistical reasons, we do not consider co-occurrence of studio audience and field study audiences deriving from studio audience applause-and-laughter (n = 5 rooms), laughter-and-booing (n = 2 rooms), or booing (n = 8 rooms). This leaves us with a total of 306 field study OAR in the seven rooms characterized based upon type of OAR (laughter = 244; mixed = 62) and cumulative strength. This allows us to consider the influence of studio audience OAR intensity and type (laughter = 109; applause = 22; no response = 175) on field study room OAR.

When this categorical data is analyzed we find a highly significant relationship between types of studio audience and field study OAR co-occurring, $\chi^2(2,306) = 27.790$, $p < 0.001$, with a moderately strong relationship (Cramer's $V = 0.301$). Specifically, marginally more field study audience laughter was observed in the seven rooms than was expected when studio audience laughter occurred (5.1) or there was no studio audience OAR (4.5). However, there were substantially fewer laughter responses in the field study room when there was studio audience applause (−9.5).

To assess the effect of the studio audience OAR on cumulative strength of field study OAR, we ran 3 (type of studio audience OAR: laughter, other, no response) × 7 (field study room) ANOVA on cumulative strength of OAR in field study rooms. Findings suggest the difference in the type of OAR was highly significant and had a strong effect [$F(2,285) = 28.904$, $p < 0.001$, $\eta_p^2 = 0.806$]. Neither the field study room [$F(6,285) = 0.535$, $p = 0.780$, $\eta_p^2 = 0.050$] nor was the interaction between OAR and field study room [$F(12,285) = 0.570$, $p = 0.865$, $\eta_p^2 = 0.023$] significant.

Post hoc analysis of the effect of the different types of OAR (applause–cheering vs. laughter) on the standardized cumulative strength of response in the field study rooms found that studio audience applause–cheering ($p < 0.01$; $M = 0.352$, $SD = 0.233$) and laughter ($p < 0.001$; $M = 0.352$, $SD = 0.090$) was significantly stronger than when there was no studio audience OAR ($M = -0.295$, $SD = 0.072$). At the same time, there was no difference between applause–cheering and laughter (p = ns).

Co-occurrence of Studio and Field Study OAR: Truncated Sample

Finally, to assess the influence of the intensity of the studio audience OAR on field study room OAR we considered only those cases in which there was a co-occurrence of studio audience and field study OAR. This leaves us with a truncated sample of 131 events. To consider the effect of the studio audience OAR type and intensity on the field study's OAR type and the standardized cumulative strength, we carried out a binary logistic regression and an ANCOVA, respectively. Both equations include the studio audience OAR intensity index as a covariate with the type of studio audience OAR (laughter or other) as a between-subjects factor.

The binary logistic regression analysis considered the field study audience rooms laughter or other OAR type was predicted by studio audience laughter or applause and the intensity of their response. The full model was significant $\chi^2(1) = 20.495$, $p < 0.001$, and moderately strong (Cox and Snell $R^2 = 0.145$ and Nagelkerke $R^2 = 0.218$). Analysis of the variables suggest that while the intensity index was not significant, Wald $\chi^2 = 0.347$, $p = 0.556$, studio audience OAR type was significant Wald $\chi^2 = 9.285$, $p < 0.01$. Studio audience OAR predicted field study laughter correctly 92% of the time (92/100) and other types of response 45.2% (14/31).

Analysis of the effect of studio audience OAR type and intensity on field study room OAR standardized cumulative strength, on the other hand, suggest both variables have influence. Findings show the studio audience OAR intensity index was significant, had a small effect, and was positively related to field study OAR ($F = 18.179$, $p < 0.001$, $\eta_p^2 = 0.124$). The effect of the studio audience OAR type was likewise significant and had a small effect ($F = 12.117$, $p < 0.01$, $\eta_p^2 = 0.086$) with studio audience OAR laughter ($M = 0.343$, $SD = 1.128$) having a stronger influence than applause–cheering ($M = 0.224$, $SD = 0.873$).

Discussion

Despite taking a conservative approach regarding our analysis of co-occurring studio and field study OAR by not including those studio audience events where applause–cheering followed and combined with laughter, our findings indicate laughter was more evident in the field study rooms than in the studio audience. When co-occurring with studio audience OAR, there was a moderately strong relationship between the type of studio audience OAR (laughter or applause)/non-response and the field study audiences OAR type, with applause–cheering significantly less likely to co-occur with laughter.

Furthermore, the more stereotypical signaling nature of laughter, when compared with other types of OAR, is apparent even when taking into account the "success" of candidate utterances (as indexed through studio audience audible intensity). This may be seen as indicating laughter, even when aggregated in OAR, being more automatic and stereotyped when compared with all other responses, even when considering observational judgments.

While the findings are illuminating, it should be noted that younger audiences such as studied here will likely laugh more due to social pressures, such as the implicit lack of knowledge concerning the status/rank of those around them (Mehu and Dunbar, 2008; Mehu, 2011). Younger individuals might be more likely to behaviorally mimic others (Sachisthal et al., 2016; Moody et al., 2017), especially if they appraise themselves as belonging to the implicit in-group (Platow et al., 2005; Sachisthal et al., 2016). As can be seen in **Figure 2**, the greatest amount of laughter, both concurrently with the studio audience and independent of them, occurred across all seven field study rooms after 5 min of relative quiet and appeared to be clustered in the first 20–25 min of the debate. In this case, participants likely signaled themselves as belonging to the peer group as a fellow student by laughing (relatively) early and often. While student participants might be more likely to mimic others around them, they do not necessarily experience the emotional contagion resulting in attitudinal change toward the candidates. To assess this, Study 3 considers the influence of studio audience OAR on how well individuals like the candidates.

STUDY 3

Methods

Participants were recruited from a west Texas community as part of an election study announced on the local newspaper's website. Due to continuous response theater using dedicated wireless dials, sample size was limited to 34 participants—the maximum number the room could accommodate during the debate. Partisan identification was divided between 14 Republican Party identifiers, 11 Independents, and 9 Democratic Party identifiers. Participants received a small monetary inducement in exchange for their participation. Age ranged from 18 to 73 ($M = 36.60$, $SD = 17.88$) with a slight majority of participants ($n = 19$, 54.3%) male.

The dependent variable, *candidate evaluation*, was derived from participants' moment-to-moment (MTM) response to the speaking candidate using the DialSmith Perception Analyzer 8.0 through wireless handheld response dials. When watching the debate, participants used their dial to indicate their agreement to the statement, "I like the candidate who is speaking," with response options ranging from 0 (*Strongly Disagree*) to 100 (*Strongly Agree*). Prior to the debate beginning, participants were asked to set their dials to the scale's mid-point of 50.

To calculate participant response to studio audience laughter and applause–cheering during the debate, the MTM responses 10 s prior to the onset of studio audience OAR provided a baseline average from which deviations up to 5 s afterward were considered. Thus, positive MTM change scores represent a more favorable attitude toward the candidate. The first 5 s after the onset of OAR was analyzed in order to account for potential delayed MTM reaction to OAR, as well as the average duration of OAR lasting roughly 2–3 s.

Nineteen studio audience OAR comprised of laughter and 11 of applause–cheering identified in Study 1 are considered, with overlapping or indistinct OAR removed from analysis. Of these, nine studio audience laughter segments and five applause–cheering OAR occurred during or after Hillary Clinton's

comments, while 10 laughter and 6 applause–cheering OAR occurred during or after Donald Trump's comments.

Findings

To address the research questions, an omnibus 2 (studio audience OAR: Laughter v. Applause) × 3 (partisan affiliation: Democratic v. Republican v. Independent) × 5 (Time) repeated-measures ANOVA was conducted. Because all participants evaluated every studio audience OAR, studio audience OAR and time (i.e., change scores for the 5 s after onset of laughter or applause) served as the within-subjects repeated measure. Political affiliation served as the sole between-subjects variable.

The main effect of studio audience OAR on MTM response in the continuous response theater was not significance [$F(2,980) = 3.14$, $p = 0.08$, $\eta_p^2 = 0.003$]. As seen in **Figure 3**, although studio audience applause elicited more positive MTM response than did laughter, this difference was not statistically significant. This is possibly due to the fact that participants were instructed to give a general evaluation of the speaking candidate, and studio audience OAR was only one of the many factors that influenced real-time candidate evaluation during the presidential debates. Also, this finding suggests that laughter is not necessarily associated with candidate evaluation. While contrary to expectations from Fein et al. (2007), the context is different with strong feelings already held toward the two candidates likely affecting MTM response.

The main effect of political affiliation on MTM responses was significant [$F(2,980) = 4.94$, $p = 0.007$, $\eta_p^2 = 0.01$]. *Post hoc* analysis showed Independent participants' MTM responses didn't significantly differ from Democrat ($p = 0.21$) and Republican participants ($p = 0.15$). The significant main effect of political affiliation on MTM responses was primarily driven by the difference between Republican and Democrat participants ($p = 0.007$). To provide a closer examination on the impact of participants' political affiliation on MTM responses, a series of follow-up analyses were conducted. When the studio audience applauded-cheered, a significant difference in MTM response was found between participants based upon political party affiliation [$F(2,337) = 3.34$, $p = 0.04$, $\eta_p^2 = 0.02$]. Specifically, when studio audience applause occurred in response to Clinton's comments, a significant difference in the continuous response theater participants was found between the three political affiliations [$F(2,167) = 11.83$, $p < 0.001$, $\eta_p^2 = 0.12$]. As seen in **Figure 4**, while studio audience applause–cheering elicited more positive MTM responses among Democrat and Independent participants, Republican participants' MTM responses became more negative when studio audience applauded-cheered for Clinton. Interestingly, no significant difference between participant MTM response based upon political party affiliation when studio audience applause–cheering occurred in reaction to Trump's comments, $F(2,167) = 1.56$, $p = 0.21$.

After studio audience laughter, participant MTM response didn't significantly differ between the three political affiliations [$F(1,643) = 1.22$, $p = 0.30$]. These follow-up analyses indicated the main effect of political affiliation on MTM response was primarily

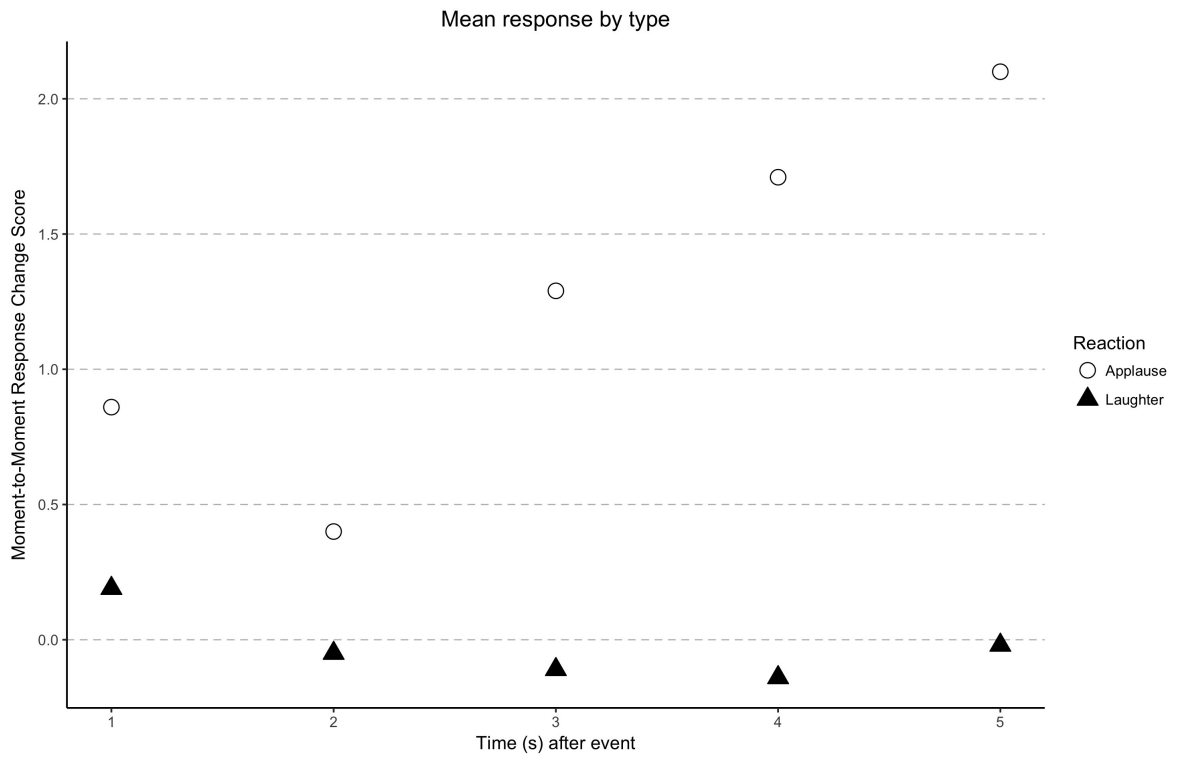


FIGURE 3 | Main effect of studio audience OAR on moment-to-moment response 1–5 seconds after onset.

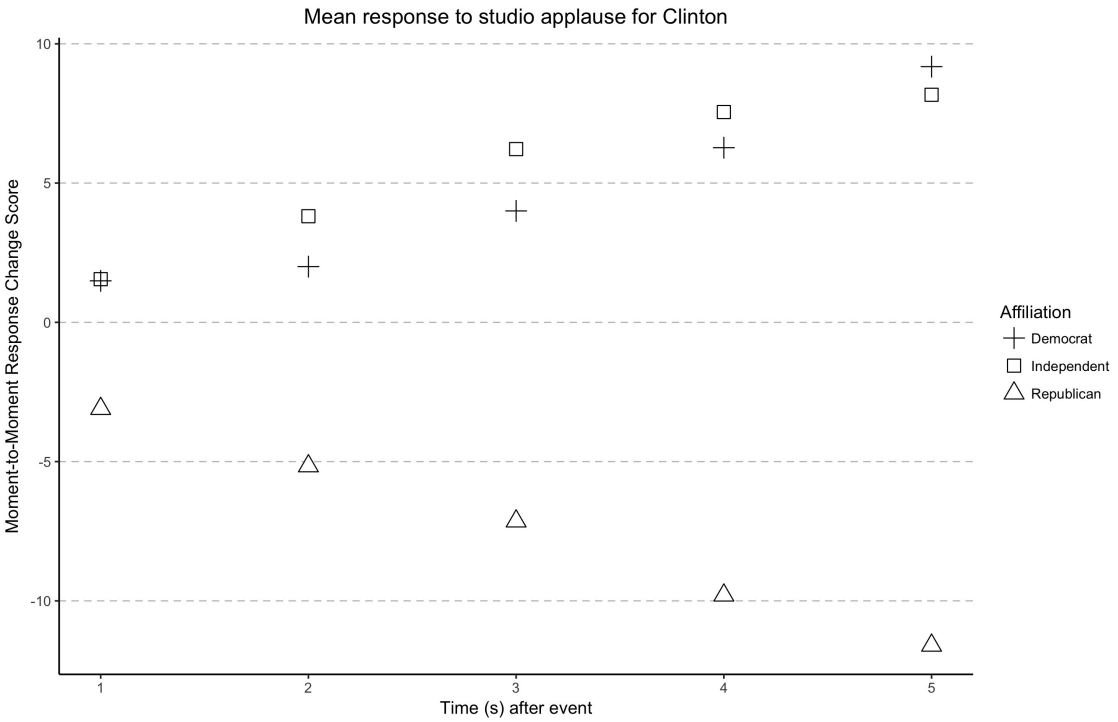


FIGURE 4 | Partisan moment-to-moment response to applause-cheering for Clinton 1–5 seconds after onset.

due to the difference between Democrat and Republican participants, especially their MTM response to applause instead of laughter. Finally, the studio audience OAR by political affiliation interaction was not significant [$F(2,980) = 2.38$, $p = 0.09$].

Discussion

Our findings suggest that studio audience applause–cheering had an effect on continuous response study participant candidate evaluations, whereas laughter did not, and that political party affiliation further clarified differences in how likeable the candidates were perceived; however, these findings might not adequately reflect the influence of OAR type. First, this study's sample was quite small at less than one-third of the comparable studies by Fein et al. (2007) (2 and 3), with statistical power diminished further by small numbers of partisans. Second, our study was carried out during a high stakes election high in a polarized political environment where both candidates were equally likely to win. Finally, and perhaps most important, as can be seen in **Figure 1**, the intensity of studio audience applause–cheering was stronger for most all of their response to both Trump and Clintons' comments than was that of laughter, making direct comparisons difficult. Fein and colleague's laboratory studies, while comparable by using continuous response measurement to evaluate response to United States President Ronald Reagan and Minnesota Senator Walter Mondale during their 1984 debates, considered only two studio audience OAR with combined laughter and applause–cheering, and an observable audible and visible reaction from the moderator in one of the instances. Thus, while not as easily parsed as planned laboratory experiments, Study 3 in combination with findings from Study 2, provide real-time evidence of the differential effects of studio OAR on mediated viewers.

GENERAL DISCUSSION

While not as tidy as laboratory experiments, we believe that the enhanced generalizability of our analyses of multiple studies by using ethological methods most proximately building on those pioneered by Robert Provine in his research on laughter (Provine, 2001, 2015) allows for greater and unique insights than provided by other more traditional approaches. Here, we take the position initially promoted by John Wahlke in his 1979 American Political Science Association presidential address and echoed most recently by de Gelder (2017) by asserting that the "prebehavioral" tendencies in social science research, with an emphasis on self-report, miss what the "small data" we use captures (Wahlke, 1979). While we use both approaches throughout our project to triangulate our findings, by focusing on behavioral responses which are more visceral, automatic, and tied to our primate ancestors' behavior, audible non-verbal utterances such as laughter, applause–cheering, and booing might best reflect behavioral intent of individuals as part of a group.

Observable audience responses such as laughter, applause–cheering, and booing are important because they reflect the

emergent properties of individuals becoming groups. While the research reported here does not purport to explain OAR or appraise intent, it makes an important first step in providing evidence concerning individual humans engaging in the group behaviors of applause–cheering, laughter, and (to an extent) booing. In addition to serving the more theoretical purposes of understanding social identity with its evolutionary roots of followership and in-group vs. out-group identities (Haslam et al., 2010; Van Vugt and Ahuja, 2011) and with it the reliability of non-verbal signals (Mehu et al., 2011) inherent in laughter, applause–cheering, and booing (as well as mixtures of these), the research carried out here serves the more proximate and practical needs of understanding the appeal of populist politicians such as Donald Trump, especially in comparison with more traditional candidates. And while we did not systematically explore the booing that occurred, by focusing on the occurrence and effect of laughter and applause–cheering, we have been able to better discriminate between them in terms of form and function.

Findings regarding the specific research questions posited and evaluated in Study 2 suggests that there is a moderately strong relationship between not just the studio audience and the field study audience OAR, answering Research Question 1, but also between laughter occurring in the studio audience and in the field study rooms. When the truncated model was considered, allowing for us to control for studio audience OAR intensity, we found laughter in the studio audience was more strongly related with field study room laughter than applause was with the "all other types" category we used for the field study rooms. This provides evidence responding to Research Question 3. However, while there was modest evidence for Research Question 2, as studio audience OAR intensity was weakly related with field study room cumulative OAR strength, we find, regarding Research Question 4, that there is not a significant relationship between studio audience intensity and OAR type.

The differential response to studio audience OAR was further probed by continuous response measurement (CRM) of MTM liking of the speaking candidate. This allows us to move beyond our research questions to more directly draw inferences. The greater amounts of studio audience laughter elicited by Trump in comparison with Clinton may have affected unaffiliated viewer perceptions by evoking the behavioral mimicry that presumably occurs before social contagion. However, the applause–cheering evoked by Trump may have mattered more, as well as the intensity of the evoked studio audience OAR. Specifically, it appears that the likability of Trump was positively affected by audience applause–cheering to a significantly greater extent than laughter with the CRM study, and that the applause–cheering for Trump was more effective than that elicited by Clinton. In combination with the observational studies regarding the field study, the lack of studio audience control by the moderator may have affected viewer perceptions not just through the stereotypical laughter that is mimicked near automatically, but also by the applause–cheering and mixed audience responses that increase their likability to partisans.

FUTURE RESEARCH

While the information found through the three studies regarding the first general election debate of 2016 helps clarify the role group response in the form of OAR plays, a series of broader questions remain. Specifically, it has been established that individual laughter is a "costly signal" involving abrupt eruptions of distinctive vocalizations concomitant with physiological and emotional change (Grammer and Eibl-Eibesfeldt, 1990; Weisfeld, 1993; Ruch and Ekman, 2001; Gervais and Wilson, 2005; Panksepp, 2007). This might be due to the multi-channel nature of this display; in addition to the vocalic qualities of laughter, distinct facial display signature become evident and co-occur with the laughter (Platow et al., 2005; Mehu and Dunbar, 2008; Mehu, 2011; Stewart et al., 2015). Together, the amusement smiles and audible utterances may be used to differentiate between different types of positive emotional states by how often they occur (Hofmann et al., 2017). Cheering and booing, for their part, both seemingly involve distinctive facial displays co-occurring with the audible utterances. Despite their being more consciously chosen, these types of OAR likely may still lead to change in emotional state and behavioral intent by engaging in two non-verbal channels. However, the influence of applause – which involves only rhythmic hand-and-arm movements – may not necessarily be as reliable an index of individual involvement. At the very least, research should consider more fully the facial display behavior co-occurring with all vocalizations inherent in OAR.

Likewise, questions still remain regarding how individual responses aggregate into a group response. In other words, applause–cheering, laughter, and booing apparently are mimicked, albeit at different levels based upon the audience, and may potentially be socially contagious. As seen in this study, the shared, and potentially mimicked and contagious experience of co-occurring OAR between the studio audience and the field study rooms raises questions. The first, and perhaps foremost, concerns which form of OAR is more likely to lead to group coordination in the form of greater support for goals as stated by the speaker, as well as support for the leader herself or himself. Specifically, while laughter appears to be more likely to be shared than applause–cheering, the nature of booing is not as well established due in great part to its rarity.

At the very least, Studies 1 and 2 suggest a high level of mimicry by individuals, especially regarding laughter. Here, mimicry is defined as the quick and spontaneous matching (within 1 s) of another person's display behavior and linked with empathy and prosocial behavior (Sachisthal et al., 2016; Moody et al., 2017). Mimicry is thus highly important for social functioning such as group coordination. Social contagion, on the other hand, may be seen as a higher order concept with mimicry being an initial step in an appraisal process whereas individuals assess not just the behavior they are mimicking but also consider their social context (Hatfield et al., 2014). What happened with both the studio and field study audiences with their laughter, however, may reflect mimicry more so than social contagion. This is because social contagion involves appraisal of such factors as social context and group membership

(Hatfield et al., 1994, 2014; Lakin et al., 2003), as was seen in the CRM study. On the other hand, field study participants laughing at comments by candidates they did not support (or indeed were predisposed against), could merely be considered mimicry. Whether this ultimately led to social contagion is beyond the purview of this research project; however, it is an important next step in research best considered through more diverse and precise measurement.

A further question concerns whether there are optimal audience sizes for these different forms of OAR; in other words, there tends to be a greater likelihood of applause–cheering, laughter, and booing based upon the increasing size of a group in a form of mutual "grooming" (Dezecache and Dunbar, 2012). However, while evidence suggests that laughter can be a form of mutual grooming amongst two and more individuals (Provine, 2001, 2015) questions remain concerning the numbers of individuals requisite for applause–cheering and booing to occur. Furthermore, there is the question concerning when the group reaches a threshold, will there be a greater likelihood of groups "factioning off" – especially if they are proximate with each other as identifiable entities with separate putative leaders. Furthermore, and related to all the foregoing questions, the mechanism by which individuals are influenced, whether physiological, appraisal-oriented, or emotionally driven group contagion, provides questions to explore in greater detail with a range of different methodologies.

Future research thus should be able to better disambiguate not only the audible signal of group response, but also understand attitudinal and behavioral change. Advances in technology should allow for more precise measurement than that carried out here by naïve judges with limited training. Specifically, audio recorders (including smart phones) placed throughout the room might allow for more accurate notation of OAR timing, type, and intensity, even to the individual level. Indeed, as seen with acoustic research regarding laughter, the different utterances might have a range of signal qualities that are not being considered in needed detail. Just as laughter itself may embody many different emotional messages by reflecting the responses of many different individuals, the resulting message may "get lost in the crowd." Therefore, by understanding more perfectly the union in OAR such as laughter, applause–cheering, booing, and their combinations, we may be able to divine a greater understanding of the most fundamental of human social activities – politics.

NOTES

Previously presented at the 75th Annual Midwest Political Science Association Conference, April 5–9, 2017, Chicago, IL, United States.

ETHICS STATEMENT

This study (IRB Protocol #: 16-07-029: "The 2016 Presidential Election: Attitudinal Change in Response to Campaign Events, Debates, and Electoral Results") was carried out in accordance with United States Federal Regulations concerning research [45 CFR 46.102(d)] and human subjects [45 CFR 46.102(f)] as implemented by the University of Arkansas, Fayetteville Office of Research Compliance Institutional Review Board. All participants were given written informed consent in accordance with these United States Federal Regulations and the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

PS: data collection, data cleaning, theory building, writing, and data analysis. AE: data collection, data cleaning, data analysis, and figures. RD: data collection, data cleaning, and editing. ZG: data collection, data cleaning, data analysis, writing, and figures. EB: data collection, data analysis, writing, and editing. RW: data collection and editing. SE: data collection.

REFERENCES

Atkinson, M. (1984). *Our Masters' Voices: The Language and Body-Language of Politics*. New York, NY: Routledge.

Axsom, D., Yates, S., and Chaiken, S. (1987). Audience response as a heuristic cue in persuasion. *J. Pers. Soc. Psychol.* 53, 30–40. doi: 10.1037/0022-3514.53.1.30

Bachorowski, J., and Owren, M. J. (2001). Not all laughs are alike: voiced but not unvoiced laughter readily elicits positive affect. *Psychol. Sci.* 12, 252–257. doi: 10.1111/1467-9280.00346

Benoit, W. L. (2013). *Political Election Debates: Informing Voters About Policy and Character*. New York, NY: Lexington Books.

Benoit, W. L., McKinney, M. S., and Holbert, R. L. (2001). Beyond learning and persona: extending the scope of presidential debate effects. *Commun. Monogr.* 68, 259–273. doi: 10.1080/03637750128060

Bonin, F., Campbell, N., and Vogel, C. (2014). Time for laughter. *Knowl. Based Syst.* 71, 15–24. doi: 10.1016/j.knosys.2014.04.031

Bucy, E. P., and Stewart, P. A. (Forthcoming). "The personalization of campaigns: non-verbal cues in presidential debates," in *Oxford Research Encyclopedia of Politics*, ed. W. R. Thompson (New York, NY: Oxford University Press).

Bull, P. (2003). *The Microanalysis of Political Communication: Claptrap and Ambiguity*. New York, NY: Routledge.

Bull, P., and Miskinis, K. (2014). Whipping it up! An analysis of audience responses to political rhetoric in speeches from the 2012 american presidential elections. *J. Lang. Soc. Psychol.* 34, 521–538. doi: 10.1177/0261927X14564466

Bull, P., and Wells, P. (2002). By invitation only? an analysis of invited and uninvited applause. *J. Lang. Soc. Psychol.* 21, 230–244. doi: 10.1177/0261927X14564466

Cavari, A., Powell, R. J., and Mayer, K. R. (2017). "The 2016 presidential election," in *The Causes and Consequences of a Political Earthquake*, eds A. Cavari, R. J. Powell, and K. R. Mayer (Lanham, MD: Lexington Books). doi: 10.1177/0261927X02021003002

Choi, H., Bull, P., and Reed, D. (2016). Audience responses and the context of political speeches. *J. Soc. Polit. Psychol.* 4, 601–622. doi: 10.5964/jspp.v4i2.618

Clayman, S. E. (1992). Caveat orator: audience disaffiliation in the 1988 presidential debates. *Q. J. Speech* 78, 33–60. doi: 10.1080/00335639209383980

Clayman, S. E. (1993). Booing: the anatomy of a disaffiliative response. *Am. Sociol. Rev.* 58, 110–130. doi: 10.2307/2096221

Clayman, S. E. (1995). Defining moments, presidential debates, and the dynamics of quotability. *J. Commun.* 45, 118–147. doi: 10.1111/j.1460-2466.1995.tb00746.x

Cummins, R. G., and Gong, Z. (2017). Mediated intra-audience effects in the appreciation of broadcast sports. *Commun. Sport* 5, 27–48. doi: 10.1177/2167479515593418

Dailey, W. O., Hinck, E. A., and Hinck, S. S. (2005). Audience perceptions of politeness and advocacy skills in the 2000 and 2004 presidential debates. *Argument. Advocacy* 41, 196–210. doi: 10.1080/00028533.2005.11821630

Davis, C. J., Bowers, J. S., and Memon, A. (2011). Social influence in televised election debates: a potential distortion of democracy. *PLoS One* 6:e18154. doi: 10.1371/journal.pone.0018154

de Gelder, B. (2017). Going native: emotion science in the 21th century. *Front. Psychol.* 8:1212. doi: 10.3389/fpsyg.2017.01212

Devereux, P. G., and Ginsburg, G. P. (2001). Sociality effects on the production of laughter. *J. Gen. Psychol.* 128, 227–240. doi: 10.1080/00221300109598910

Dezecache, G., and Dunbar, R. (2012). Sharing the joke: the size of natural laughter groups. *Evol. Hum. Behav.* 33, 775–779. doi: 10.1016/j.evolhumbehav.2012.07.002

Duck, S. W., and Baggaley, J. (1975). Audience reaction and its effect on perceived expertise. *Commun. Res.* 2, 79–85. doi: 10.1177/009365027500200105

Dunbar, R. I. M. (1993). Co-evolution of neocortex size, group size and language in humans. *Behav. Brain Sci.* 16, 681–735. doi: 10.1017/S0140525X00032325

Eibl-Eibesfeldt, I. (1989). *Human Ethology*. New York, NY: Aldine De Gruyter.

Ekman, P., and Friesen, W. V. (2003). *Unmasking the Face*. Cambridge, MA: Malor Books.

Fein, S., Goethals, G. R., and Kugler, M. B. (2007). Social influence on political judgments: the case of presidential debates. *Polit. Psychol.* 28, 165–192. doi: 10.1111/j.1467-9221.2007.00561.x

Fridlund, A. J. (2017). "The behavioral ecology view of facial displays, 25 years later," in *The Science of Facial Expression*, eds J. M. Fernández-Dols and J. A. Russell (New York, NY: Oxford University Press), 77–92. doi: 10.1093/acprof:oso/9780190613501.003.0005

Gervais, M., and Wilson, D. S. (2005). The evolution and functions of laughter and humor: a synthetic approach. *Q. Rev. Biol.* 80, 395–430. doi: 10.1086/498281

Gilmartin, E., Bonin, F., Vogel, C., and Campbell, N. (2013). "Laugher and topic transition in multiparty conversation," in *Paper Presented at the SIGDIAL Conference* (Saarland), 304–308.

Gong, Z. H., and Bucy, E. P. (2015). "Image bite analysis of presidential debates," in *Exploring the C-SPAN Archives: Advancing the Research Agenda*, ed. X. Browning Robert (West Lafayette, IN: Purdue University Press), 45–75.

Grammer, K., and Eibl-Eibesfeldt, I. (1990). "The ritualisation of laughter," in *Natürlichkeit Der Sprache Und Der Kultur*, ed. W. Koch (Bochum: Brock Meyer), 192–214.

Haslam, S. A., Reicher, S. D., and Platow, M. J. (2010). *The New Psychology of Leadership: Identity, Influence and Power*. New York, NY: Psychology Press.

Hatfield, E., Bensman, L., Thornton, P. D., and Rapson, R. L. (2014). New perspectives on emotional contagion: a review of classic and recent research on facial mimicry and contagion. *Interpersona* 8, 159–179. doi: 10.5964/ijpr.v8i2.162

Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1994). *Emotional Contagion (Studies in Emotion & Social Interaction)*. New York, NY: Cambridge University Press.

Heritage, J., and Greatbatch, D. (1986). Generating applause: a study of rhetoric and response at party political conferences. *Am. J. Sociol.* 92, 110–157. doi: 10.1086/228465

Hofmann, J., Platt, T., and Ruch, W. (2017). Laughter and smiling in 16 positive emotions. *IEEE Trans. Affect. Comput.* 8, 495–507. doi: 10.1109/TAFFC.2017.2737000

Hylton, C. (1971). Intra-audience effects: observable audience response. *J. Commun.* 21, 253–265. doi: 10.1111/j.1460-2466.1971.tb00922.x

Kipp, M. (2012). "Multimedia annotation, querying and analysis in ANVIL, chapter 19, multimedia information extraction," in *Video, Audio, and Imagery Analysis for Search, Data Mining, Surveillance and Authoring*, ed. M. T. Maybury (Hoboken, NJ: Wiley), 351–386.

Lakin, J. L., Jefferis, V. E., Cheng, C. M., and Chartrand, T. L. (2003). The chameleon effect as social glue: evidence for the evolutionary significance of nonconscious mimicry. *J. Nonverbal Behav.* 27, 145–162. doi: 10.1023/A:1025389814290

Mann, R. P., Faria, J., Sumpter, D. J., and Krause, J. (2013). The dynamics of audience applause. *J. R. Soc. Interface* 10:20130466. doi: 10.1098/rsif.2013.0466

Masters, R. D. (1989). *The Nature of Politics*. New Haven, CT: Yale University Press.

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., et al. (2015). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cereb. Cortex* 25, 246–257. doi: 10.1093/cercor/bht227

McKinney, M. S., and Warner, B. R. (2013). Do presidential debates matter? Examining a decade of campaign debate effects. *Argt. Advoc.* 49, 238–258. doi: 10.1080/00028533.2013.11821800

Mehu, M. (2011). Smiling and laughter in naturally occurring dyadic interactions: relationship to conversation, body contacts, and displacement activities. *Hum. Ethol. Bull.* 26, 10–28.

Mehu, M., and Dunbar, R. I. M. (2008). Relationship between smiling and laughter in humans (*Homo sapiens*): testing the power asymmetry hypothesis. *Folia Primatol.* 79, 269–280. doi: 10.1159/000126928

Mehu, M., Mortillaro, M., Bänziger, T., and Scherer, K. R. (2011). Reliable facial muscle activation enhances recognizability and credibility of emotional expression. *Soc. Psychol. Personal. Sci.* 2, 262–271. doi: 10.1037/a0026717

Moody, E. J., Reed, C. L., Van Bommel, T., App, B., and McIntosh, D. N. (2017). Emotional mimicry beyond the face? rapid face and body responses to facial expressions. *Soc. Psychol. Personal. Sci.*

Nagel, F., Maurer, M., and Reinemann, C. (2012). Is there a visual dominance in political communication? how verbal, visual, and vocal communication shape viewers' impressions of political candidates. *J. Commun.* 62, 833–850. doi: 10.1111/j.1460-2466.2012.01670.x

Néda, Z., Ravasz, E., Brechet, Y., Vicsek, T., and Barabási, A. (2000a). Self-organizing processes: the sound of many hands clapping. *Nature* 403, 849–850. doi: 10.1038/35002660

Néda, Z., Ravasz, E., Vicsek, T., Brechet, Y., and Barabási, A. (2000b). Physics of the rhythmic applause. *Phys. Rev. E* 61:6987.

Norton, M. I., and Goethals, G. R. (2004). Spin (and pitch) doctors: campaign strategies in televised political debates. *Polit. Behav.* 26, 227–248. doi: 10.1023/B:POBE.0000043454.25971.6a

Owren, M. J., and Bachorowski, J. (2003). Reconsidering the evolution of nonlinguistic communication: the case of laughter. *J. Nonverbal Behav.* 27, 183–200. doi: 10.1023/A:1025394015198

Panksepp, J. (2007). Neuroevolutionary sources of laughter and social joy: modeling primal human laughter in laboratory rats. *Behav. Brain Res.* 182, 231–244. doi: 10.1016/j.bbr.2007.02.015

Pätynen, J., and Lokki, T. (2016). Concert halls with strong and lateral sound increase the emotional impact of orchestra music. *J. Acoust. Soc. Am.* 139, 1214–1224. doi: 10.1121/1.4944038

Pellis, S. M., Pellis, V. C., and Himmler, B. T. (2014). How play makes for a more adaptable brain: a comparative and neural perspective. *Am. J. Play* 7:73.

Platow, M. J., Haslam, S. A., Both, A., Chew, I., Cuddon, M., Goharpey, N., et al. (2005). It's not funny if they're laughing: self-categorization, social influence, and responses to canned laughter. *J. Exp. Soc. Psychol.* 41, 542–550. doi: 10.1016/j.jesp.2004.09.005

Provine, R. R. (1992). Contagious laughter: laughter is a sufficient stimulus for laughs and smiles. *Bull. Psychon. Soc.* 30, 1–4. doi: 10.3758/BF03330380

Provine, R. R. (1993). Laughter punctuates speech: linguistic, social and gender contexts of laughter. *Ethology* 95, 291–298. doi: 10.1111/j.1439-0310.1993.tb00478.x

Provine, R. R. (2001). *Laughter: A Scientific Investigation*. New York, NY: Penguin press.

Provine, R. R. (2015). Laughter as a scientific problem: an adventure in sidewalk neuroscience. *J. Comp. Neurol.* 524, 1532–1539. doi: 10.1002/cne.23845

Rhea, D. M. (2012). There they go again: the use of humor in presidential debates 1960–2008. *Argument. Adv.* 49, 115–131. doi: 10.1080/00028533.2012.11821783

Ruch, W., and Ekman, P. (2001). "The expressive pattern of laughter," in *Emotion, Qualia, and Consciousness*, ed. A. W. Kaszniak (Tokyo: Word Scientific Publisher), 426–443.

Sachisthal, M. S., Sauter, D. A., and Fischer, A. H. (2016). Mimicry of ingroup and outgroup emotional expressions. *Compr. Results Soc. Psychol.* 1, 86–105. doi: 10.1080/23743603.2017.1298355

Salter, F. K. (2007). *Emotions in Command: Biology, Bureaucracy, and Cultural Evolution*. Piscataway, NJ: Transaction Pub.

Schroeder, A. (2016). *Presidential Debates: 40 Years of High Risk TV 3rd Edn*. New York, NY: Columbia University Press.

Schrott, P. R., and Lanoue, D. J. (2008). Debates are for losers. *PS Polit. Sci. Polit.* 41, 513–518. doi: 10.1017/S1049096508080682

Schubert, J. N. (1988). Age and active-passive leadership style. *Am. Polit. Sci. Rev.* 82, 763–772. doi: 10.2307/1962489

Schweingruber, D., and McPhail, C. (1999). A method for systematically observing and recording collective action. *Sociol. Methods Res.* 27, 451–498. doi: 10.1177/0049124199027004001

Scott, S. K., Lavan, N., Chen, S., and McGettigan, C. (2014). The social life of laughter. *Trends Cogn. Sci.* 18, 618–620. doi: 10.1016/j.tics.2014.09.002

Shah, D. V., Hanna, A., Bucy, E. P., Wells, C., and Quevedo, V. (2015). The power of television images in a social media age: linking biobehavioral and computational approaches via the second screen. *Ann. Am. Acad. Polit. Soc. Sci.* 659, 225–307. doi: 10.1177/0002716215569220

Silverthorne, C. P., and Mazmanian, L. (1975). The effects of heckling and media of presentation on the impact of a persuasive communication. *J. Soc. Psychol.* 96, 229–236. doi: 10.1080/00224545.1975.9923288

Sloan, L. R., Love, R. E., and Ostrom, T. M. (1974). Political heckling: who really loses? *J. Personal. Soc. Psychol.* 30, 518–525. doi: 10.1037/h0037023

Stewart, P. A. (2012). *Debatable Humor: Laughing Matters on the 2008 Presidential Primary Campaign*. Lanham, MD: Lexington Books.

Stewart, P. A. (2015). Polls and elections: do the presidential primary debates matter? measuring candidate speaking time and audience response during the 2012 primaries. *Pres. Stud. Q.* 45, 361–381. doi: 10.1111/psq.12191

Stewart, P. A., Bucy, E. P., and Mehu, M. (2015). Strengthening bonds and connecting with followers: a biobehavioral inventory of political smiles. *Polit. Life Sci.* 34, 73–92. doi: 10.1017/pls.2015.5

Stewart, P. A., Eubanks, A. D., and Miller, J. (2016). "Please clap": applause, laughter, and booing during the 2016 GOP presidential primary debates. *PS Polit. Sci. Polit.* 49, 696–700. doi: 10.1017/S1049096516001451

Szameitat, D. P., Alter, K., Szameitat, A. J., Darwin, C. J., Wildgruber, D., Dietrich, S., et al. (2009). Differentiation of emotions in laughter at the behavioral level. *Emotion* 9, 397–405. doi: 10.1037/a0015692

Szameitat, D. P., Darwin, C. J., Wildgruber, D., Alter, K., and Szameitat, A. J. (2011). Acoustic correlates of emotional dimensions in laughter: arousal, dominance, and valence. *Cogn. Emot.* 25, 599–611. doi: 10.1080/02699931.2010.508624

van Hooff, J. A. R. A. M., and Preushoft, S. (2003). "Laughter and smiling: the intertwining of nature and culture," in *Animal Social Complexity: Intelligence, Culture, and Individualized Societies*, eds F. B. De Waal and P. L. Tyack (Cambridge, MA: Harvard University Press), 260–287.

Van Vugt, M., and Ahuja, A. (2011). *Naturally Selected: The Evolutionary Science of Leadership*. New York, NY: Harper Collins.

Van Vugt, M., Hardy, C., Stow, J., and Dunbar, R. (2014). "Laughter as social lubricant: a biosocial hypothesis about the pro-social functions of laughter and humor," in *Paper Presentation Centre for the Study of Group Processes Working Paper* (Canterbury: University of Kent).

Wahlke, J. C. (1979). Pre-behavioralism in political science. *Am. Polit. Sci. Rev.* 73, 9–31. doi: 10.2307/1954728

Weisfeld, G. E. (1993). The adaptive value of humor and laughter. *Ethol. Sociobiol.* 14, 141–169. doi: 10.1016/0162-3095(93)90012-7

Wells, P., and Bull, P. (2007). From politics to comedy a comparative analysis of affiliative audience responses. *J. Lang. Soc. Psychol.* 26, 321–342. doi: 10.1177/0261927X07306978

West, D. M. (1984). Cheers and jeers candidate presentations and audience reactions in the 1980 presidential campaign. *Am. Polit. Res.* 12, 23–50.

Wiegman, O. (1987). Attitude change in a realistic experiment: the effect of party membership and audience reaction during an interview with a Dutch politician. *J. Appl. Soc. Psychol.* 17, 37–49. doi: 10.1111/j.1559-1816.1987.tb00291.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Stewart, Eubanks, Dye, Gong, Bucy, Wicks and Eidelman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 February 2018
doi: 10.3389/fpsyg.2018.00150

Extraversion Is a Mediator of Gelotophobia: A Study of Autism Spectrum Disorder and the Big Five

*Meng-Ning Tsai, Ching-Lin Wu, Lei-Pin Tseng, Chih-Pei An and Hsueh-Chih Chen**

Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei, Taiwan

Previous research has shown that individuals with autism are frequently mocked in their childhood and are consequently more anxious about being ridiculed. Research has also shown that autistic individuals have a higher level of gelotophobia (fear of being laughed at) compared to typically developed individuals. However, recent studies have also found that gelotophobia is strongly related to personality, which suggests that personality is a factor that helps to create a higher level of gelotophobia in autistic individuals. To investigate whether this is the case, we recruited 279 Taiwanese high school students, 123 with autism spectrum disorder (ASD) and 156 typically developed students as a control group. Self-reporting questionnaires were used to gather data on the Big Five personality traits and on the gelotophobia-related traits of gelotophobia, gelotophilia, and katagelasticism. The results were analyzed and the two groups were compared for differences in gelotophobia and personality. The ASD group was found to have a higher level of gelotophobia than the typically developed group, but lower levels of gelotophilia and katagelasticism. Additionally, the ASD group was found to have lower levels of extraversion and agreeableness than the typically developed group, but no significant difference was found between the two groups in terms of conscientiousness, openness, and emotional stability. We then investigated the possible correlations between gelotophobia-related traits and the Big Five, and consequently the mediation effect of the Big Five on gelotophobia. The results show, firstly, that extraversion rather than ASD is a direct factor in gelotophobia. Secondly, the level of gelotophilia was partly influenced by autism but also to a certain extent by the level of extraversion. Lastly, the results indicate that autism and the level of agreeableness are in conflict when predicting the level of katagelasticism.

Keywords: autism spectrum disorder, laughter, gelotophobia, gelotophilia, katagelasticism, Big Five personality traits

OPEN ACCESS

Edited by:
Tim Bogg,
Wayne State University, United States

Reviewed by:
Jennifer Hofmann,
University of Zurich, Switzerland
Angelina Sutin,
Florida State University, United States

****Correspondence:***
Hsueh-Chih Chen
chcjyh@gmail.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *17 January 2017*
Accepted: *29 January 2018*
Published: *20 February 2018*

Citation:
Tsai M-N, Wu C-L, Tseng L-P, An C-P and Chen H-C (2018) Extraversion Is a Mediator of Gelotophobia: A Study of Autism Spectrum Disorder and the Big Five. Front. Psychol. 9:150. doi: 10.3389/fpsyg.2018.00150

INTRODUCTION

When a person is teased or mocked by others, they usually experience negative feelings such as anger, sadness, shame, disgust, or fear (Platt, 2008; Platt and Ruch, 2009). Generally, however, most people can cope with such situations and modify their responses accordingly (Chen et al., 2011). Nevertheless, there remain some people who cannot tell the difference between playful teasing and malicious ridicule. Such individuals perceive all jokes to be hostile and cannot respond lightheartedly or cheerfully to jokes or laughter when interacting socially (Titze, 2009). People

who are paranoid about laughter have *gelotophobia*—a fear of being laughed at (Ruch and Proyer, 2008a). They are sensitive to laughter in all social situations, dread being laughed at, and regard others' smiles only as scornful (Hofmann et al., 2015). Their extreme fear results in maladjusted behaviors (Ruch and Proyer, 2009a), certain cases of which are often the subject of clinical research (Titze, 1996, 2009; Ruch et al., 2009). The term "gelotophobia" (from *gelos*, Greek for laughter) was proposed by Titze (1995, 1996, 1997). Based on his clinical observations, some people seem to be excessively concerned about being laughed at by others. They cannot distinguish the difference between playful teasing and ridicule, and perceive all types of laughter as hostile (Titze, 1995, 1996, 1997). Titze (2009) claimed that gelotophobes might be accustomed to seeing poker-faced or apathetic expressions in childhood, which leads them to respond fearfully to any jokes, even to good-humored, playful banter. In such circumstances, gelotophobes allay their state of uneasiness non-verbally; out of anxiety, their faces become stiff (Titze, 2009).

In a series of empirical studies, Ruch and his colleagues (Ruch and Proyer, 2008a,b, 2009a,b; Ruch et al., 2010, 2014) found that gelotophobia was not only seen among otherwise healthy people but also more generally in different cultures, and they developed two other concepts that were related to laughter, but were contrary to the fear of being laughed at Ruch and Proyer (2009a) described two emotional aspects of laughter (fear and joy) and their relationship to laughter's object (the self and others). Thus they found three related traits: gelotophobia, where the individual fears being laughed at; gelotophilia, where the individual enjoys being laughed at; and katagelasticism, where the individual enjoys laughing at others. In contrast to gelotophobes, "gelotophiles" feel positive and happy when they are being laughed at. They are not fearful or afraid of being ridiculed; in fact, they enjoy it when others laugh at them. Moreover, they actively seek situations in which others may laugh at them; for example, by sharing embarrassing things that have happened to them, or speaking openly about misfortunes and mishaps, thus provoking laughter in their audience. It is noticeable that gelotophobes and gelotophiles are not two extremes of the same type; rather, they are two distinct types. Their responses to being ridiculed are opposite. Gelotophiles are not only unafraid of being laughed at, but also gain pleasure from it. Gelotophobes and gelotophiles exhibit entirely different characteristics (Ruch and Proyer, 2009a). Besides self-directed laughter, Ruch and Proyer (2009a) found that some people seek situations in which they can laugh at others; such individuals are termed "katagelasticists." Katagelasticists initially look for the chance to mock others and then revel in seeing others fall victim to embarrassing or unfortunate events; furthermore, they continue to search for any opportunity to ridicule these same people, and make fun of them by insulting them directly or by using offensive words to describe them. In addition, katagelasticists never make fun of themselves to please others, and will defend themselves if others laugh at them.

As for the causes of gelotophobia, Ruch (2004, 2009) proposed the "model of the putative causes and consequences of gelotophobia," which was later revised (Titze, 2009; Ruch et al., 2014). He claimed that children who did not feel loved or appreciated within the parent–child relationship do not develop a sense of belonging, and then withdraw socially to avoid being ridiculed. The experience of being mocked is possibly the origin of their fear of being laughed at, and this fear extends into adulthood, as expressed by the styles identified in the PhoPhiKat questionnaire (Chen et al., 2011). According to the study conducted by Platt and Ruch (2009), a positive correlation exists between the experience of being bullied and gelotophobia. Moreover, a study by Samson et al. (2011) also confirmed an association between gelotophobia and a past experience of ridicule. All of these studies revealed that the fear of being laughed is more marked when the past experience of being derided or bullied was serious or frequent, and that it is hard to "shake off" such a fear once it has taken root (Liu et al., 2014). The revision of this model included the external conditions and internal factors about the fear of being laughed at, where external conditions referred to the peer group norm, societal structure, cultural factors and so forth, and internal factors referred to genetics, personality, emotional dispositions and so on (Ruch et al., 2014). The revision of the model indicates that not only early experience, but also personality, social skills, and external conditions, are all potential factors in the development of gelotophobia. The revised model is more effective at explaining the cause–effect relationship of gelotophobia and how the findings can be used to help individuals.

Research into the reasons for, and prevention of, gelotophobia indicates that individuals with gelotophobia usually have problems with emotional adjustment and social skills (Papousek et al., 2009, 2014). It is also seen among some patients with psychological defects (Weiss et al., 2012); in particular, some researchers are interested in the connection between autism spectrum disorder (ASD) and gelotophobia (Samson et al., 2011; Wu et al., 2015).

ASD is a neurodevelopmental disorder that is usually characterized by weak social and communication skills, along with the tendency to show fixed repeated behaviors and activities (APA, 2013). For those with autism, such weaknesses affect their social functioning and the ability to empathize, which makes it extremely difficult for them to recognize, or identify with, the mindset of other people, including their beliefs, thoughts, and emotions (Baron-Cohen et al., 1985; Baron-Cohen, 1989, 2001). In addition, such individuals also find it difficult to interpret the non-verbal cues of other people, such as body language (Asperger, 1944; Attwood, 2000). The lack of these interpersonal skills means that peers perceive those with autism as odd, unsociable, stubborn, and self-centered. Past research has claimed that when those with autism fail to grasp the latent agreement or implicit rules within group interactions, they become frustrated and experience increasing psychological pressure, which then results in emotional disturbance (Myles and Simpson, 2003; Attwood, 2007). Some such individuals even tend to use aggressive humor (Wu et al., 2014), which can lead to those with autism being singled out or teased.

Children with autism are frequently mocked or teased for their clumsy or odd behaviors (Carter, 2009). They do not know how to make friends and consequently become isolated from

their peers. Samson et al. (2011) indicated that people with autism are very fearful of being ridiculed owing to an early experience of being bullied and being laughed at. The study by Samson et al. (2011) showed that the proportion of people with gelotophobia was higher among the autism group (45%) than among the typically developed group (6%). The level of gelotophobia was positively correlated with the frequency of being ridiculed and the severity of the ridicule. In addition, people with autism did not like self-ridiculing (i.e., they had a low level of gelotophilia), but liked to laugh at others, as did those in the typically developed group. Similar results have been found in a study of Taiwanese people (Wu et al., 2015): those with autism exhibited a higher level of gelotophobia but a lower level of gelotophilia. Moreover, in terms of the level of katagelasticism, no significant difference was found between the Taiwanese individuals with autism and the typically developed group (Wu et al., 2015).

However, while those with autism have a higher level of gelotophobia than the level seen in a typically developed group, individual difference is significant: not all those with autism have gelotophobia (Wu et al., 2015). To understand why some people with ASD have gelotophobia while others with ASD do not, researchers have investigated the relationship between personality and gelotophobia.

Personality is stable and built into the early stages of life, and is usually considered a higher hierarchical trait (Furnham, 2008). Many researchers concur that personality could explain why only some children experience emotional and behavioral problems (Eaves et al., 1994; Wing, 1997; Hepburn, 2003; Leyfer et al., 2006). An individual usually shows a disposition in early childhood (Rothbart et al., 2006), and research has revealed that personality is generally relevant to maladapted behaviors, both in individuals with autism and in typically developed individuals (Mervielde et al., 2005, 2006). Previous research has revealed that personality and the fear of being laughed at are related traits, but discussions about the causes of gelotophobia in those with autism have almost exclusively focused on the influence of childhood experience; for example, early experiences of being mocked (Samson et al., 2011; Proyer and Neukom, 2013) and parental attachment (Wu et al., 2015).

However, the literature has also highlighted a strong association between gelotophobia and personality: since personality determines how individuals cope with social situations and subsequent events, and since gelotophobia is related to many sub-factors, such as fear and anxiety, personality is also a source of gelotophobia (Ruch and Proyer, 2009b). Ruch et al. (2013) found that gelotophobia could be predicted by the personal traits of the individual, and was also correlated to the "Big Five" personality traits. In particular, gelotophobia was positively correlated with neuroticism, but negatively correlated with extraversion and an openness to experience. Gelotophilia was found to be positively correlated with extraversion and openness, but negatively correlated with conscientiousness and neuroticism. Finally, katagelasticism was found to be negatively correlated with levels of agreeableness and conscientiousness. Other research has revealed similar findings (Chen et al., 2011; Proyer et al., 2012a,b,c). Studies that have investigated the personalities of those with ASD and those in a typically developed group also found that the level of neuroticism was higher in those with autism than in those in the typically developed group, while the levels of extraversion, agreeableness, openness, and conscientiousness were lower; these findings were consistent for both children and adults, male or female (Schriber et al., 2014). In their revised model of gelotophobia, Ruch et al. (2014) claim that personality is an antecedent risk factor for gelotophobia. We assume therefore that personality may be the key trait of gelotophobia in those with ASD.

To determine the potential causes of gelotophobia, the present study aims to understand the connection between the tendency toward gelotophobia in individuals with autism, and personality. We investigated the difference in the degree of gelotophobia between those with autism and a typically developed group, and how personality mediates the extent of gelotophobia. As teenagers with autism are more likely to be interacting with peers at school, and thus to encounter situations in which they are laughed at, or they laugh at others (Van Roekel et al., 2010), the present study focused on high school students.

We investigated the differences between the Big Five personality traits and gelotophobia among teenagers with autism and without autism, and then investigated the relationship between the mediating effects of the Big Five personality traits and gelotophobia in the autism group and the typically developed group. The research structure is shown in **Figure 1**. We duplicated past research by establishing two groups as independent variables (teenagers with autism vs. the typically developed teenagers as the control group) and traits for fear of being laughed at (gelotophobia, gelotophilia, and katagelasticism) as dependent variables, with the Big Five personality traits as mediators between the two sets of variables.

We hypothesized that teenagers with autism would have a greater tendency toward gelotophobia and would have a greater dislike of being laughed at than the members of the typically developed group. The study conducted by Ruch and Proyer (2009b) found that gelotophobia was positively related to extraversion but negatively related to neuroticism as defined in Eysenck's psychoticism, extraversion, and neuroticism (PEN) model of personality (Eysenck et al., 1985); therefore, we also hypothesized that extraversion and neuroticism were associated with a tendency toward gelotophobia for individuals with autism.

MATERIALS AND METHODS

Participants

The present study recruited students with autism from high schools in Taiwan. These students had been diagnosed by doctors or by the municipal special education identification and counseling committees composed of special education professionals, who confirmed the presence of autism or Asperger syndrome as defined by the following *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-V) criteria: (1) having notable impaired verbal and non-verbal communication; (2) having notable impaired social interaction; (3) having restricted and repetitive behavior; and (4) having intelligence

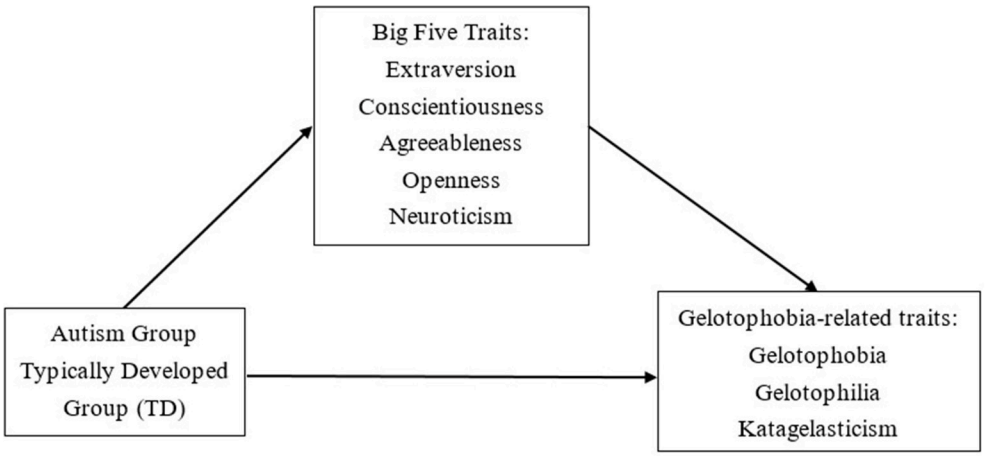


FIGURE 1 | Research structure.

quotients (IQs) of 70 or above according to the Wechsler Intelligence Scale for Children. For the typically developed group, we recruited high school students without autism from the same area as the high school students with autism. Excluding invalid questionnaires (i.e., those with a ratio > 1:8 of incomplete material or inconsistent answers), the sample group comprised 123 high school students with autism while the typically developed group comprised 156 high school students without autism; the total sample size was 279 participants. In the autism group, 87% of the participants were male, which matches the gender ratio (8:1 male to female) of individuals with autism in Taiwan aged 14–18 years ($M = 15.67$, $SD = 1.33$). In the typically developed group, 87% of the participants were again male, but aged 15–18 years ($M = 15.84$, $SD = 0.65$). There were no differences in gender or age between the two groups [$X^2 = 0.002$, $p = 0.963$; $t_{(277)} = 1.361$, $p = 0.175$]. The study was approved by the Institution Review Board of Taipei Medical University. All participants were informed of the study procedure and provided informed consent before commencement of the study.

Materials

The research tools were the "Big Five mini-markers" and the PhoPhiKat-TC scale. These are described below.

Big Five Mini-Markers

We used the Big Five mini-markers test employed in the study by Chen et al. (2011). This study was translated into Chinese by Saucier (1994). It includes five constructs: *extraversion, conscientiousness, agreeableness, openness*, and *emotional stability*. Saucier used *emotional stability* to replace *neuroticism* to ensure consistency between the five constructs. Every construct is composed of eight adjectives, totaling 40 items. The participants rate each item using a seven-point scale; a higher rating was recorded when the participant believed the item to be consistent with their own characteristics. The results of factor analysis by Saucier (1994) showed that the loading for all factors was >0.40, and the Cronbach's alpha for the five factors ranged from 0.76 to 0.85. Both reliability and validity were found to be satisfactory.

PhoPhiKat-TC Scale

We used the PhoPhiKat-TC scale described by Chen et al. (2011). This scale encompasses three concepts: gelotophobia, gelotophilia, and katagelasticism. Every concept is measured by 15 items, and the entire scale comprises 45 items. The participants rate each item on a four-point scale; the more the participants agree with the item, the higher they rate it. With regard to the construct validity, the indices of model fit were found to be higher than 0.90 according to the results of confirmatory factor analysis. The construct validities of the three factors were found to be satisfactory. The Cronbach's alpha of each sub-scale was 0.85, and the test-retest reliability ranged from 0.87 to 0.92. The reliability of the scale was also found to be satisfactory.

Procedure

At the initial stage of the research, we investigated the distribution of Taiwanese high school students with autism. We invited the relevant high schools to survey their students regarding their willingness to participate in the research. We then recruited the participants and carried out the research in the high schools where the students with autism were studying; all participants received detailed information about the content of the research and provided their written consent prior to participation. Research was implemented by means of the test and the questionnaire. Researchers first introduced the purpose of the research and offered guidance on how to fill in the questionnaire. When all the participants fully understood the procedure, they began to complete the Big Five mini-markers test and the PhoPhiKat-TC questionnaire in sequential order. The total time for completing the questionnaire was 15 min. All the participants received a set of stationery items upon completion.

RESULTS

Comparison of Gelotophobia in Students with ASD and Those in the Typically Developed Group

To classify the level of gelotophobia, Ruch and Proyer (2008b) designated the cut-off points of the PhoPhiKat as follows: no gelotophobia 1.0–2.5; slight gelotophobia 2.5–3.0; marked gelotophobia 3.0–3.5; and extreme gelotophobia 3.5–4.0. The result of a Chi-square test revealed a significant difference in gelotophobia between the two groups ($X^2 = 8.597$, $p = 0.035$): 73.7% of the typically developed group had no gelotophobia, compared with 57.7% of the ASD group; and 35.8% of the ASD group had slight gelotophobia, compared with 20.5% in the typically developed group. No difference was found between the ASD group and the typically developed group with regard to extreme gelotophobia (**Table 1**).

TABLE 1 | Comparison of gelotophobia in ASD group and typically developed group.

			Group		Total
			TD	ASD	
Level of gelotophobia	No	Count	115[a]	71[b]	186
		% within group	73.7	57.7	66.7
	Slight	Count	32[a]	44[b]	76
		% within group	20.5	35.8	27.2
	Marked	Count	7[a]	6[a]	13
		% within group	4.5	4.9	4.7
	Extreme	Count	2[a]	2[a]	4
		% within group	1.3	1.6	1.4
Sub-total		Count	156	123	279
		% within group	100.0	100.0	100.0

[a,b]Significant difference between ASD group and typically developed group, $p < 0.05$.

Differences in Personality Traits and Gelotophobia-Related Traits

The multivariate analysis of variance (MANOVA) results are reported in **Table 2**. In terms of the traits related to gelotophobia, the ASD group had a higher level of gelotophobia than the typically developed group [$F_{(1, 277)} = 3.938$, $p = 0.048$, $\eta^2 = 0.14$], but lower levels of gelotophilia [$F_{(1, 277)} = 47.752$, $p < 0.001$, $\eta^2 = 0.147$] and katagelasticism [$F_{(1, 277)} = 5.208$, $p = 0.023$, $\eta^2 = 0.18$] than the typically developed group.

Regarding personality, the ASD group was found to have a lower level of extraversion [$F_{(1, 277)} = 14.604$, $p < 0.001$, $\eta^2 = 0.050$] and agreeableness [$F_{(1, 277)} = 14.344$, $p < 0.001$, $\eta^2 = 0.049$] than the typically developed group; but no significant difference was found between the two groups in terms of conscientiousness, openness, and emotional stability.

TABLE 2 | Mean and SD of Big Five and gelotophobia-related traits between groups.

	ASD group *M(SD)* *N* = 123	TD group *M(SD)* *N* = 156	*F*	*P*	η^2
PHOPHIKAT-TC SCALE					
Gelotophobia	2.42 (0.48)	2.30 (0.47)	3.938*	0.048	0.014
Gelotophilia	1.98 (0.47)	2.38 (0.50)	47.752**	<0.001	0.147
Katagelasticism	2.00 (0.50)	2.13 (0.47)	5.208*	0.023	0.018
BIG FIVE PERSONALITY					
Extraversion	3.85 (0.99)	4.34 (1.11)	14.604**	<0.001	0.050
Conscientiousness	3.76 (0.83)	3.94 (0.83)	3.578	0.060	0.013
Agreeableness	4.56 (0.91)	4.97 (0.89)	14.344**	<0.001	0.049
Openness	4.36 (0.94)	4.55 (0.95)	2.788	0.096	0.010
Emotional stability	3.91 (0.97)	4.10 (0.95)	2.764	0.098	0.010

*$p < 0.05$, **$p < 0.01$.*

Canonical Correlation of Gelotophobia-Related Traits and Personality

To investigate the correlation between gelotophobia-related traits and the Big Five, as well as to diminish the interference between traits, we used a canonical correlation to determine the difference between the Big Five and gelotophobia-related traits both within groups and between groups. According to our hypotheses, personality is the antecedent variable of gelotophobia; therefore, Big Five personality traits are predictive variables, and gelotophobia-related traits are criterion variables.

The results are shown in **Table 3**. Within the typically developed group, for Big Five and gelotophobia-related traits, three sets of generalized *F* coefficients of roots were found to be significant. The coefficients of canonical correlation for the three sets were 0.666 ($p < 0.001$), 0.543 ($p < 0.001$), and 0.380 ($p = 0.003$). In order, the first coefficient indicates that extraversion (−0.863) had an influence on the level of gelotophobia (0.971) (the lower the level of extraversion, the higher the gelotophobia); the second coefficient indicates that agreeableness (0.789) had an influence on the level of katagelasticism (−0.983) (the higher the level of agreeableness, the lower the katagelasticism); and the third coefficient indicates that emotional stability (0.563) had an influence on the level of gelotophilia (−0.889) (the higher the level of emotional stability, the lower the gelotophilia).

Within the ASD group, for Big Five and gelotophobia-related traits, two sets of generalized F coefficients of roots were found to be significant. The coefficients of canonical correlation for the two sets were 0.626 ($p < 0.001$) and 0.488 ($p < 0.001$). In order, the first coefficient indicates that agreeableness (−0.966) had an influence on the level of katagelasticism (0.891) (the lower the level of agreeableness, the higher the katagelasticism); and the second coefficient indicates that extraversion (−0.901) had an influence on the level of gelotophobia (0.775) (the lower the extraversion, the higher the gelotophobia).

In terms of the results of the canonical correlation for Big Five and gelotophobia-related traits, similar findings were recorded for both groups; i.e., that extraversion is the best predictor of gelotophobia, and agreeableness is the best predictor of katagelasticism. The main difference between the groups was

TABLE 3 | Correlation of Big Five and gelotophobia-related traits.

Group	TD			ASD		
Root	**1**	**2**	**3**	**1**	**2**	**3**
Eigenvalue	0.796	0.419	0.098	0.645	0.312	0.051
Variance	60.593	31.891	7.517	63.954	30.941	5.104
Canonical	0.666	0.543	0.380	0.626	0.488	0.221
Wilks	0.357	0.642	0.910	0.441	0.725	0.951
F	12.320	9.258	4.936	7.326	5.063	2.008
p	<0.001	<0.001	0.003	<0.001	<0.001	0.117
PREDICTIVE VARIABLES						
Extraversion	**−0.863**	−0.206	−0.208	0.103	**−0.901**	
Conscientiousness	−0.206	0.633	0.148	−0.475	−0.282	
Agreeableness	−0.492	**0.789**	−0.291	**−0.966**	−0.062	
Openness	−0.530	−0.006	−0.487	−0.051	−0.456	
Emotional stability	−0.612	0.480	**0.563**	−0.695	−0.277	
CRITERION VARIABLES						
Gelotophobia	**0.971**	−0.018	−0.240	0.176	**0.775**	
Gelotophilia	−0.320	−0.327	**−0.889**	0.042	−0.575	
Katagelasticism	0.142	**−0.983**	−0.117	**0.891**	−0.028	

*$p < 0.05$, **$p < 0.01$. The bold values are highest correlation in the root.*

that emotional stability was proven to be the best predictor of gelotophobia for the typically developed group, but not for the ASD group.

Mediation Analysis of the Big Five Personality Markers

To investigate the mediation effect of the Big Five personality markers on gelotophobia for people with autism and the typically developed group, we used gender and age as the control variables and the group (the autism group vs. the typically developed group) as the independent variable. For the dummy coding of the group, those with autism were assigned a value of 1 and those in the typically developed group were assigned a value of 0. We then used the Big Five personality markers as mediators and performed a mediation analysis using gelotophobia, gelotophilia, and katagelasticism as dependent variables. For the mediation analysis, we used a bootstrapping approach (Preacher and Hayes, 2008), which simulates the data of a large sample by re-sampling from current data. Analyzing a large sample of data means that we can obtain a more precise prediction. In the present study, our simulation sample comprised 5,000 people. First, we made a prediction for the groups in terms of personality, and then we investigated the predictive power of personality by using gelotophobia, gelotophilia, and katagelasticism as dependent variables. Finally, we analyzed the predictions for the groups in terms of gelotophobia, gelotophilia, and katagelasticism, and the change in the predictions after using personality as a mediator.

Predictions for Groups in Terms of Personality

Because the predictive power of groups in terms of personality is not affected by using gelotophobia, gelotophilia, and katagelasticism as dependent variables, we began the analysis with a path analysis. The findings revealed that groups are powerful in terms of predicting extraversion ($\beta = -0.499$, $p < 0.01$) and agreeableness ($\beta = -0.394$, $p < 0.01$). The β-values were negative for both groups and indicated that individuals with autism have lower levels of extraversion and agreeableness.

Using Gelotophobia as a Dependent Variable

Only extraversion ($\beta = -0.171$, $p < 0.001$) and emotional stability ($\beta = -0.175$, $p < 0.001$) were found to be statistically significant with regard to predicting gelotophobia by personality. In other words, those with lower levels of extraversion and emotional stability had higher levels of gelotophobia.

Concerning the predictions for groups in terms of gelotophobia and the change in the prediction after using personality as a mediator, the findings revealed that groups significantly predict the level of gelotophobia ($\beta = 0.114$, $p = 0.047$). However, after using personality as a mediator, the power of the groups disappears ($\beta = 0.006$, $p = 0.910$), and the power of prediction is completely overtaken by personality. Upon further examination, the mediation effect of personality was found to derive mainly from extraversion (95% CI [0.041, 0.145]) and shows that gelotophobia is significantly influenced by extraversion; i.e., individuals with lower levels of extraversion are more fearful of being laughed at (**Figure 2**). To conclude, extraversion, rather than the groups (with or without autism), influences the level of gelotophobia.

Using Gelotophilia as a Dependent Variable

Our findings showed that extraversion ($\beta = 0.105$, $p < 0.001$) and openness ($\beta = 0.081$, $p = 0.016$) were found to be significant in predicting gelotophilia. Individuals with higher levels of extraversion or openness had higher levels of gelotophilia.

We then investigated the predictions for groups and the change after using personality as a mediator. Our results showed that the group significantly predicted the level of gelotophilia ($\beta = -0.411$, $p < 0.001$). The group effect remains but decreases after using personality as a mediator ($\beta = -0.351$, $p < 0.01$). The Sobel test clearly showed statistical significance ($p < 0.05$), and the mediation effect of personality was only partial. The mediation effect of personality was found to derive from extraversion (95% CI [−0.106, −0.015]). In summary, those with autism had lower levels of gelotophilia, as did those with lower levels of extraversion (as shown in **Figure 3**).

Using Katagelasticism as a Dependent Variable

Extraversion ($\beta = 0.063$, $p = 0.015$), conscientiousness ($\beta = -0.085$, $p = 0.012$), agreeableness ($\beta = -0.219$, $p < 0.001$), and emotional stability ($\beta = -0.065$, $p = 0.044$) were found to be statistically significant in predicting the level of katagelasticism. Individuals with higher levels of extraversion had higher levels of katagelasticism. In contrast, individuals with higher levels of conscientiousness, agreeableness, and emotional stability had lower levels of katagelasticism.

Concerning the predictions for groups in terms of katagelasticism and the change after using personality as a mediator, our findings show that the group is powerful in

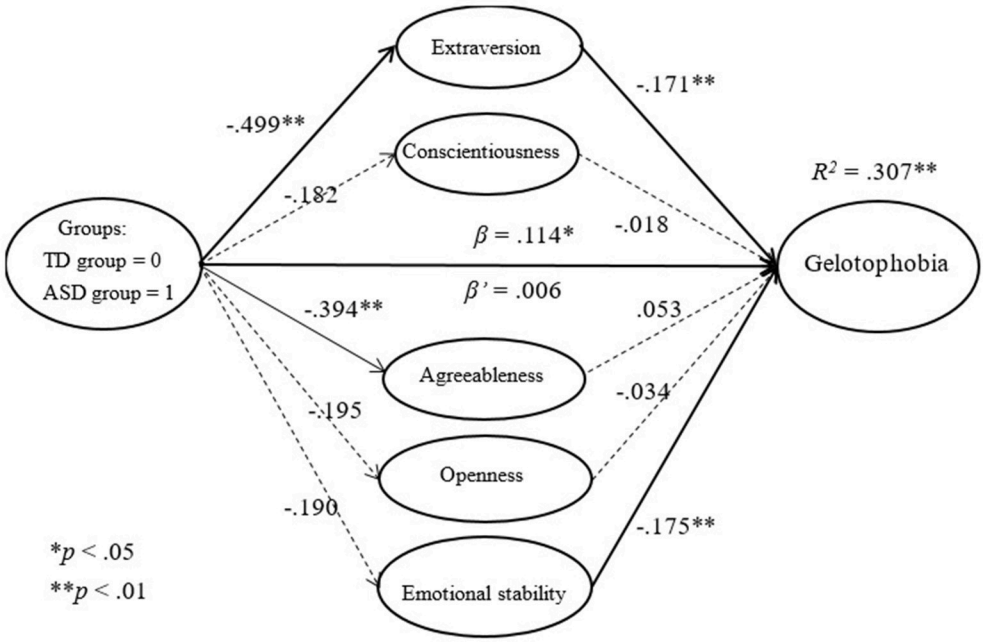


FIGURE 2 | The mediation effect of personality on gelotophobia.

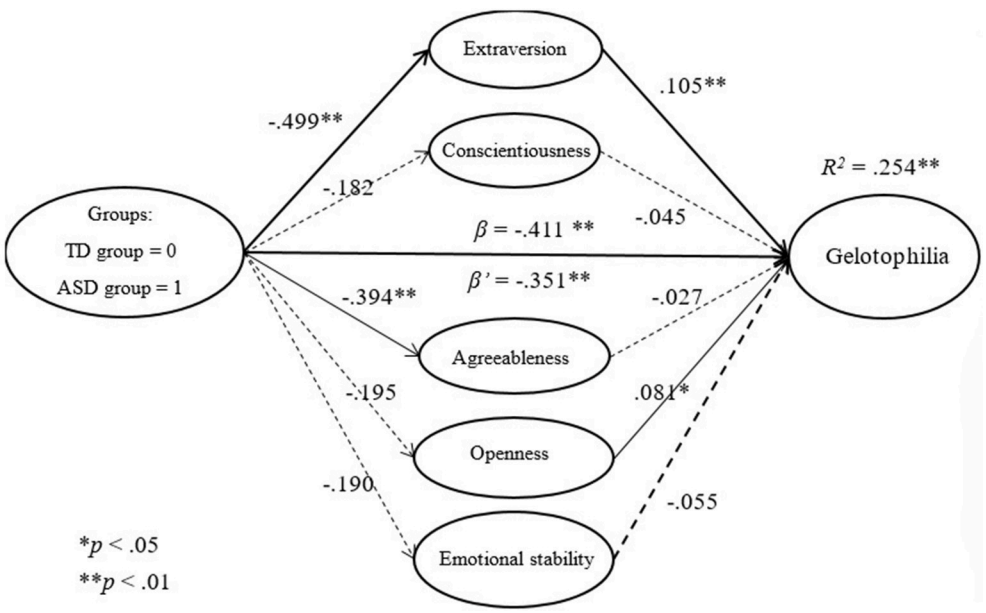


FIGURE 3 | The mediation effect of personality on gelotophilia.

predicting the level of katagelasticism ($\beta = -0.136$, $p = 0.021$); moreover, its power slightly increases after using personality as a mediator ($\beta = -0.211$, $p < 0.001$). Personality is not a mediator, but rather a suppressor. Upon further analysis, the suppression effect of personality was found to derive from agreeableness (95% CI [0.041, 0.151]). Individuals with lower levels of agreeableness had higher levels of katagelasticism. Those with autism were found to have lower levels of agreeableness, but also lower levels of katagelasticism, than those in the typically developed group. Our results indicate that autism and the level of agreeableness are in conflict when predicting the level of katagelasticism. After using personality as a mediator, the effect of the group is more powerful in predicting the level of katagelasticism (as shown in **Figure 4**).

DISCUSSION

Conclusion and Implications

Past research has indicated that early experience affects whether gelotophobia develops in those with autism. Such individuals were thought to be lacking in communication and empathy skills and were often ridiculed by peers, which ultimately resulted in a fear of being laughed at. Personality was seldom considered as a contributory factor in terms of the fear of being laughed at in those with autism. Moreover, in the earlier studies of the personalities of those with autism, and the associations of personality with gelotophobia, the sample sizes were usually fewer than 50 participants, such as the research undertaken by Samson et al. (2011) and Schriber et al. (2014).

The present study recruited participants from high schools in Taiwan, selecting a typically developed group of teenagers and a group of teenagers with autism. We used a self-reporting questionnaire as a research tool, and investigated the difference in the level of gelotophobia between those teenagers with autism and those without. We then considered personality as a mediator to verify the relationship of autism to gelotophobia.

Our findings show that those individuals with ASD had higher levels of gelotophobia than those in the typically developed group. Further, we found that the most significant difference between the two groups was that the ASD group had a higher percentage of slight gelotophobia, and a lower percentage of no gelotophobia, than those in the typically developed group. In general, those with ASD had a higher level of gelotophobia than those in the typically developed group, particularly at the "slight" level, though no difference was found between the "marked" level and the "extreme" level between the two groups. The ASD group was found to have lower levels of katagelasticism and gelotophilia; i.e., those with ASD dislike being laughed at, but also are not interested in laughing at others. This finding is consistent with the findings of previous research (Wu et al., 2015).

Regarding personality, our results indicate that those with ASD had lower levels of extraversion and agreeableness than those in the typically developed group, possibly due to the fact that Chinese culture strongly emphasizes interpersonal harmony (Markus and Kitayama, 2010). Individuals with ASD tend to have poor social skills; consequently, those in the ASD group lagged behind those in the typically developed group in certain situations. Thus the difference between the ASD group and the typically developed group can perhaps be explained by cultural factors, and suggests that cultural factors should be taken into account in future studies. The other difference between the ASD group and the typically developed group was the canonical correlation results. We found that, for both groups, extraversion can predict the level of gelotophobia while agreeableness can predict the level of katagelasticism, though emotional stability can predict the level of gelotophilia only for the typically developed group. This suggests that personality is not the reason for the level of gelotophilia in those with ASD.

Lastly, the mediation analysis revealed that the level of gelotophobia was completely mediated by extraversion, indicating that those in the ASD group had a higher level of gelotophobia than those in the typically developed group, which was mainly driven by extraversion. This finding supports the claim by Ruch and Proyer (2009b) that an individual with a lower level of extraversion is less adept at social interaction and is less able to engage in humor with others. This, in turn, makes it less likely for such individuals to experience the positive influence of laughter and humor. Moreover, they are more ill at ease in teasing situations and more afraid of being laughed

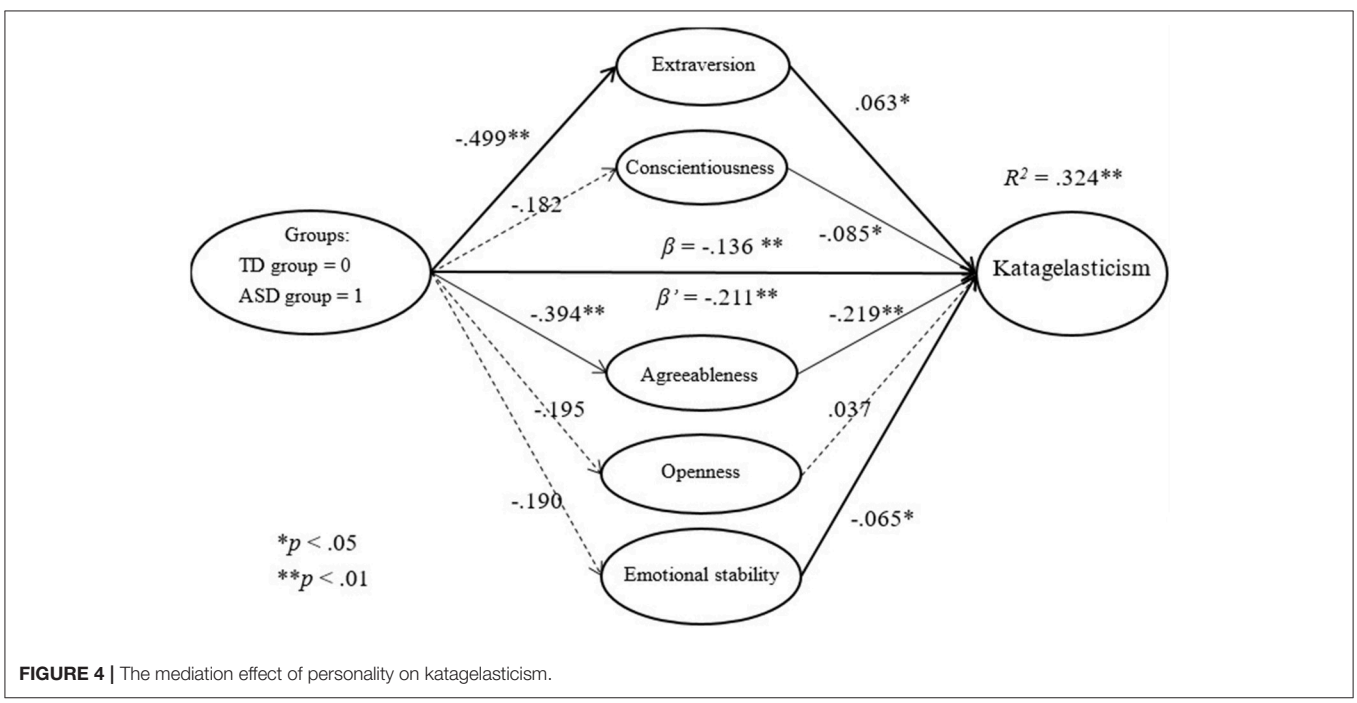


FIGURE 4 | The mediation effect of personality on katagelasticism.

at. In addition, extraversion was found to have an influence on the level of gelotophilia, although the mediated effect was partial. Those with higher levels of extraversion are keen to interact with others to create interesting situations, and even to engage in self-mockery. They do not feel embarrassed or uncomfortable at being laughed at; rather, they enjoy it (Ruch and Proyer, 2009a). Hence individuals with higher levels of extraversion usually have higher levels of gelotophilia. Further, the Big Five personality traits were found to confer no mediated effect on katagelasticism. We found that agreeableness inhibits the relationship between autism and katagelasticism, so that individuals with lower levels of agreeableness also have higher levels of katagelasticism (Ruch et al., 2013). As those with autism were found to have lower levels of agreeableness and also lower levels of katagelasticism than those in the typically developed group, this indicates that katagelasticism is not driven by personality but by other, unknown, factors.

Our results suggest that gelotophobia develops not from autism, but from a lower level of extraversion. For individuals with autism, a group therapy dealing with social skills (Glinski and Page, 2010) or a training program in emotional competencies will help (Nelis et al., 2011). Those with autism can learn interpersonal skills and thus learn how to engage in humor and how to cope with teasing situations. This, in turn, may alter their negative perception of interpersonal interaction, and thereby improve their fear of being laughed at.

Our findings suggest that future studies should investigate the reasons for katagelasticism and gelotophilia in order to understand what influences levels of gelotophilia and katagelasticism for those with ASD, and why those with ASD have both lower levels of agreeableness and lower levels of katagelasticism. They also suggest that it would be beneficial to create a comprehensive model of mockery styles based on the model of the putative causes and consequences of gelotophobia.

Limitations

The present study has several limitations. First, we did not manipulate variables in the current study, so we cannot draw any inferences about the causal directions of the relationships. Titze (1995, 1996, 1997) considers that past experience of being laughed at during childhood influences personality development. More empirical studies of the relationships between gelotophobia and personality are needed, particularly longitudinal studies. Second, the present study used a test and a self-reporting questionnaire as research tools but lacked any other source of information; for example, observations by parents, teachers, or peers. Social desirability may have been a source of bias, indicating that more objective data should be used in future research. Third, the results of the current study were possibly restricted to males, because purposive sampling was used in this study. The gender ratio of individuals with autism in Taiwan is 8:1 (males to females), so the proportion of female participants in our study was only 13%. Therefore, if females and males had been included equally in the sample, it would have been necessary to examine gender as a factor. Fourth, the present study did not collect demographics of participants, e.g., parental socioeconomic status. Past research has indicated that a positive relationship has been observed between socioeconomic status and ASD prevalence (Durkin et al., 2010). Socioeconomic status might be a confounder of any associations between ASD status, personality traits, and gelotophobia-related traits. To investigate the index of socioeconomic status in future research, it would be more helpful to understand the causal mechanisms or confounding factors associated with ASD. Finally, because the participants of this study were Taiwanese students, the results may only be applicable to participants from an Asian cultural background. Thus, we recommend that future studies include participants from various cultures to examine whether the findings of this study differ culturally.

AUTHOR CONTRIBUTIONS

L-PT and C-PA designed and conducted the research. M-NT and C-LW analyzed the data and drafted the manuscript under the supervision of H-CC. All authors approved the final version of the manuscript for submission.

FUNDING

This research is sponsored by the "Building of the Chinese Sentiment WordNet Dictionary and Development of the Mental Assessment Index via Chinese Written Material-A study based on linguistic big data" of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 104-2511-S-003-019-MY3.

REFERENCES

American Psychiatric Association (APA) (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edn.* Washington, DC: American Psychiatric Association.

Asperger, H. (1944). Die "Autistischen Psychopathen" im Kindesalter. *Eur. Arch. Psychiatry Clin. Neurosci.* 117, 76–136. doi: 10.1007/BF01837709

Attwood, T. (2000). Strategies for improving the social integration of children with Asperger syndrome. *Autism* 4, 85–100. doi: 10.1177/1362361300004001006

Attwood, T. (2007). Guida alla *Sindrome* di Asperger. *Diagnosi e Caratteristiche Evolutive*. Austin, TX: Edizioni Erickson.

Baron-Cohen, S. (1989). The autistic child's theory of mind: a case of specific developmental delay. *J. Child Psychol. Psychiatry* 30, 285–297. doi: 10.1111/j.1469-7610.1989.tb00241.x

Baron-Cohen, S. (2001). "Theory of mind and autism: a review," in *Special Issue of the International Review of Mental Retardation,* Vol. 23, ed L. M. Glidden (San Diego, CA: Academic Press), 169–184.

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Carter, S. (2009). Bullying of students with Asperger syndrome. *Iss. Comprehen. Pediat. Nurs.* 32, 145–154. doi: 10.1080/01460860903062782

Chen, H. C., Chan, Y., Ruch, W., and Proyer, R. T. (2011). Evaluating the reliability and validity of a traditional Chinese version of the PhoPhiKat-45. *Psychol.*

Test. Taiwan 58, 119–145. Available online at: http://www.epc.ntnu.edu.tw/files/writing/2775_7a4ea1ed.pdf

Durkin, M. S., Maenner, M. J., Meaney, F. J., Levy, S. E., DiGuiseppi, C., Nicholas, J. S., et al. (2010). Socioeconomic inequality in the prevalence of autism spectrum disorder: evidence from a US cross-sectional study. *PLoS ONE* 5:e11551 doi: 10.1371/journal.pone.0011551

Eaves, L. C., Ho, H. H., and Eaves, D. M. (1994). Subtypes of autism by cluster analysis. *J. Autism Develop. Disord.* 24, 3–22. doi: 10.1007/BF02172209

Eysenck, S. B. G., Eysenck, H. J., and Barrett, P. (1985). A revised version of the psychoticism scale. *Person. Indiv. Differ.* 6, 21–29. doi: 10.1016/0191-8869(85)90026-1

Furnham, A. (2008). The Eysenck personality measures: fifty years of scale development. *Sage Handb. Person. Theory Assess.* 2, 199–218. doi: 10.4135/9781849200479.n10

Glinski, K., and Page, A. C. (2010). Modifiability of neuroticism, extraversion, and agreeableness by group cognitive behaviour therapy for social anxiety disorder. *Behav. Change* 27, 42–52. doi: 10.1375/bech.27.1.42

Hepburn, S. L. (2003). Clinical implications of temperamental characteristics in young children with developmental disabilities. *Infants Young Child.* 16, 59–76. doi: 10.1097/00001163-200301000-00007

Hofmann, J., Platt, T., Ruch, W., and Proyer, R. R. (2015). Individual differences in gelotophobia predict responses to joy and contempt. *Sage Open* 5, 1–12 doi: 10.1177/2158244015581191

Leyfer, O. T., Folstein, S. E., Bacalman, S., Davis, N. O., Dinh, E., Morgan, J., et al. (2006). Comorbid psychiatric disorders in children with autism: interview development and rates of disorders. *J. Aut. Develop. Disord.* 36, 849–861. doi: 10.1007/s10803-006-0123-0

Liu, C. H., Chiu, F. C., Chen, H. C., and Lin, C. Y. (2014). Helpful but insufficient: incremental theory on challenge-confronting tendencies for students who fear being laughed at. *Motiv. Emot.* 38, 367–377. doi: 10.1007/s11031-013-9386-x

Markus, H. R., and Kitayama, S. (2010). Cultures and selves: a cycle of mutual constitution. *Perspect. Psychol. Sci.* 5, 420–430. doi: 10.1177/1745691610375557

Mervielde, I., De Clercq, B., De Fruyt, F., and Van Leeuwen, K. (2005). Temperament, personality, and developmental psychopathology as childhood antecedents of personality disorders. *J. Person. Disord.* 19, 171–201. doi: 10.1521/pedi.19.2.171.62627

Mervielde, I., De Clercq, B., De Fruyt, F., and Van Leeuwen, K. (2006). "Temperament and personality as broad-spectrum antecedents of psychopathology in childhood and adolescence," in *Dimensional Models of Personality Disorders. Refining the Research Agenda for DSM-V*, eds T. A. Widiger, E. Simonsen, P. J. Sirovatka, and D. A. Regier (Washington, DC: American Psychiatric Association), 85–109.

Myles, B. S., and Simpson, R. L. (2003). *Asperger Syndrome: A Guide for Educators and 94 Parents*. Austin, TX: PRO-ED Inc.

Nelis, D., Kotsou, I., Quoidbach, J., Hansenne, M., Weytens, F., Dupuis, P., et al. (2011). Increasing emotional competence improves psychological and physical well-being, social relationships, and employability. *Emotion* 11:354. doi: 10.1037/a0021554

Papousek, I., Aydin, N., Lackner, H. K., Weiss, E. M., Bühner, M., Schulter, G., et al. (2014). Laughter as a social rejection cue: gelotophobia and transient cardiac responses to other persons' laughter and insult. *Psychophysiology* 51, 1112–1121. doi: 10.1111/psyp.12259

Papousek, I., Ruch, W., Freudenthaler, H. H., Kogler, E., Lang, B., and Schulter, G. (2009). Gelotophobia, emotion-related skills and responses to the affective states of others. *Person. Indiv. Differ.* 47, 58–63. doi: 10.1016/j.paid.2009.01.047

Platt, T. (2008). Emotional responses to ridicule and teasing: should gelotophobes react differently? *Humor Int. J. Humor Res.* 21, 105–128. doi: 10.1515/HUMOR.2008.005

Platt, T., and Ruch, W. (2009). The emotions of gelotophobes: shameful, fearful, and joyless? *Humor Int. J. Humor Res.* 22, 91–110. doi: 10.1515/HUMR.2009.005

Preacher, K. J., and Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* 40, 879–891. doi: 10.3758/BRM.40.3.879

Proyer, R. T., Estoppey, S., and Ruch, W. (2012a). An initial study on how families deal with ridicule and being laughed at: parenting styles and parent–child relations with respect to gelotophobia, gelotophilia, and katagelasticism. *J. Adult Develop.* 19, 228–237. doi: 10.1007/s10804-012-9150-6

Proyer, R. T., Flisch, R., Tschupp, S., Platt, T., and Ruch, W. (2012b). How does psychopathy relate to humor and laughter? Dispositions toward ridicule and being laughed at, the sense of humor, and psychopathic personality traits. *Int. J. Law Psychiatry* 35, 263–268. doi: 10.1016/j.ijlp.2012.04.007

Proyer, R. T., and Neukom, M. (2013). Ridicule and being laughed at in the family: gelotophobia, gelotophilia, and katagelasticism in young children and their parents. *Int. J. Psychol.* 48, 1191–1195. doi: 10.1080/00207594.2013.775448

Proyer, R. T., Neukom, M., Platt, T., and Ruch, W. (2012c). Assessing gelotophobia, gelotophilia, and katagelasticism in children: an initial study on how six to nine-year-olds deal with laughter and ridicule and how this relates to bullying and victimization. *Child Indic. Res.* 5, 297–316. doi: 10.1007/s12187-011-9127-1

Rothbart, M. K., Bates, J. E., and Eisenberg, N. (2006). *Handbook of Child Psychology,* Vol. 3. *Social, Emotional, and Personality Development, 6th Edn.* Hoboken, NJ: John Wiley & Sons Inc, 99–166.

Ruch, W. (2004). *Gelotophobia: A Useful New Concept? The meeting of IPSR Spring 2004 Colloquium Series.* Berkeley, CA: University of California at Berkeley.

Ruch, W. (2009). Fearing humor? Gelotophobia: the fear of being laughed at. Introduction and overview. *Humor Int. J. Humor Res.* 22, 1–25. doi: 10.1515/HUMR.2009.001

Ruch, W., Altfreder, O., and Proyer, R. T. (2009). How do gelotophobes interpret laughter in ambiguous situations? An experimental validation of the concept. *Humor Int. J. Humor Res.* 22, 63–89. doi: 10.1515/HUMR.2009.004

Ruch, W., Harzer, C., and Proyer, R. T. (2013). Three by five: Three dispositions towards ridicule and being laughed at meet the Big Five. *Israeli J. Humor Res. Int. J.* 2, 2–20.

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. (2014). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor Int. J. Humor Res.* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., and Proyer, R. T. (2008a). The fear of being laughed at: individual and group differences in gelotophobia. *Humor Int. J. Humor Res.* 21:47. doi: 10.1515/HUMOR.2008.002

Ruch, W., and Proyer, R. T. (2008b). Who is gelotophobic? Assessment criteria for the fear of being laughed at. *Swiss J. Psychol.* 67, 19–27. doi: 10.1024/1421-0185.67.1.19

Ruch, W., and Proyer, R. T. (2009a). Extending the study of gelotophobia: on gelotophiles and katagelasticists. *Humor Int. J. Humor Res.* 22, 183–212. doi: 10.1515/HUMR.2009.009

Ruch, W., and Proyer, R. T. (2009b). Who fears being laughed at? The location of gelotophobia in the Eysenckian PEN-model of personality. *Person. Indiv. Differ.* 46, 627–630. doi: 10.1016/j.paid.2009.01.004

Ruch, W., Proyer, R. T., and Ventis, L. (2010). The relationship of teasing in childhood to the expression of gelotophobia in adults. *Psychol. Sci. Quart.* 52, 77–93. Available online at: https://pdfs.semanticscholar.org/97d7/85a22b29bfaa689de581b5be5c064fc826a8.pdf

Samson, A. C., Huber, O., and Ruch, W. (2011). Teasing, ridiculing and the relation to the fear of being laughed at in individuals with Asperger's syndrome. *J. Aut. Develop. Disord.* 41, 475–483. doi: 10.1007/s10803-010-1071-2

Saucier, G. (1994). Mini-markers: a brief version of Goldberg's unipolar Big-Five markers. *J. Person. Assess.* 63, 506–516. doi: 10.1207/s15327752jpa6303_8

Schriber, R. A., Robins, R. W., and Solomon, M. (2014). Personality and self-insight in individuals with autism spectrum disorder. *J. Person. Soc. Psychol.* 106:112. doi: 10.1037/a0034950

Titze, M. (1995). *Die heilende Kraft des Lachens [The Healing Power of Laughter]*. Munich: Kosel.

Titze, M. (1996). The Pinocchio Complex: overcoming the fear of laughter. *Humor Health J.* 5, 1–11.

Titze, M. (1997). "Das Komische als schamauslo" sende Bedingung [The Funny as an Elicitor of Shame]," in *Scham - ein Menschliches Gefühl [Shame - A Human Emotion]*, eds R. Kühn, M. Raub, and M. Titze (Opladen: Westdeutscher Verlag), 169–178.

Titze, M. (2009). Gelotophobia: the fear of being laughed at. *Humor Int. J. Humor Res* 22, 27–48. doi: 10.1515/HUMR.2009.002

Van Roekel, E., Scholte, R. H., and Didden, R. (2010). Bullying among adolescents with autism spectrum disorders: prevalence and perception. *J. Autism Develop. Disord.* 40, 63–73. doi: 10.1007/s10803-009-0832-2

Weiss, E. M., Schulter, G., Freudenthaler, H. H., Hofer, E., Pichler, N., and Papousek, I. (2012). Potential markers of aggressive behavior: the fear of other persons' laughter and its overlaps with mental disorders. *PLoS ONE* 7:e38088. doi: 10.1371/journal.pone.0038088

Wing, L. (1997). The autistic spectrum. *Lancet* 350, 1761–1766. doi: 10.1016/S0140-6736(97)09218-0

Wu, C. L., An, C. P., Tseng, L. P., Chen, H. C., Chan, Y. C., Cho, S. L., et al. (2015). Fear of being laughed at with relation to parent attachment in individuals with autism. *Res. Autism Spectr. Disord.* 10, 116–123. doi: 10.1016/j.rasd.2014.11.004

Wu, C. L., Tseng, L. P., An, C. P., Chen, H. C., Chan, Y. C., Shih, C. I., et al. (2014). Do individuals with autism lack a sense of humor? A study of humor comprehension, appreciation, and styles among high school students with autism. *Res. Autism Spectr. Disord.* 8, 1986–1393. doi: 10.1016/j.rasd.2014.07.006

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Tsai, Wu, Tseng, An and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 November 2017
doi: 10.3389/fpsyg.2017.01954

Eye Contact and Fear of Being Laughed at in a Gaze Discrimination Task

Jorge Torres-Marín[1], Hugo Carretero-Dios[2], Alberto Acosta[1] and Juan Lupiáñez[1]*

[1] Mind, Brain and Behavior Research Center, Department of Experimental Psychology, University of Granada, Granada, Spain, [2] Mind, Brain and Behavior Research Center, Department of Methodology of Behavioral Sciences, University of Granada, Granada, Spain

Current approaches conceptualize gelotophobia as a personality trait characterized by a disproportionate fear of being laughed at by others. Consistently with this perspective, gelotophobes are also described as neurotic and introverted and as having a paranoid tendency to anticipate derision and mockery situations. Although research on gelotophobia has significantly progressed over the past two decades, no evidence exists concerning the potential effects of gelotophobia in reaction to eye contact. Previous research has pointed to difficulties in discriminating gaze direction as the basis of possible misinterpretations of others' intentions or mental states. The aim of the present research was to examine whether gelotophobia predisposition modulates the effects of eye contact (i.e., gaze discrimination) when processing faces portraying several emotional expressions. In two different experiments, participants performed an experimental gaze discrimination task in which they responded, as quickly and accurately as possible, to the eyes' directions on faces displaying either a happy, angry, fear, neutral, or sad emotional expression. In particular, we expected trait-gelotophobia to modulate the eye contact effect, showing specific group differences in the happiness condition. The results of Study 1 ($N = 40$) indicated that gelotophobes made more errors than non-gelotophobes did in the gaze discrimination task. In contrast to our initial hypothesis, the happiness expression did not have any special role in the observed differences between individuals with high vs. low trait-gelotophobia. In Study 2 ($N = 40$), we replicated the pattern of data concerning gaze discrimination ability, even after controlling for individuals' scores on social anxiety. Furthermore, in our second experiment, we found that gelotophobes did not exhibit any problem with identifying others' emotions, or a general incorrect attribution of affective features, such as valence, intensity, or arousal. Therefore, this bias in processing gaze might be related to the global processes of social cognition. Further research is needed to explore how eye contact relates to the fear of being laughed at.

Keywords: gelotophobia, gaze discrimination, eye contact, emotional expression, emotional categorization

OPEN ACCESS

Edited by:
Willibald Ruch,
University of Zurich, Switzerland

Reviewed by:
Kay Brauer,
Martin Luther University of Halle-Wittenberg, Germany
Karl-Heinz Renner,
Universität der Bundeswehr München, Germany

****Correspondence:***
Jorge Torres-Marín
jtorresmarin@ugr.es

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *28 July 2017*
Accepted: *24 October 2017*
Published: *08 November 2017*

Citation:
Torres-Marín J, Carretero-Dios H, Acosta A and Lupiáñez J (2017) Eye Contact and Fear of Being Laughed at in a Gaze Discrimination Task. Front. Psychol. 8:1954. doi: 10.3389/fpsyg.2017.01954

INTRODUCTION

The term gelotophobia (*gelos* in Greek means *laughter*) refers to a personality trait characterized by a disproportionate fear of being laughed at by others (Ruch, 2009). Although this phenomenon was originally conceptualized as a psychopathological disorder (Titze, 2009), recent approaches have operationalized gelotophobia as an individual differences variable that also shows considerable

variation in non-clinical samples (e.g., Ruch and Proyer, 2008). In this sense, those individuals scoring high in trait-gelotophobia —or gelotophobes— are described as neurotic and introverted and as having a paranoid tendency to anticipate derision and mockery situations (Ruch and Proyer, 2009). This misinterpretation of humor related-situations undermines their social interactions, as they are constantly expecting contempt and rejection from others individuals (Ruch et al., 2014a). Research on gelotophobia has gradually progressed over the past two decades (Ruch et al., 2008; Titze, 2009; Platt et al., 2012, 2013; Wu et al., 2016), leading to a theoretical framework of reference that includes major findings concerning both potential triggering causes and moderating factors (e.g., bullying, parental influences or sociocultural factors), as well as consequences (e.g., humourlessness or social withdrawal) linked to gelotophobia predisposition (Ruch, 2009; Ruch et al., 2014a). Nevertheless, it is important to note that still, the nature of the predisposing factors of gelotophobia remains unclear. Contrary to traditional assumptions about the appearance/origin of gelotophobia (Titze, 2009), the presence of the traumatic experiences of teasing during childhood and adolescence does not seem to be a differentiating or invariant aspect of the development of this humor-related trait (Ruch et al., 2010). Therefore, additional research areas such as, for example, perceptual biases toward relevant affective or social cues (e.g., gaze or eye contact) need to be explored. Indeed, it has been stressed that gelotophobia research needs to move toward a more comprehensive and accurate theoretical model (Ruch et al., 2014a).

Recent experimental research on the fear of being laughed at has advanced our knowledge about this phenomenon. For instance, Papousek et al. (2009) designed an experimental task in which participants were exposed to several emotionally contagious films displaying a positive (e.g., cheerfulness), negative (e.g., anxiety or sadness), or neutral mood, with the purpose of comparing gelotophobes' and non-gelotophobes' responses to the emotional states of other individuals. The results revealed that individuals with gelotophobia did not show a reduced emotional induction to positive emotions compared with non-gelotophobes; interestingly, however, they showed a higher degree of affective induction to negative emotions, that is, high scores of subjective anxiety or sadness after watching anxiety- or sadness-causing films, respectively. In line with the analysis of gelotophobes' reactions concerning the affective states of others, Ruch et al. (2015) used the Facial Action Coding System to analyze the potential differences between gelotophobes and non-gelotophobes in joy and contempt responses to videos of laughter-eliciting emotions (e.g., amusement or relief). In particular, they found that gelotophobes exhibited reduced facial expressions of joy (i.e., joyful smiles) and more expressions of contempt when they were exposed to laughter-eliciting emotions. In a different study, Ruch et al. (2014b), by using interactions with virtual agents (i.e., human-like figures or avatars) investigated which features of avatar laughter were considered to be not genuine, threatening, or malicious among individuals who score high on gelotophobia. Their results indicated that, among other factors, a low or mid-level intensity of laughter, an inhibited facial expression, and exaggerated body movements that accompany the laughter may be perceived as more malicious among gelotophobes. In a further investigation, Papousek et al. (2014) developed a realistic and socially relevant context in which participants were interrupted while performing an arithmetic task. The nature of the interruption was manipulated in three experimental conditions: anger provocation together with laughter, anger provocation together with white noise, and no interruption. The cardiac responses of the participants were recorded during the experiment, with a specific reaction of individuals with gelotophobia emerging, that is, a heart rate deceleration in response to others' laughter. According to these authors, this psychophysiological response would be associated with a higher inclination in gelotophobes to interpret laughter as a cue of social rejection. To sum up, gelotophobes, compared with non-gelotophobes, seem to exhibit differentiated emotional manifestations. They are more sensitive to the contagion of negative emotions, show fewer facial expressions related to positive affective states as joyful smiles, and exhibit specific physiological reactions to potential threatening laugher. However, despite the undeniable progress made in the understanding of gelotophobia, further experimental research and new research topics are necessary for deepening the role of the fear of being laughed at in gelotophobes' processing of emotional information.

Smiles, Eye Contact, and Gelotophobia

Numerous authors have discussed the variability of meanings ascribed to a smile as well as the implication of its degree of genuineness or authenticity (Ekman et al., 1990; Ekman, 2003; Johnston et al., 2010). Although a smile is generally labeled as an indicator of a positive affective state, this emotional expression may hide other motivations as to denote, for example, social hierarchy or to mask negative feelings (Niedenthal et al., 2010). Evidence exists that a smile perceived as false or as a non-enjoyment smile is evaluated more negatively and can even lead the perceiver to show less cooperation or trust in comparison with a genuine or enjoyment smile (Johnston et al., 2010). One of the main features related to the fear of being laughed at is the tendency to interpret benevolent or neutral humor-related situations as threatening or malicious (Titze, 2009). Consistent with these findings, gelotophobes also tend to perceive others' smiles as less joyful and more scornful than non-gelotophobes do (Hofmann et al., 2015). This smile misattribution may disturb the adequate social integration of these individuals, thus constituting to the persistence of gelotophobia (Ruch et al., 2014a). Exploring all different cues that may support the recognition of smiles and that may facilitate correct access to the meaning of smiles, especially among individuals with a higher inclination to gelotophobia, is therefore important.

Previous research has indicated that gaze and eye contact play relevant roles in the processes of recognition and inference making with regard to the meanings of others' smiles (Niedenthal et al., 2010). Indeed, gaze entails an essential information source for enhancing our understanding of other people's intentions, facilitating adaptation to our environment and being particularly relevant during social interactions (Argyle and Cook, 1976; Baron-Cohen, 1994; Cañadas and Lupiáñez, 2012). In particular,

according to the simulation of smiles (SIMS) model, eye contact could act as a trigger of an embodied simulation process by which an individual obtains information to identify and interpret smiles (Niedenthal et al., 2010). Another theoretical approach that has highlighted the importance of the gaze direction when individuals have to interpret the intentions or anticipate the actions of others is *theory of mind* (ToM). According to Baron-Cohen (1994, 1995), the capacity to make inferences about others' states of mind, or the "mind reading" system, would consist of a set of modular components, among which would be an eye direction detector (EDD). This module would be involved in the identification of a gaze direction (e.g., direct or averted) and therefore in the subjective perception of being looked at (Cañadas and Lupiáñez, 2012).

It has already been proposed—as a tentative explanation—that an atypically developed ToM could be related to the underlying wrong attribution present in gelotophobes, which would lead them to interpret that people are not laughing with them but rather laughing at them during social interactions (Ruch et al., 2008). Given that gaze discrimination is associated with both access to adequate meanings of smiles as well as expectations about how someone is going to behave, providing useful information for interpreting their objectives or intentions (Hudson et al., 2009; Niedenthal et al., 2010; Hudson and Jellema, 2011), we decided to explore whether higher trait-gelotophobia could be associated with potential bias processing gaze discrimination or eye contact, especially when the looking face portrays a smile. In this sense, a fundamental difficulty in gaze discrimination might underlie interpretation biases, leading gelotophobes to wrongly interpret others' smiles as malicious or false.

To test this point, we used a novel gaze discrimination task that Cañadas and Lupiáñez (2012) developed, with the objective of exploring the importance of social stimuli (i.e., eye contact) in spatial Stroop paradigms. These authors discovered that the identification of a gaze direction is quicker when a face is located to the left but looking to the right, or vice versa (incongruent condition), in comparison with when the face location and eyes' direction match (congruent condition). This reverse congruency effect—classical results with non-social stimuli, such as arrows, show faster responses for congruency trials—was interpreted in terms of eye contact (e.g., responses are faster when a face located to the left looks to the right, i.e., at us). Moreover, further investigation revealed that the emotional charge of the facial expression modulated this eye contact effect (Jones, 2015). More specifically, Jones's results indicated that the effect was stronger for happy and angry faces (approach-oriented emotions) than for neutral faces, and it was non-existent for fearful faces (avoidance-oriented emotions). According to Adams and Kleck (2003), approach-oriented emotions (i.e., happiness and anger) are those that are identified more quickly when the faces displaying these emotions feature direct gazes rather than averted gazes. On the contrary, avoidance-oriented emotions (i.e., fear and sadness) are those that are recognized more quickly when the faces feature averted gazes vs. direct gazes. In this sense, Jones (2015) pointed out that the differences in the observed eye contact effect could be due to the differential facilitation of the processing of each emotion depending on the eye contact condition (e.g., a direct gaze would facilitate the processing of anger or happiness, and an adverted gaze would facilitate the processing of fear or sadness).

EXPERIMENT 1

The purpose of the first experiment was to explore the performance of individuals scoring high vs. low in trait-gelotophobia in a gaze direction discrimination task, which has been previously shown to index an eye contact effect. The emotional expression of the face whose gaze direction had to be discriminated was also manipulated to investigate whether emotion affected the observed eye contact effect as a function of the gelotophobia levels of the participants. We expected trait-gelotophobia to modulate the eye contact effect data, showing specific group differences in the happiness condition. It may be possible for gelotophobes to respond to happy faces in the same way they would respond to fear faces, that is, as an avoidance-oriented emotion. Additionally, to corroborate the adequacy of the "approach or avoidance oriented emotions" interpretation for the reverse congruency (i.e., eye contact) effect that Jones (2015) proposed, and to extend our understanding of the role of emotional expression in the modulation of gaze discrimination, we decided to incorporate faces portraying sadness into our experiment. In accordance with Jones (2015), we expected to replicate the previous results in happiness, anger, neutral, and fear stimuli; regarding sadness, we expected to find a pattern similar to fearful faces and different from angry or happy faces.

Materials and Methods

Participants

From a total sample ($N = 202$) of undergraduate students, 40 (32 females, 8 males; age ranging from 17 to 34; $M = 19.80$, $SD = 2.94$) were selected on the basis of their either extremely high or extremely low scores in trait-gelotophobia, and they were assigned to one of two comparison groups (gelotophobes vs. non-gelotophobes). All participants took part in the experiment voluntarily and received course credits in exchange for their collaboration. They reported normal or corrected-to-normal vision and hearing.

In particular, the selection criterion was the participant's score on the Spanish version of Geloph <15> (Ruch and Proyer, 2008; Carretero-Dios et al., 2010a). The gelotophobes group consisted of the 20 participants who had the highest trait-gelotophobia scores (18 females; 17–25 years; $M_{\text{Geloph}} = 2.76$; $SD_{\text{Geloph}} = 0.35$; $\text{Min}_{\text{Geloph}} = 2.20$; $\text{Max}_{\text{Geloph}} = 3.27$). According to a transcultural investigation (Proyer et al., 2009), gelotophobia scores can be set in the following categories: 1.0–2.0: no gelotophobia; 2.0–2.5: borderline fearful; 2.5–3.0: slight expression of gelotophobia; 3.0–3.5: marked expression of gelotophobia; and 3.5–4.0: extremely fearful of being laughed at. Therefore, of the 20 participants, five were classified as borderline fearful, seven as slight expression of gelotophobia, and eight as marked expression of gelotophobia. Meanwhile, the non-gelotophobes group was also made up of 20 participants but, in this case, with the lowest trait-gelotophobia scores. (14 females;

18–34 years; $M_{Geloph} = 1.38$; $SD_{Geloph} = 0.25$; $Min_{Geloph} = 1.00$; $Max_{Geloph} = 1.80$). These 20 participants were classified as having no gelotophobia. It should be noted that, in order to improve the comparability of the results, we ensured that both comparison groups had the same number of participants ($n = 20$).

The two reported experiments were conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki, following an ethical protocol approved by the University of Granada. All participants participated voluntarily in the studies and provided signed written consent before participating in the experiment.

Instruments

The Spanish version of the Geloph <15> (Ruch and Proyer, 2008; Carretero-Dios et al., 2010a) consists of a self-report questionnaire that assesses trait-gelotophobia, A sample item is "when others laugh in my presence I get suspicious." It includes 15 positively keyed items in a 4-point answer format ranging from 1 (*Strongly disagree*) to 4 (*Strongly agree*). Test reliability (Cronbach's alpha) was $\alpha = 0.94$ in the present sample.

Apparatus and Stimuli

In this experiment, stimuli presentation, timing, and data collection were controlled by using E-Prime 2.0 run on a standard personal computer (PC). Stimuli were presented on a 17″ screen running at a 1024 pixel × 768 pixel resolution. The stimulus material consisted of 40 different full-color photographs (dimensions = 152 pixels × 186 pixels or 5.5 cm × 6.0 cm) of four males and four females portraying either a happy, angry, fearful, neutral, or sad emotional expression. All faces were selected from the Karolinska Directed Emotional Faces (KDEF; Lundqvist et al., 1998). As the original photos featured faces that looked straight ahead, they were manipulated via Adobe Photoshop CS6 for the purpose of changing the gaze directions to the left and right sides. The main selection criteria for the faces were as follows: (a) The gaze was clearly visible while displaying each facial expression (Jones, 2015), and (b) the global hit rate accuracy scores of each individual displaying an emotional expression was higher than 0.49 ($M = 0.66$; $SD = 0.10$) (Goeleven et al., 2008).

Procedure

We used a paradigm similar to that used in previous research (Cañadas and Lupiáñez, 2012; Jones, 2015). Participants performed an experimental task in which they had to discriminate the gaze directions (left or right) of faces that were presented to the left or to the right of fixation, by pressing, as quickly and accurately as possible, the corresponding key on the keyboard. Participants sat approximately 60 cm away from the monitor in a dimly illuminated testing room. Each trial began with the onset of a fixation point (a white cross: 0.5° × 0.5°) located in the center of a black computer screen for 500 ms. Then, a face portraying different emotional expressions was presented either to the left or to the right of the fixation point (approximately at 3.02° away from fixation to the inner edge of the face) and gazing either to the left or to the right (see **Figure 1**). Thus, considering that participants were in the middle, and following the interpretation by Cañadas and Lupiáñez (2012), the gaze direction could be either direct (e.g., a left-looking face presented to the right of fixation, i.e., potentially producing eye contact) or averted (e.g., a left-looking face presented to the left of fixation). Participants had to identify the face's gaze direction by pressing, respectively, the "Z" or "M" key of the computer keyboard when the correct answer was left or right. Feedback on no-response or incorrect response trials was provided via a 220-Hz tone for 700 ms and a short text message. All possible combinations of stimuli, 8 (face identity) × 5 (emotional expression) × 2 (presentation side) × 2 (gaze direction), formed a total of 160 trials. Two blocks of trials with all combinations were presented for a total of 320 trials. Participants completed a practice block of 16 randomly selected trials to familiarize themselves with the task, followed by eight experimental subblocks of 40 randomly selected trials each, with a rest period between blocks. Participants could determine the duration of each rest period.

After performing the experimental task, participants had to fill out, again, the Geloph <15> (Carretero-Dios et al., 2010a) to ensure that they had been assigned to the appropriate groups and hence to enhance the validity of the obtained results.

Design

A 2 (gelotophobia: participants scoring high vs. low on Geloph <15>) × 5 (emotional expression: happiness, anger, fear, neutral, or sadness) × 2 (gaze direction: direct or averted) mixed design was used to analyze the data, with 32 observations per experimental condition. Response times (RTs) and error rates were used as dependent variables. The gelotophobia level was treated as a between-participant variable, and emotional expression and gaze direction as within-participant factors. A two-tailed significance level of $p < 0.05$ was used for all analyses.

Results

Response Time

Taking into account the procedure followed in the original study by Cañadas and Lupiáñez (2012), those trials with RTs shorter than 200 ms or slower than 1300 ms were eliminated from the RT analyses. Mean corrected RTs were submitted to a 2 (gelotophobia) x 5 (emotional expression) x 2 (gaze direction) mixed ANOVA. All response times (RTs) are measured and reported in ms. The results showed a main effect of emotional expression, $F(4, 152) = 29.75$, $p < 0.001$, $\eta_p^2 = 0.44$, with the lowest reaction times being for fearful faces ($M = 618$; $SD = 55.86$) and the highest for angry faces ($M = 651$; $SD = 60.72$). Replicating Cañadas and Lupiáñez (2012), a main effect of gaze direction was also found, $F(1,38) = 68.02$, $p < 0.001$, $\eta_p^2 = 0.64$, with shorter RTs for direct gaze stimuli ($M = 616$; $SD = 53.87$) than for averted gaze stimuli ($M = 655$; $SD = 65.40$). Furthermore, as Jones (2015) showed, the interaction between emotional expression and gaze direction was significant, $F(4,152) = 2.67$, $p = 0.035$, $\eta_p^2 = 0.07$. However, in contrast to Jones's (2015) conclusions regarding the interaction, paired *t*-tests showed that RTs were lower in the direct gaze than in the averted gaze condition for all emotional expressions [$8.11 > t(39) > 5.54$; all $ps < 0.001$, $d = 0.47$–0.79] (see **Figure 2**).

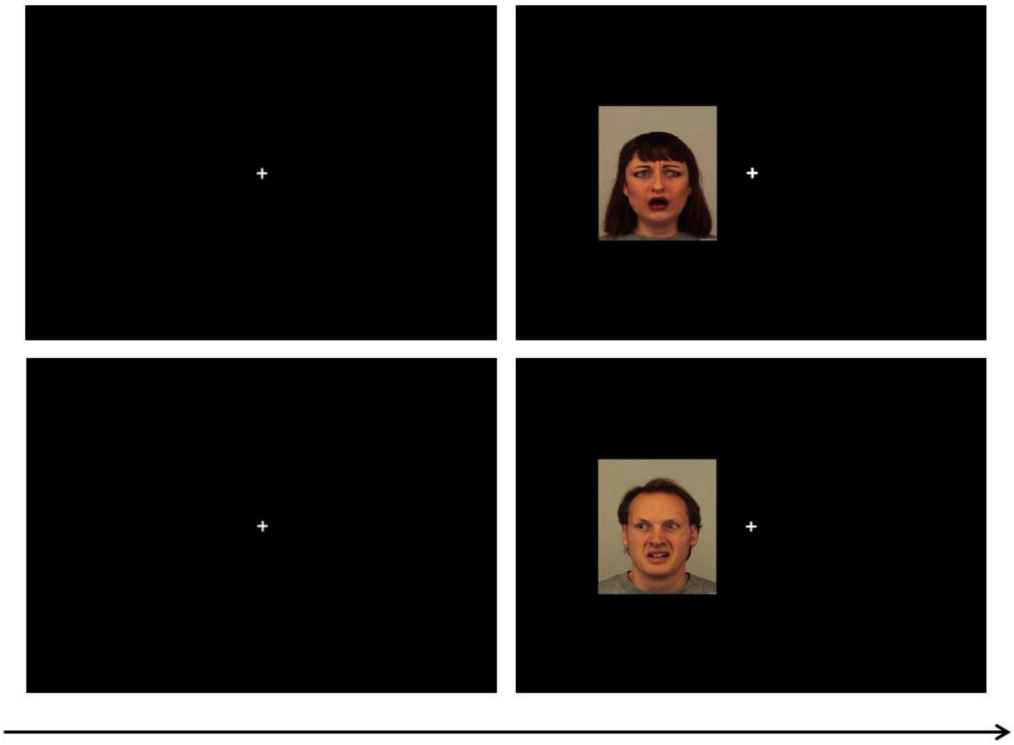


FIGURE 1 | Procedure used in Experiments 1 and 2. The pictures located above illustrate a direct gaze trial (incongruent condition: a left-looking face located to the right), whereas the bottom pictures illustrate an averted gaze trial (congruent condition: a left-looking face located to the left).

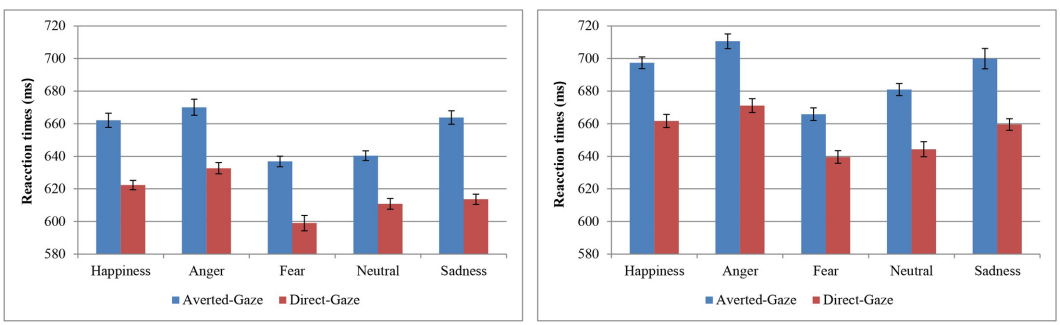


FIGURE 2 | Responses time for gaze discrimination by emotional expression and gaze condition. The results obtained in the Experiment 1 are on the left and Experiment 2 on the right. Error bars represent standard error of the mean computed following Cousineau (2005) method.

Regarding trait-gelotophobia, no main effect of group was observed, $F(1,38) = 0.15$, $p = 0.697$, $\eta_p^2 = 0.004$. Furthermore, and importantly for our hypotheses, gelotophobia did not modulate any effect, especially the emotional expression × gaze direction interaction, $F(4,152) = 0.40$, $p = 0.807$, $\eta_p^2 = 0.01$.

Error Rates of Responses

In a similar pattern to the RT data, the obtained results showed a significant main effect of emotional expression, $F(4,152) = 4.21$, $p = 0.003$, $\eta_p^2 = 0.10$, with a higher error rate for responses to angry ($M = 0.07$; $SD = 0.09$) compared with fearful faces ($M = 0.04$; $SD = 0.06$). A main effect of gaze direction, $F(1,38) = 10.17$, $p = 0.003$, $\eta_p^2 = 0.21$, was also found, with lower error rates for direct gaze ($M = 0.04$; $SD = 0.05$) than for averted gaze stimuli ($M = 0.07$; $SD = 0.10$). Finally, as in the RT analysis, the interaction between emotional expression and gaze direction was significant, $F(4,152) = 3.14$, $p = 0.016$, $\eta_p^2 = 0.08$. To explore this interaction (see **Figure 3**), paired-samples t-tests were employed, and a greater error rate for averted gaze stimuli emerged for happiness, $t(39) = 3.22$, $p = 0.003$, $d = 0.54$; anger, $t(39) = 3.17$, $p = 0.003$, $d = 0.29$; and sadness, $t(39) = 2.76$, $p = 0.011$, $d = 0.41$. Furthermore, in spite of the results just bordered on a statistically significant value, $t(39) = 1.98$, $p = 0.054$, $d = 0.29$, a low effect size was observed for

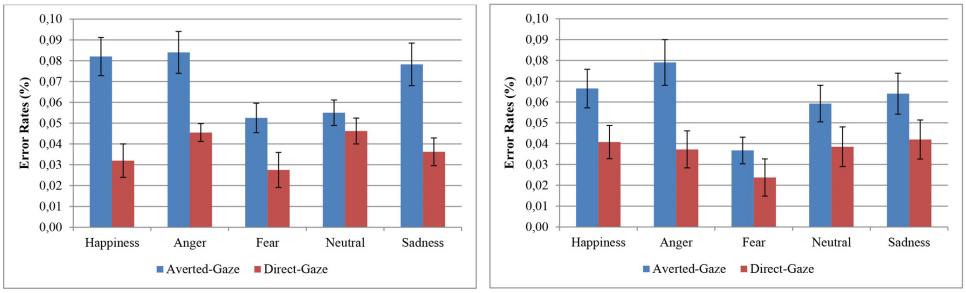


FIGURE 3 | Error rates for gaze discrimination by emotional expression and gaze condition. The results obtained in the Experiment 1 are on the left and Experiment 2 on the right. Error bars represent standard error of the mean computed following Cousineau (2005) method.

fearful faces in accordance with Cohen (1988) criteria. Lastly, no differences were found for neutral faces, $t(39) = 0.84$, $p = 0.404$, $d = 0.12$.

Concerning gelotophobia effects, our data revealed that this predisposition did not modulate the gaze discrimination error rate, $F(1,38) = 2.32$, $p = 0.136$, $\eta_p^2 = 0.06$, but an interaction close to statistical significance between gelotophobia and gaze direction appeared, $F(1,38) = 3.56$, $p = 0.067$, $\eta_p^2 = 0.09$. To explore this interaction, an independent analysis was performed on each gaze direction condition (direct vs. averted). Although our results failed to attain statistical significance at conventional levels, the Cohen values suggested that gelotophobes had higher error rates than non-gelotophobes, specially for averted gaze, $F(1,38) = 2.91$, $p = 0.096$, $d = 0.50$, compared to direct gaze, $F(1,38) = 0.92$, $p = 0.343$, $d = 0.40$.

Given that Ruch and Proyer (2008) derived empirical cut-off points for gelotophobia ($\geq$2.50), and with the aim of avoiding potential limitations of our participant selection, we decided to repeat the above analyses but remove those participants classified as borderline fearful ($n = 5$) in the gelotofobia group. In addition, to balance the two comparison groups, we also removed the five participants of the no-gelotophobia group with the highest scores on the Geloph <15>. Thirty individuals composed our new test sample. Again, two comparison groups were created: 15 gelotophobes (14 females; 17–21 years; $M_{Geloph} = 2.91$; $SD_{Geloph} = 0.24$; $Min_{Geloph} = 2.53$; $Max_{Geloph} = 3.27$) and 15 non-gelotophobes (nine females; 18–34 years; $M_{Geloph} = 1.26$; $SD_{Geloph} = 0.16$; $Min_{Geloph} = 1.00$; $Max_{Geloph} = 1.53$). The RT analysis on the data from the more extremely selected sample did not change from the results with the whole sample. However, with regard to error rates, the new analysis indicated that gelotophobes had significantly higher error rates ($M = 0.05$; $SD = 0.05$) compared with non-gelotophobes ($M = 0.02$; $SD = 0.01$), $F(1,28) = 8.59$, $p = 0.007$, $\eta_p^2 = 0.24$. Additionally, the interaction between gelotophobia and gaze direction was also significant, $F(1,28) = 4.99$, $p = 0.034$, $\eta_p^2 = 0.15$. Again, an independent analysis was performed on each gaze direction condition (direct vs. averted). A between-participant effect emerged for the averted gaze condition, $F(1,28) = 7.65$, $p = 0.010$, $d = 1.03$, showing that gelotophobes had higher error rates ($M = 0.08$; $SD = 0.08$) compared with non-gelotophobes ($M = 0.02$; $SD = 0.02$). Along the same lines, a trend that approached significance and a low effect size, $F(1,38) = 3.71$, $p = 0.064$, $d = 0.39$, emerged for direct gaze stimuli, with higher error rates for gelotophobes ($M = 0.03$; $SD = 0.03$) than for non-gelotophobes ($M = 0.02$; $SD = 0.02$). Finally, and importantly for our hypotheses, the third-order interaction among gelotophobia, emotional expression, and gaze direction (see **Table 1**) did not reach statistical significance, $F(4,112) = 1.33$, $p = 0.263$, $\eta_p^2 = 0.05$.

Discussion

As we expected, the results of the present experiment confirmed that gaze direction and emotion modulate reaction time in gaze discrimination. In line with Cañadas and Lupiáñez (2012), participants were faster and more accurate at identifying a gaze direction when the face was presented to the left but looking to the right (direct gaze) than the same face location but looking to the left (averted gaze). These data reinforce the eye contact interpretation of this reversed congruency effect and entail new evidence regarding its robustness. Importantly, our results indicated a similar pattern for RT and accuracy data in contrast to other authors who have reported that gaze direction does not modulate accuracy in a gaze-cueing paradigm (Prinzmetal et al., 2008). Furthermore, we found that emotional expression influenced our eye contact effect, but in a way that is inconsistent with the "approach and avoidance oriented emotions" interpretation by Jones (2015). In fact, the expression of sadness, which has been considered an avoidance-oriented emotion (Adams and Kleck, 2003), showed a pattern similar to that of approach-oriented emotions (e.g., happiness and anger). Similarly, emotional expression modulated the gaze direction effect in error rates as well. Participants showed lower error rates in identifying gaze directions with fearful faces compared with angry faces. In addition, more interestingly, direct gaze facilitated performance leading to higher accuracy, i.e., lower error rates for all emotional expressions with the exception of neutral faces. Therefore, although the "approach and avoidance oriented emotions" interpretation by Jones (2015) was not supported, the pattern of results supported the social nature of the reverse congruency effect observed, and therefore its interpretation in terms of eye contact (Cañadas and Lupiáñez, 2012). Eye contact is important in human communications (Doherty-Sneddon and Phelps, 2005), particularly in those

TABLE 1 | Means of reaction times (in ms) and error rates for gaze discrimination in Experiments 1 and 2 for each condition and gelotophobia group.

			High gelotophobia		Low gelotophobia	
	Emotional expression	**Eye contact**	**RT**	**% Errors**	**RT**	**% Errors**
Experiment 1	Happiness	Averted	669	0.10	655	0.07
		Direct	623	0.04	622	0.03
	Anger	Averted	673	0.11	668	0.06
		Direct	635	0.06	630	0.04
	Fear	Averted	644	0.08	630	0.02
		Direct	600	0.03	598	0.02
	Neutral	Averted	647	0.08	634	0.03
		Direct	616	0.06	606	0.04
	Sadness	Averted	666	0.12	662	0.04
		Direct	616	0.04	612	0.03
Experiment 2	Happiness	Averted	675	0.10	721	0.04
		Direct	641	0.06	682	0.03
	Anger	Averted	693	0.12	729	0.04
		Direct	643	0.05	699	0.02
	Fear	Averted	646	0.06	685	0.01
		Direct	613	0.04	666	0.01
	Neutral	Averted	655	0.09	707	0.03
		Direct	621	0.06	668	0.02
	Sadness	Averted	679	0.09	721	0.04
		Direct	640	0.06	679	0.02

interactions where emotional expression is present (Milders et al., 2011).

With respect to gelotophobia, and in relation to RT, we found no evidence for any modulation of gelotophobia in gaze discrimination. However, and interestingly, individuals with high trait-gelotophobia tend to make more errors when they have to discriminate gaze direction. The ability to detect correctly gaze direction is associated with the appropriate interpretation of others' intentions (Baron-Cohen, 1994; Hudson and Jellema, 2011). Given that wrong attributions on the motivations and goals of other individuals could be considered one of the main components of gelotophobia (Ruch et al., 2008; Titze, 2009), this potential bias related to gaze identification could be a relevant finding to better understand the fear of being laughed at. Furthermore, the interaction between gelotophobia and gaze direction was significant, showing that the higher error rates observed in gelotophobes was larger in averted gaze trials than in direct gaze trials. Nevertheless, the independent analyses of the direct gaze condition also showed a low effect size for gelotophobia, so it seems necessary to explore this interaction further.

Finally, and in contrast to our hypothesis, happiness did not seem to have any special role in the observed differences between individuals with high and low trait-gelotophobia. Previous research has shown that gelotophobia may influence reactions to others' affective states but not just those related to happiness (Papousek et al., 2009). However, inasmuch as we had to reduce our testing sample to adjust it to the reported cut-off points for gelotophobia ($\geq$2.50) (Ruch and Proyer, 2008), we carried out an additional experiment to confirm the observed pattern of data, thus avoiding this potential limitation of the research. In addition, and importantly, to test whether our findings are specific to gelotophobia, in the next experiment, we controlled for social phobia as an alternative explanation of the observed effect of gelotophobia.

EXPERIMENT 2

In Experiment 2, we tried to replicate the results observed in the preceding experiment, but controlling for the potential limitations highlighted above. The newly recruited participants for the high- and low-gelotophobia groups showed greater differences in their trait-gelotophobia scores. Then, we tested again whether individuals scoring high in trait-gelotophobia indeed have higher error rates in detecting gaze direction compared with individuals with lower trait-gelotophobia scores. Moreover, we explored the interaction between gelotophobia and gaze direction with the aim of confirming our previous finding that a larger gelotophobia predisposition could be associated with poorer performance, especially with averted gaze in comparison with direct gaze conditions. Finally, we were interested in analyzing the third-order interaction among gelotophobia, emotional expression, and gaze direction once again to corroborate that the happiness condition does not play any specific role in the eye contact effect that gelotophobes show.

Another main objective in gelotophobia research is to determine its differential features in relation to other disorders with similar symptomatology (e.g., social phobia) (Carretero-Dios et al., 2010b). Actually, previous research studies have reported that a high percentage of gelotophobes are also assessed as individuals with social phobia and/or Cluster A (i.e., schizoid,

paranoid, or schizotypal) personality disorder (Weiss et al., 2012). Consequently, we included social phobia as a control variable to investigate whether the effects of gelotophobia could be explained on the basis of differences in social phobia.

In addition, we added a new experimental phase related to the identification of others' emotional expressions. Although previous research indicated that gelotophobes did not have a general deficit in interpersonal emotion-related skills so as to categorize the emotions of others (Papousek et al., 2009), the goal of this second phase was to test whether eye contact conditions (direct vs. averted) modulate gelotophobes' capacity to identify others' emotional expressions. The manipulation of gaze direction in our preceding experiment seemed to be relevant, so we were interested in knowing whether gelotophobes would show a different pattern of emotion categorization depending on gaze conditions. Furthermore, all participants evaluated the intensity of the expressed emotion together with the valence and arousal of each face.

Materials and Methods

Participants

Undergraduate students ($N = 241$) were screened using the Geloph <15>. The Sample included a total of 40 participants (32 females, 8 males; mean age of 21.18, $SD = 6.34$; range from 18 to 49) who were selected on the basis of their either extremely high or extremely low scores in trait-gelotophobia and assigned to one of the two comparison groups (gelotophobes and non-gelotophobes). As in Experiment 1, all participants reported normal or corrected-to-normal vision and hearing, and participants' collaboration was in exchange for course credit. None of the participants had participated in Experiment 1.

The gelotophobes group was made up of 20 participants who had the highest trait-gelotophobia scores (16 females; 17–27 years; $M = 20.00$; $SD = 3.06$; $M_{\text{Geloph}} = 2.93$; $SD_{\text{Geloph}} = 0.39$; $\text{Min}_{\text{Geloph}} = 2.53$; $\text{Max}_{\text{Geloph}} = 3.60$). In contrast to Experiment 1, all participants in this study exceeded the cut-off point for gelotophobia (>2.50; see Ruch and Proyer, 2008). Thus, of these 20 participants with high trait-gelotophobia scores, none was classified as borderline fearful, 11 were classified as slight expression of gelotophobia, seven were classified as marked expression of gelotophobia, and two were classified as extremely fearful of being laughed at. Likewise, the non-gelotophobes group was made up of 20 participants whose scores were the lowest in the GELOPH <15> (16 females; 18–49 years; $M = 22.35$; $SD = 8.39$; $M_{\text{Geloph}} = 1.24$; $SD_{\text{Geloph}} = 0.17$; $\text{Min}_{\text{Geloph}} = 1.00$; $\text{Max}_{\text{Geloph}} = 1.53$). As in Experiment 1, these individuals were classified as having no gelotophobia.

Instruments

The Spanish version of the Geloph <15> was also used in this experiment with test reliability (Cronbach's alpha) $\alpha = 0.96$.

The Spanish version of the Social Interaction Anxiety Scale (SIAS; Mattick and Clarke, 1998; Olivares et al., 2001) consists of 20 items rated on a Likert-type scale ranging from 0 (*Not at all*) to 4 (*Totally*). A sample item is "I get nervous if I have to speak with someone in authority (teacher, boss, etc.)." In this study, the SIAS showed adequate good internal consistency (Cronbach's alpha = 0.94).

Apparatus, Stimuli, and Procedure

The same procedure as in Experiment 1 was used in the first phase of this experiment. Additionally, a second experimental task was added in which participants had to identify the emotional expressions of faces with direct vs. averted gaze. For this new task, 160 photographs of 16 individuals, eight males and eight females, portraying either a happy, angry, fearful, neutral, or sad emotional expression, were also selected from the KDEF (Lundqvist et al., 1998). Stimuli were different from those used in the gaze discrimination task. Photographs did not have to be modified to recreate eye contact conditions. Each target face was presented for an unlimited time at the center of the monitor either with a direct gaze (i.e., the eyes looking straight ahead) or an averted gaze (i.e., the eyes looking left or right). Participants had to categorize the emotional expression by pressing the corresponding key on the keyboard ("1 = happiness," "2 = anger," "3 = fear," "4 = neutral," or "5 = sadness"). After each categorization, and while the picture remained visible, participants indicated their estimation of different affective dimensions—valence, intensity, and arousal—for that facial expression based on the Self-Assessment Manikin (SAM: Lang, 1980). Only one experimental block composed of 160 trials, 16 (faces) × 5 (emotion) × 2 (gaze direction), was created. Hence, we obtained 16 observations per gaze direction condition displaying each emotional expression. Trials were presented randomly for each participant. Finally, participants responded to gelotophobia and social phobia questionnaires, in that order.

Design

For the gaze discrimination task, the same design was used as in Experiment 1. For the analysis of the ratings in the emotional expression task, a similar design was used, 2 (gelotophobia: high trait-gelotophobia vs. low trait-gelotophobia) × 5 (emotional expression: happiness, anger, fear, neutral, or sadness) × 2 (gaze direction: direct or averted), with the following dependent variables (DVs): (a) *reaction time*; (b) *accuracy of responses* in the emotional categorization task; (c) *intensity* or magnitude of the emotion expressed (high vs. low); (d) *valence* or pleasantness of the faces displaying either emotional expression (positive vs. negative); and (e) the *arousal* or activation of these faces (active vs. calm). Again, gelotophobia predisposition was manipulated between participants, whereas the other variables were manipulated within participants. Furthermore, in all analyses, social phobia scores were introduced as a covariate to determine whether the specific effects are related to gelotophobia independently of social phobia. A two-tailed significance level of $p < 0.05$ was used for all analyses.

Results

Gaze Direction Discrimination Task

Response time data showed, again, a main effect of emotional expression, $F(4,152) = 30.16$, $p < 0.001$, $\eta_p^2 = 0.44$, with the lowest reaction times for fearful faces ($M = 653$; $SD = 64.71$) and the highest for angry faces ($M = 691$; $SD = 67.24$). As

in the previous experiment, a main effect of gaze direction was also found, $F(1,38) = 41.18$, $p < 0.001$, $\eta_p^2 = 0.52$, with participants having shorter RTs for direct gaze ($M = 655$; $SD = 62.42$) than averted gaze faces ($M = 691$; $SD = 69.61$). However, the interaction between emotional expression and gaze direction was not significant in this case, $F(4,152) = 1.33$, $p = 0.26$, $\eta_p^2 = 0.03$. Furthermore, there was a main effect of group, $F(1,38) = 5.58$, $p = 0.023$, $\eta_p^2 = 0.13$, with gelotophobes ($M = 651$; $SD = 70.90$) being faster compared with non-gelotophobes ($M = 696$; $SD = 48.83$). Nevertheless, this effect disappeared after controlling for social phobia scores, $F(1,37) = 1.96$, $p = 0.170$, $\eta_p^2 = 0.05$. As in the Experiment 1, the interaction between emotional expression and gaze direction was not modulated by gelotophobia, $F(4,152) = 1.48$, $p = 0.210$, $\eta_p^2 = 0.04$.

As in our previous experiment, the analysis of *error rate data* also showed a main effect of emotional expression, $F(4,152) = 7.60$, $p < 0.001$, $\eta_p^2 = 0.17$. Again, participants had the lowest error rate for fearful ($M = 0.03$; $SD = 0.05$) and the highest for angry faces ($M = 0.06$; $SD = 0.07$). However, the difference between direct gaze and averted gaze did not reach significance this time, $F(1,38) = 2.70$, $p = 0.109$, $\eta_p^2 = 0.07$, and neither was an interaction found between emotional expression and gaze direction, $F(4,152) = 1.60$, $p = 0.179$, $\eta_p^2 = 0.04$. Concerning gelotophobia, our results replicated the significant main effect of group, $F(1,38) = 6.68$, $p = 0.014$, $\eta_p^2 = 0.15$, with gelotophobes having higher error rates ($M = 0.07$; $SD = 0.08$) compared with non-gelotophobes ($M = 0.03$; $SD = 0.03$). Interestingly, this effect remained significant after controlling for individual social phobia scores, $F(1,37) = 5.54$, $p = 0.024$, $\eta_p^2 = 0.13$. Additionally, and in contrast to Experiment 1, the interaction between gelotophobia and gaze (see **Figure 4**) was not statistically significant, $F(1,38) = 0.14$, $p = 0.708$, $\eta_p^2 = 0.004$. Notwithstanding, a trend close to being significant and a medium effect size according to Cohen' (1988) criteria, were found for the averted gaze condition, $F(1,38) = 3.54$, $p = 0.067$, $d = 0.62$, with gelotophobes having higher error rates ($M = 0.09$; $SD = 0.13$) compared with non-gelotophobes ($M = 0.03$; $SD = 0.04$), and it was significant for the direct gaze condition, $F(1,38) = 8.65$, $p = 0.006$, $d = 0.79$, with gelotophobia predispositions being associated, again, with higher error rates ($M = 0.05$; $SD = 0.05$) in comparison with lower gelotophobia ($M = 0.02$; $SD = 0.02$). Both effects remained after controlling for social phobia, $F(1,37) = 3.59$, $p = 0.066$, and $F(1,37) = 4.80$, $p = 0.036$, respectively.

Emotional Expression Categorization Task

The *analysis of RTs* showed a main effect of emotional expression, $F(4,152) = 32.35$, $p < 0.001$, $\eta_p^2 = 0.46$, with happiness faces identified significantly more quickly ($M = 2137$; $SD = 614.52$) than all other emotions were. Gaze direction and the interaction between emotional expression and gaze direction did not modulate any effect. It should be noted that RT was not limited in this phase. Furthermore, individuals with gelotophobia showed a tendency to respond more quickly when they identified others' emotional expressions, $F(1,38) = 5.87$, $p = 0.020$; $d = 0.77$, $\eta_p^2 = 0.13$. Nevertheless, this effect disappeared after the inclusion of social phobia as a covariate, $F(1,37) = 0.99$, $p = 0.327$; $\eta_p^2 = 0.03$. On the other hand, the interaction between gelotophobia and emotional expression, $F(4,152) = 2.02$, $p = 0.095$, $\eta_p^2 = 0.05$, did not reach statistical significance. Finally, and interestingly with regard to our hypothesis, individuals with higher gelotophobia scores did not differ in their RTs due to gaze direction conditions, $F(1,38) = 0.29$, $p = 0.865$; $\eta_p^2 = 0.001$.

On the other hand, the *analysis of accuracy* data indicated a main effect of emotional expression, $F(4,152) = 8.47$, $p < 0.001$; $\eta_p^2 = 0.18$. The highest accuracy was observed for faces displaying happiness ($M = 0.97$; $SD = 0.05$) and the lowest for neutral faces ($M = 0.86$; $SD = 0.15$). On the contrary, no main effect of gaze direction, $F(1,38) = 0.34$, $p = 0.566$; $\eta_p^2 = 0.01$, was found, and the interaction between emotional expression and gaze direction did not reach statistical significance, $F(4,152) = 1.47$, $p = 0.213$, $\eta_p^2 = 0.04$. Importantly, the results showed that gelotophobia predisposition did not have any effect, $F(1,38) = 0.40$, $p = 0.531$; $\eta_p^2 = 0.01$, and did not modulate the effects of emotional expression, $F(4,152) = 1.17$, $p = 0.328$, $\eta_p^2 = 0.18$, or gaze, $F(1,38) = 0.62$, $p = 0.805$; $\eta_p^2 = 0.002$.

Emotional Expression Rating Task: Intensity, Valence, and Arousal

Emotional expression, $F(4,152) = 15.80$, $p < 0.001$, $\eta_p^2 = 0.29$, modulated the reported *intensity*. Particularly, the neutral expression had the lowest reported levels ($M = 5.42$; $SD = 1.62$) and the happiness expression the highest ones ($M = 6.75$; $SD = 1.14$). Gaze direction also modulated intensity, $F(4,38) = 12.84$, $p < 0.001$, $\eta_p^2 = 0.25$, with greater intensity levels reported for direct gaze ($M = 6.20$; $SD = 1.03$) than for averted gaze ($M = 6.08$; $SD = 1.11$). Furthermore, the interaction between emotional expression and gaze direction was also significant, $F(4,152) = 3.95$, $p = 0.004$, $\eta_p^2 = 0.09$. Paired t-tests showed that happy faces, $t(39) = 5.30$, $p < 0.001$, $d = 0.34$, with direct gazes ($M = 6.94$; $SD = 1.09$) were assessed with a greater level of intensity in comparison with happy faces with averted gazes ($M = 6.54$; $SD = 1.23$). A similar pattern was found for angry faces, $t(39) = 1.98$, $p = 0.055$, $d = 0.14$, with larger intensity reports for direct gaze ($M = 6.41$; $SD = 1.19$) than for averted gaze ($M = 6.23$; $SD = 1.36$). No differences were found for the other emotional expressions. The main effect of the gelotophobia group did not reach statistical significance, $F(1,38) = 0.24$, $p = 0.628$, $\eta_p^2 = 0.01$, although an interaction close to statistical significance between emotional expression and gelotophobia was observed, $F(4,152) = 2.28$, $p = 0.063$, $\eta_p^2 = 0.06$. Nevertheless, it completely disappeared after controlling for social phobia, $F(4,148) = 0.93$, $p = 0.446$, $\eta_p^2 = 0.03$.

Concerning valence and arousal, as expected, the *valence* ratings were modulated by emotional expression, $F(4,152) = 172.83$, $p < 0.001$, $\eta_p^2 = 0.82$. No evidence was found, however, that gaze direction modulated the valence ratings, $F(1,38) = 0.21$, $p = 0.653$, $\eta_p^2 = 0.01$. Interestingly, a significant interaction between emotional expression and gaze direction was also observed, $F(4,152) = 3.37$, $p = 0.011$, $\eta_p^2 = 0.08$. Paired *t*-tests, $t(39) = 2.85$, $p = 0.007$, $d = 0.21$, showed that happy faces with direct gazes ($M = 7.27$; $SD = 0.98$) were evaluated as

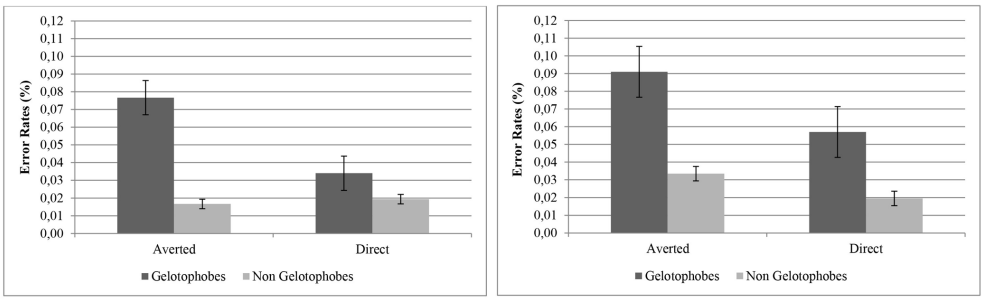


FIGURE 4 | Error rates for gaze discrimination as a function of gaze condition and Gelotophobia group. The results from Experiment 1 are plotted on the left and those for Experiment 2 on the right. Error bars represent standard error of the mean computed following Cousineau (2005) method.

more positive than happy faces with averted gazes ($M = 7.06$; $SD = 1.02$). In contrast, angry faces with direct gazes were evaluated as less positive ($M = 3.05$; $SD = 1.02$) than angry faces with averted gazes ($M = 3.20$; $SD = 0.94$), $t(39) = -2.09$, $p = 0.044$, $d = 0.15$. No differences for the other emotional expressions were found. On other hand, neither the main effect of gelotophobia, $F(1,38) = 1.34$, $p = 0.255$, $\eta_p^2 = 0.03$, nor its modulation over emotional expression, $F(4,152) = 0.07$, $p = 0.992$, $\eta_p^2 = 0.002$, or gaze direction, $F(1,38) = 0.58$, $p = 0.451$, $\eta_p^2 = 0.02$, reached statistical significance.

Finally, emotional expression also influenced the participants' perceptions of *arousal*, $F(4,152) = 48.64$, $p < 0.001$, $\eta_p^2 = 0.56$, whereas gaze direction did not, $F(1,38) = 0.31$, $p = 0.583$, $\eta_p^2 = 0.01$. Nevertheless, the interaction between emotion and gaze was also significant, $F(4,152) = 2.67$, $p = 0.034$, $\eta_p^2 = 0.07$, and paired *t*-tests were used to explore this interaction. Differences in arousal were found for angry faces, $t(39) = 2.53$, $p = 016$, $d = 0.20$, with a greater arousal associated with direct gaze ($M = 6.52$; $SD = 0.97$) vs. averted gaze ($M = 6.33$; $SD = 1.07$). Finally, no main effect of group, $F(1,38) = 0.42$, $p = 0.522$, $\eta_p^2 = 0.01$, or interaction involving gelotophobia and emotional expression, $F(4,152) = 0.10$, $p = 0.984$, $\eta_p^2 = 0.003$, or gaze, $F(1,38) = 0.48$, $p = 0.491$, $\eta_p^2 = 0.01$, reached statistical significance for the arousal ratings. Those results concerning the abovementioned third interaction can be seen in **Table 2**.

GENERAL DISCUSSION

In this study, we explored the modulation that a higher trait-gelotophobia produced in a task in which individuals were asked to discriminate the directions of the gazes of faces portraying different emotions. In particular, we were interested in examining the RTs and the error rates of gelotophobes to discriminate adequately the left-right direction of others' eyes, as a function of whether they conformed direct vs. averted gaze conditions. To our knowledge, this is the first empirical study investigating the potential effects of gelotophobia in reaction to eye contact. In contrast to our initial hypothesis, and compared with non-gelotophobes, gelotophobes did not show a differential eye contact effect for happy faces. However, our results revealed a potential tendency among individuals with a greater degree of gelotophobia to make more error rates when identifying gaze direction. Interestingly, this potential bias in gaze discrimination is rather general, as it does not seem to be associated with a specific emotion or, according to our second experiment, the eye contact condition (direct or averted gaze). In fact, gelotophobes constantly exceeded—in terms of error rates—non-gelotophobes

TABLE 2 | Means RTs and percentages of correct responses, and affective dimensions evaluations, for each condition and gelotophobia group, in the emotional categorization task of Experiment 2.

		High gelotophobia					Low gelotophobia				
Emotional expression	**Eye contact**	**RT**	**% Correct**	**Intensity**	**Valence**	**Arousal**	**RT**	**% Correct**	**Intensity**	**Valence**	**Arousal**
Happiness	Averted	2108	0.97	6.57	7.15	5.30	2236	0.97	6.52	6.97	5.64
	Direct	2113	0.96	7.05	7.37	5.43	2091	0.99	6.84	7.17	5.73
Anger	Averted	2975	0.88	6.17	3.28	6.36	3672	0.91	6.29	3.13	6.31
	Direct	2823	0.91	6.40	3.15	6.43	3590	0.93	6.43	2.94	6.63
Fear	Averted	3416	0.86	6.12	3.60	6.13	3876	0.88	6.20	3.32	6.18
	Direct	3369	0.84	6.12	3.48	5.90	3828	0.90	6.30	3.30	6.21
Neutral	Averted	2799	0.89	5.93	4.83	3.75	3575	0.84	4.99	4.81	3.93
	Direct	3042	0.87	5.70	5.02	3.76	3614	0.83	5.04	4.75	3.82
Sadness	Averted	3113	0.89	5.97	3.30	4.83	3638	0.93	6.04	3.24	5.01
	Direct	2905	0.87	6.16	3.31	4.82	3812	0.90	5.97	3.26	4.96

when they had to detect correctly the eyes' directions of the different faces.

Detecting correctly gaze direction or eye contact is widely considered as a crucial factor in the communication of social intentions or desires (e.g., Argyle and Cook, 1976), to modulate social cognition processes as person categorization (e.g., Macrae et al., 2002) and also to obtain key elements concerning the mental states of others (e.g., Baron-Cohen, 1995). Traditional conceptualizations of gelotophobia have included a poorly developed social competence among its features (Titze, 2009). These limited social skills are characterized, for example, by a widespread fear of acting in a socially inadequate way ("maybe funny"), a feeling of insecurity, hypervigilance toward all possible contempt manifestations of social partners, and a general belief in the negative intentions of others (Platt et al., 2012; Ruch et al., 2014a).

In accordance with Baron-Cohen (1994), difficulties with discriminating others' gaze direction could lead to wrong interpretations of others' intentions or mental states. This is because individuals use the information provided by gaze in order to clarify ambiguous situations and, thus, to judge correctly intentions or acts of others (Phillips et al., 1992). Accordingly, a greater difficulty in knowing where other people are exactly looking at could be connected with the misattributions of others' intentions that gelotophobes make during social interactions, as well as the incorrect access to the real meanings of some more complex emotional expressions. In this sense, given that gelotophobes seem to be less able to identify accurately the direction of other's eyes (i.e., and therefore whether the attention is focused at a particular point or not), may contribute to perceive social interactions as ambiguous or, even, threatening. Furthermore, these results are in line with previous studies, which reported that gelotophobes show difficulties in adequately interpreting facially expressed communication (Ruch et al., 2014b).

In addition, it has been demonstrated that ambiguous eye contact conditions influence the subjective feeling of being observed (Senju and Hasegawa, 2006). Theoretical considerations and empirical data have supported the importance of studying eye contact and, more specifically, the feeling of being observed in relation to several disorders, such as social anxiety. For this reason, the goal of previous research was to determine the contextual cues that exacerbate the feeling of being looked at (Gamer et al., 2011). These authors found that a higher social phobia inclination would be associated only with a greater tendency to judge a "mutual gaze" in situations with a light social pressure (i.e., a second observer is present during the interaction), but not in one-to-one conditions. Given that our experimental setting recreated a one-to-one interaction, this may help with explaining why social anxiety cannot explain the bias revealed for gaze discrimination, i.e., why this bias rather seems to be specific to gelotophobia. Additionally, in our second experiment, we found that a higher gelotophobia predisposition could be related to faster responses regardless of eye contact conditions—direct or averted—in the gaze discrimination task. However, this effect disappeared after controlling for social phobia scores. These results could be due to the tendency of individuals with high social anxiety to be hypervigilant toward threatening social cues (Eysenck, 1992; Boll et al., 2016), such as eye contact, which is an indicator of the beginning of a social interaction.

Consistent with previous research (Cañadas and Lupiáñez, 2012; Jones, 2015), we replicated a reversed congruency effect. Furthermore, in general, the emotional expressions of faces modulate this effect: Although the interaction was not significant in Experiment 2, the same tendency was observed, and the combined analysis of the two experiments showed a significant interaction for both RT, $F(4,316) = 2.48$, $p = 0.044$, $\eta_p^2 = 0.03$, and error rates, $F(4,316) = 3.30$, $p = 0.011$, $\eta_p^2 = 0.04$. This is important, as it favors the interpretation of the reversed congruency effect in terms of eye contact. Nevertheless, it should be noted that in contrast to the pattern of results that Jones (2015) reported, the effect was also observed for the fearful expression; furthermore, sadness (theoretically, an avoidance-oriented emotion) showed a pattern similar to those of happiness and anger (approach-oriented emotion) in both experiments. For this reason, we cannot corroborate the "approach and avoidance oriented emotions" interpretation that Jones (2015) suggested. Thus, additional studies of our eye contact effect should look into other different frameworks used to explain the interaction between facial expression and gaze direction, as the *appraisal* theory (Sander et al., 2007). This theory focuses on the importance of the observer's goals or intentions when interpreting or evaluating (appraisal process) the meaning of all of the external social clues (Sander et al., 2007; Milders et al., 2011). Perhaps, sadness could trigger avoidance motivation in others but also feelings of compassion or approach behavior to offer occasional help to the observer. Nevertheless, it is important to note the need for developing further research to elucidate the relationship between sadness and the reverse congruency effect data and, more generally, the role of emotional expression in this unusual effect. Furthermore, in this research, we incorporated the data of error rates in the gaze discrimination task. We observed a main effect of emotional expression, which was replicated in both experiments, with the highest error rates in faces expressing anger and the lowest in faces expressing fear. Importantly, the joint analysis of the error data in these two experiments revealed that this eye contact effect was stronger in faces displaying emotional expressions—with the exception of fearful faces—compared with neutral faces.

Concerning the emotional expression categorization task, we found that gelotophobes were faster when they had to categorize others' emotional expressions, but as in the previous gaze discrimination task, this effect disappeared after controlling for social phobia scores. No accuracy differences between gelotophobes and non-gelotophobes in identifying others' emotional expressions were found, with our results being consistent with the notion that gelotophobes do not present difficulties in the use of interpersonal emotion-related skills (Papousek et al., 2009). Interestingly, neither did we find any interaction among gelotophobia, emotional expression, and gaze direction for intensity, valence, or arousal. In sum, our results seem to indicate that eye contact conditions do not modulate the gelotophobes' ratings of these affective dimensions.

Aside from gelotophobia effects, however, it should be noted that in this categorization task emotional expression modulated both RT and accuracy. More specifically, we found that happiness trials produced the fastest RTs and highest accuracy rates. Furthermore, the lowest accuracy rates were found in neutral faces. This data could be in line with previous studies reporting that emotional expressions, in comparison with neutral faces, facilitated processes such as face detectability (de Jong and Martens, 2007; Calvo and Nummenmaa, 2008; Milders et al., 2011). Along the same lines, we know that neutral faces contain affective keys more ambiguously than others' emotional expressions. Indeed, authors as Zebrowitz et al. (2010) pointed out that neutral faces are often wrongly labeled as faces displaying anger in males or faces portraying surprise in females. In addition, a fewer number of neutral trials exist compared with emotional faces, which can lead to interpreting neutral faces as an emotional expression.

With respect to the affective dimensions measured, our results showed an interaction between emotion and gaze direction for intensity, valence, and arousal rates. More specifically, we found that faces displaying anger or happiness with direct gazes were evaluated as more intense than those with averted gazes. These results are consistent with other studies that have proposed that gaze direction modulates the recognition accuracy and perceived intensity of several emotions (Adams and Kleck, 2005). Consistent with intensity data, we obtained for valence an opposite pattern between anger and happiness. Whereas happy faces with direct gazes were rated as more positive than happy faces with averted gazes, angry faces with direct gazes were evaluated as more unpleasant than angry faces with averted gazes. Finally, differences in arousal were found for angry faces with a greater arousal associated with direct gaze than with averted gaze. The observed interaction between emotional expression and gaze direction fits with other empirical data suggesting that the processing of emotional expression and the processing of gaze pattern are interdependent (Ganel et al., 2005).

CONCLUSION

The current results provide the first preliminary empirical evidence that gelotophobia is related to a potential bias in gaze discrimination. The effects of gelotophobia on error rates in discriminating gaze direction were replicated in two experiments. Furthermore, in the second experiment, the effect remained when controlling for social anxiety scores. Taking into account that gelotophobes, on the other hand, did not show any difference with non-gelotophobes in discriminating emotional expression, or intensity, arousal, or valence, our results could suggest that the gaze discrimination difficulties observed in high gelotophobes are not associated with problems with identifying others' emotions or an incorrect attribution of affective features. These higher error rates in gaze direction accuracy might not be due to any limitation in processing affective information but rather might be related to global processes of social cognition. However, future research should clarify and continue exploring social cognition biases in gelotophobia to analyze the potential consequences of feeling being observed.

Several limitations of this research must be nevertheless pointed out. Firstly, due to the low prevalence of gelotophobes in non-clinical population, the sample sizes were relatively small. However, both the number of participants selected and the strategy adopted for recruiting them (i.e., construction of extreme groups) were in the line with previous research concerning gelotophobia (Papousek et al., 2014; Ruch et al., 2015). Lastly, it is important to indicate that some particular laughter (or humor)-related aspects were not included in these studies. Indeed, the use of pictures does not allow for the incorporation of key emotional components, such as sounds or movements. Therefore, it is possible that these stimuli may be insufficient to trigger some gelotophobes' specific reactions and can help with explaining the absence of a specific effect on eye contact in the happiness condition. For this reason, future research should add other materials (e.g., films, virtual reality, etc.) with the aim of creating more realistic scenarios of emotional interactions where laughter is present, which will surely be a significant step forward for the main purpose of this research.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: JT-M, HC-D, AA, and JL. Performed the experiments: JT-M. Analyzed the data: JT-M, HC-D, AA, and JL. Interpreted the data and drafted the manuscript: JT-M, HC-D, AA, and JL.

FUNDING

This research is part of the doctoral dissertation by JT-M, which is supported by the Spanish Ministerio de Educación, Cultura y Deporte with a predoctoral fellowship (FPU14/05755) and with research grants from the Spanish Ministerio de Economía, Industria y Competitividad (MINECO) (PSI2014-52764-P to JL), and Dirección General de Investigación Científica y Técnica-Ministerio de Educación y Ciencia (DGICYT-MEC) (PSI2016-78236-P to AA and PSI2016-79812-P to HC-D).

REFERENCES

Adams, R. B., and Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychol. Sci.* 14, 644–647. doi: 10.1046/j.0956-7976.2003.psci_1479.x

Adams, R. B., and Kleck, R. E. (2005). The effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion* 5, 3–11. doi: 10.1037/1528-3542.5.1.3

Argyle, M., and Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.

Baron-Cohen, S. (1994). How to build a baby that can read minds: cognitive mechanisms in mindreading. *Curr. Psychol. Cogn.* 13, 513–552.

Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.

Boll, S., Bartholomaeus, M., Peter, U., Lupke, U., and Gamer, M. (2016). Attentional mechanisms of social perception are biased in social phobia. *J. Anxiety Disord.* 40, 83–93. doi: 10.1016/j.janxdis.2016.04.004

Calvo, M. G., and Nummenmaa, L. (2008). Detection of emotional faces: salient physical features guide effective visual search. *J. Exp. Psychol. Gen.* 137, 471–494. doi: 10.1037/a0012771

Cañadas, E., and Lupiáñez, J. (2012). Spatial interference between gaze direction and gaze location: a study on the eye contact effect. *Q. J. Exp. Psychol.* 65, 1586–1598. doi: 10.1080/17470218.2012.659190

Carretero-Dios, H., Proyer, R. T., Ruch, W., and Rubio, V. J. (2010a). The Spanish version of the GELOPH <15> : properties of a questionnaire for the assessment of the fear of being laughed at. *Int. J. Clin. Health Psychol.* 10, 345–357.

Carretero-Dios, H., Ruch, W., Agudelo, D., Platt, T., and Proyer, R. T. (2010b). Fear of being laughed at and social anxiety: a preliminary psychometric study. *Psychol. Test Assess. Model.* 52, 108–124. doi: 10.5167/uzh-33204

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutor. Quant. Methods Psychol.* 1, 42–45. doi: 10.20982/tqmp.01.1.p042

de Jong, P. J., and Martens, S. (2007). Detection of emotional expressions in rapidly changing facial displays in high and low socially anxious women. *Behav. Res. Ther.* 45, 1285–1294. doi: 10.1016/j.brat.2006.10.003

Doherty-Sneddon, G., and Phelps, F. G. (2005). Gaze aversion: a response to cognitive or social difficulty? *Mem. Cogn.* 33, 727–733. doi: 10.3758/BF03195338

Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York, NY: Times Books.

Ekman, P., Davidson, R. J., and Friesen, W. V. (1990). The duchenne smile: emotional expression and brain physiology II. *J. Pers. Soc. Psychol.* 58, 342–353. doi: 10.1037/0022-3514.58.2.342

Eysenck, M. W. (1992). *Anxiety: The Cognitive Perspective*. Hove: Psychology Press.

Gamer, M., Hecht, H., Seipp, N., and Hiller, W. (2011). Who is looking at me? The cone of gaze widens in social phobia. *Cogn. Emot.* 25, 756–764. doi: 10.1080/02699931.2010.503117

Ganel, T., Goshen-Gottstein, Y., and Goodale, M. A. (2005). Interactions between the processing of gaze direction and facial expression. *Vis. Res.* 45, 1191–1200. doi: 10.1016/j.visres.2004.06.025

Goeleven, E., De Raedt, R., Leyman, L., and Verschuere, B. (2008). The Karolinska directed emotional faces: a validation study. *Cogn. Emot.* 22, 1094–1118. doi: 10.1080/02699930701626582

Hofmann, J., Platt, T., Ruch, W., and Proyer, R. (2015). Individual differences in gelotophobia predict responses to joy and contempt. *Sage Open* 5, 1–12. doi: 10.5167/uzh-110542

Hudson, M., and Jellema, T. (2011). Resolving ambiguous behavioural intentions by means of involuntary prioritisation of gaze processing. *Emotion* 11, 681–686. doi: 10.1037/a0023264

Hudson, M., Liu, C. H., and Jellema, T. (2009). Anticipating intentional actions: the effect of gaze direction on the judgment of head rotation. *Cognition* 112, 423–434. doi: 10.1016/j.cognition.2009.06.011

Johnston, L., Miles, L., and Macrae, C. N. (2010). Why are you smiling at me? Social functions of enjoyment and non-enjoyment smiles. *Br. J. Soc. Psychol.* 49, 107–127. doi: 10.1348/014466609X412476

Jones, S. (2015). The mediating effects of facial expression on spatial interference between gaze direction and gaze location. *J. Gen. Psychol.* 142, 106–117. doi: 10.1080/00221309.2015.1009822

Lang, P. J. (1980). "Behavioural treatment and bio-behavioural assessment: computer applications," in *Technology in Mental Health Care Delivery Systems*, eds J. B. Sidowski, J. H. Johnson, and T. A. Williams (Norwood, NJ: Ablex), 119–l37.

Lundqvist, D., Flykt, A., and Ohman, A. (1998). *The Karolinska Directed Emotional Faces - KDEF, CD ROM*. Stockholm: Karolinska Institute.

Macrae, C. N., Hood, B. M., Milne, A. B., Rowe, A. C., and Mason, M. F. (2002). Are you looking at me? Eye gaze and person perception. *Psychol. Sci.* 13, 460–464. doi: 10.1111/1467-9280.00481

Mattick, R. P., and Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behav. Res. Ther.* 36, 455–470. doi: 10.1016/S0005-7967(97)10031-6

Milders, M., Hietanen, J. K., Leppänen, J. M., and Braun, M. (2011). Detection of emotional faces is modulated by the direction of eye gaze. *Emotion* 11, 1456–1461. doi: 10.1037/a0022901

Niedenthal, P. M., Mermillod, M., Maringer, M., and Hess, U. (2010). The simulation of smiles (SIMS) model: embodied simulation and the meaning of facial expression. *Behav. Brain Sci.* 33, 417–433. doi: 10.1017/S0140525X10000865

Olivares, J., García-López, L. J., and Hidalgo, M. D. (2001). The social phobia scale and the social interaction anxiety scale: factor structure and reliability in a Spanish-speaking population. *J. Psychoeduc. Assess.* 19, 69–80. doi: 10.1177/073428290101900105

Papousek, I., Aydin, N., Lackner, H. K., Weiss, E. M., Buhner, M., Schulter, G., et al. (2014). Laughter as a social rejection cue: gelotophobia and transient cardiac responses to other persons' laughter and insult. *Psychophysiology* 51, 1112–1121. doi: 10.1111/psyp.12259

Papousek, I., Ruch, W., Freudenthaler, H. H., Kogler, E., Lang, B., and Schulter, G. (2009). Gelotophobia, emotion-related skills and responses to the affective states of others. *Pers. Individ. Dif.* 47, 58–63. doi: 10.1016/j.paid.2009.01.047

Phillips, W., Baron-Cohen, S., and Rutter, M. (1992). The role of eye contact in goal detection: evidence from normal infants and children with autism or mental handicap. *Dev. Psychopathol.* 4, 375–383. doi: 10.1017/S0954579400000845

Platt, T., Hofmann, J., Ruch, W., and Proyer, R. T. (2013). Duchenne display responses towards sixteen enjoyable emotions: individual differences between no and fear of being laughed at. *Mot. Emot.* 37, 776–786. doi: 10.1007/s11031-013-9342-9

Platt, T., Ruch, W., Hofmann, J., and Proyer, R. T. (2012). Extreme fear of being laughed at: components of gelotophobia. *Isr. J. Humor Res.* 1, 86–106. doi: 10.5167/uzh-51808

Prinzmetal, W., Leonhardt, J., and Garrett, R. (2008). Does gaze direction affect accuracy? *Vis. Cogn.* 16, 567–584. doi: 10.1080/13506280801981341

Proyer, R. T., Ruch, W., Ali, N. S., Al-Olimat, H. S., Amemiya, T., Andualem, T., et al. (2009). Breaking ground in cross-cultural research on the fear of being laughed at (gelotophobia): a multi-national study involving 73 countries. *Humor Int. J. Humor Res.* 22, 253–279. doi: 10.1515/HUMR.2009.012

Ruch, W. (2009). Fearing humor? Gelotophobia: the fear of being laughed at. Introduction and overview. *Humor* 22, 1–25. doi: 10.1515/HUMR.2009.001

Ruch, W., Hofmann, J., and Platt, T. (2015). Individual differences in gelotophobia and responses to laughter-eliciting emotions. *Pers. Individ. Dif.* 72, 117–121. doi: 10.1016/j.paid.2014.08.034

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. T. (2014a). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor Int. J. Humor Res.* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., Platt, T., Hofmann, J., Niewiadomski, R., Urbain, J., Mancini, M., et al. (2014b). Gelotophobia and the challenges of implementing laughter into virtual agents interactions. *Front. Hum. Neurosci.* 8:928. doi: 10.3389/fnhum.2014.00928

Ruch, W., and Proyer, R. T. (2008). Who is gelotophobic? Assessment criteria for the fear of being laughed at. *Swiss J. Psychol.* 67, 19–27. doi: 10.1024/1421-0185.67.1.19

Ruch, W., and Proyer, R. T. (2009). Who fears being laughed at? The location of gelotophoia in the Eysenckian PEN-model. *Pers. Individ. Dif.* 46, 627–630. doi: 10.1016/j.paid.2009.01.004

Ruch, W., Proyer, R. T., and Popa, D. E. (2008). "The fear of being laughed at (gelotophobia) and personality," in *Anuarul Institutului de Istorie: Series Humanistica*, Vol. 6, ed. G. Baritiu (Cluj-Napoca: Romania), 53–68.

Ruch, W., Proyer, R. T., and Ventis, L. (2010). The relationship of teasing in childhood to the expression of gelotophobia in adults. *Psychol. Test Assess. Model.* 52, 77–93. doi: 10.5167/uzh-33203

Sander, D., Grandjean, D., Kaiser, S., Wehrle, T., and Scherer, K. R. (2007). Interaction effects of perceived gaze direction and dynamic facial expression: evidence for appraisal theories of emotion. *Eur. J. Cogn. Psychol.* 19, 470–480. doi: 10.1080/09541440600757426

Senju, A., and Hasegawa, T. (2006). Do the upright eyes have it? *Psychon. Bull. Rev.* 13, 223–228. doi: 10.3758/BF03193834

Titze, M. (2009). Gelotophobia: the fear of being laughed at. *Humor* 22, 27–48. doi: 10.1515/HUMR.2009.002

Weiss, E. M., Schulter, G., Freudenthaler, H. H., Hofer, E., Pichler, N., and Papousek, I. (2012). Potential markers of aggressive behavior: the fear of other persons' laughter and its overlaps with mental disorders. *PLOS ONE* 7:e38088. doi: 10.1371/journal.pone.0038088

Wu, C. L., Zhong, S., Chan, Y. C., Chen, H. C., Gong, G., He, Y., et al. (2016). White-matter structural connectivity underlying human laughter-related traits processing. *Front. Psychol.* 7:1637. doi: 10.3389/fpsyg.2016.01637

Zebrowitz, L. A., Kikuchi, M., and Fellous, J. M. (2010). Facial resemblance to emotions: group differences, impression effects, and race stereotypes. *J. Pers. Soc. Psychol.* 98, 175–189. doi: 10.1037/a0017990

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Torres-Marín, Carretero-Dios, Acosta and Lupiáñez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 May 2017
doi: 10.3389/fpsyg.2017.00714

Assessing Dispositions Toward Ridicule and Laughter in the Workplace: Adapting and Validating the PhoPhiKat-9 Questionnaire

Jennifer Hofmann[1,2]*, *Willibald Ruch*[1,2], *René T. Proyer*[3], *Tracey Platt*[4] and *Fabian Gander*[1,2]

[1] Personality and Assessment, Department of Psychology, University of Zurich, Zurich, Switzerland, [2] Swiss National Centre of Competence in Research Lives–Overcoming Vulnerability: Life Course Perspectives, Lausanne, Switzerland, [3] Department of Psychology, Martin-Luther University Halle-Wittenberg, Halle, Germany, [4] Institute of Psychology, University of Wolverhampton, Wolverhampton, UK

The current paper addresses the measurement of three dispositions toward ridicule and laughter; i.e., gelotophobia (the fear of being laughed at), gelotophilia (the joy of being laughed at), and katagelasticism (the joy of laughing at others). These traits explain inter-individual differences in responses to humor, laughter, and social situations related to humorous encounters. First, an ultra-short form of the PhoPhiKat-45 (Ruch and Proyer, 2009) was adapted in two independent samples (Construction Sample $N = 157$; Replication Sample $N = 1{,}774$). Second, we tested the validity of the PhoPhiKat-9 in two further independent samples. Results showed that the psychometric properties of the ultra-short form were acceptable and the proposed factor structure could be replicated. In *Validation Sample 1* ($N = 246$), we investigated the relation of the three traits to responses in a ridicule and teasing scenario questionnaire. The results replicated findings from earlier studies by showing that gelotophobes assigned the same emotions to friendly teasing and malicious ridicule (predominantly low joy, high fear, and shame). Gelotophilia was mainly predicted by relating joy to both, teasing and ridicule scenarios, while katagelasticism was predicted by assigning joy and contempt to ridicule scenarios. In *Validation Sample 2* ($N = 1{,}248$), we investigated whether the fear of being laughed at is a vulnerability at the workplace: If friendly teasing and laughter of co-workers, superiors, or customers are misperceived as being malicious, individuals may feel less satisfied and more stressed. The results from a representative sample of Swiss employees showed that individuals with a fear of being laughed at are generally less satisfied with life and work and experience more work stress. Moreover, gelotophilia went along with positive evaluations of one's life and work, while katagelasticism was negatively related to work satisfaction and positively related to work stress. In order to establish good work practices and build procedures against workplace bullying, one needs to consider that individual differences impact on a person's perception of being bullied and assessing the three dispositions may give important insights into team processes.

Keywords: assessment, bullying, gelotophobia, humor, laughter, work satisfaction, work place

OPEN ACCESS

Edited by:
Thomas L Webb,
University of Sheffield, UK

Reviewed by:
Karl-Heinz Renner,
Bundeswehr University Munich, Germany
Ilona Papousek,
University of Graz, Austria
Russell Spears,
University of Groningen, Netherlands

***Correspondence:**
Jennifer Hofmann
j.hofmann@psychologie.uzh.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *17 January 2017*
Accepted: *21 April 2017*
Published: *12 May 2017*

Citation:
Hofmann J, Ruch W, Proyer RT, Platt T and Gander F (2017) Assessing Dispositions Toward Ridicule and Laughter in the Workplace: Adapting and Validating the PhoPhiKat-9 Questionnaire. Front. Psychol. 8:714. doi: 10.3389/fpsyg.2017.00714

INTRODUCTION

Although humor and laughter are commonly viewed as positively valued, empirical evidence suggests individual differences in the perception of laughter and laughter-related events (see Ruch et al., 2014). Three dispositions toward laughter and ridicule (Ruch and Proyer, 2009) have been coined to define specific inter-individual tendencies to either (a) fearing being laughed at (Ruch and Proyer, 2008a,b; gelotophobia), (b) enjoying being laughed at (gelotophilia; Ruch and Proyer, 2009), or (c) enjoying to laugh at others (katagelasticism; Ruch and Proyer, 2009).

Individuals with a fear of being laughed at display biases in their perception of humor and laughter, as well as in responses to those phenomena (see Ruch et al., 2014). They see humor and laughter as negative, aversive, and directed toward them in a malicious way (see e.g., Ruch and Proyer, 2008a; Ruch et al., 2014). For example, they respond to both, friendly teasing and malicious ridicule with higher felt shame, fear, and low joy in predefined scenarios of ridicule and teasing. They do not emotionally distinguish between the different contexts (Platt, 2008). Gelotophobes screen social interaction partners for signs of derision, and often show paranoid tendencies toward being laughed at. They further display disproportionate negative responses toward anticipated ridicule. Moreover, they respond with controlling themselves and their environment, withdrawing, or internalizing when confronted with (anticipated) ridicule (Papousek et al., 2009; Platt et al., 2012; Ruch et al., 2014). Moreover, gelotophobes experience marked heart rate deceleration when hearing laughter (indicating a "freezing-like" response; Papousek et al., 2014).

Thus, gelotophobes respond to the pro-social bonding and group building aspects of humor and laughter with aversion and misinterpretation, which can have detrimental effects on social interaction. Whereas, withdrawing from fear-evoking situations may be manageable in their personal lives, they will encounter problems in the work place where they presumably cannot avoid engaging in social interactions. It is speculated that gelotophobes will find humorous interactions with (unfamiliar) customers and staff, team members, and supervisors difficult (Ruch et al., 2014): they are likely to misinterpret friendly banter and humor in the work place more often as negative, will screen the environment for laughter and will attribute this laughter as being laughed at. In line with this, Ruch and Proyer (2008b) already predicted that higher degrees of gelotophobia should be found in victims of bullying (e.g., at the workplace see Ruch et al., 2014) and related to phenomena like aggressiveness[1] or coherence within social groups (see Samson and Meyer, 2010). Additionally, Platt et al. (2009) confirmed that gelotophobia correlated positively with reports of having been a victim of bullying. While this may be distressing for the individual, it has also implications on a broader level too. At the level of organizations, such behaviors could seriously impact on employees' well-being, be a potential financial burden when going along with increased social welfare payments, have an impact on over-stretching health service resources, and potentially add costs to spurious employment ligations.

Nevertheless, these predictions have so far not been substantiated in a working context; i.e., in representative samples of the workforce of a given country. This is relevant, as the perceived bullying and discrimination may be based on "false alarms" due to gelotophobia, while there is actually no objective evidence for it (Ruch and Proyer, 2008a). Such misperceptions may reflect in lower work and life satisfaction (see Proyer et al., 2012b), as well as higher work stress. For the co-workers and supervisors, claims of bullying assaults need to be taken seriously, but they should also take into account the individual differences in the perception of humor and laughter, if other evidence does not corroborate the claims.

While gelotophobes dread the laughter of others, *gelotophiles* actively seek it: They readily tell others of their mishaps and embarrassing situations because they enjoy the laughter of others that these stories elicit (Ruch and Proyer, 2009). They explicitly seek potentially embarrassing situations for the joy of recalling this to an audience. As expected, gelotophobia is negatively correlated to gelotophilia (Ruch and Proyer, 2009). In a work context, it is assumed that gelotophiles will be frequent elicitors of humor and laughter (particularly when it relates to them) and they will perceive friendly banter as joyful. They will be viewed as the "good cheer" of the group. Thus, we hypothesize that gelotophilia will positively related to work and life satisfaction (in line with former findings, see Ruch et al., 2014 for an overview) and negatively to work stress, due to their ability to laugh at their mishaps and ability to initiate humor and laughter.

The third disposition relates to those who experience joy when laughing at others, *katagelasticism* (Ruch and Proyer, 2009). Katagelasticists screen their peers carefully to find instances or causes of amusement. These triggers are then used for making others laugh. They actively search for situations where they can laugh at others and do not feel guilty for doing so. As the saying goes "an eye for an eye, a tooth for a tooth," their aim is for the targets of their mockery to take revenge and joke or prank back by trying to out perform the initial joke (Ruch and Proyer, 2009). While katagelasticism is positively correlated to gelotophilia, typically no relation to gelotophobia is found. Therefore, some gelotophobes might as well-enjoy laughing at others, whereas others will not. In work place contexts, katagelasticists are predicted to be seen as the "bullies" as they enjoy laughing at others and be the ones who encounter problems in the work place, as they behave socially undesirable by laughing at others frequently.

The three traits can be reliably assessed with a self-report measure, the PhoPhiKat-45 (Ruch and Proyer, 2009). Many studies have shown the reliability and validity (cf. Ruch et al., 2014). The scale allows separating the gelotophobia spectrum (with means ranging from 1 to 4) into groups of no fear (<2.5 on the gelotophobia scale), a slight fear (>2.5), a marked fear (>3.0), and extreme fear of being laughed at (>3.5; see Ruch and Proyer, 2008a). While a 30-item short form (Ruch and Proyer, 2009) exists, an ultra-short version is required for research and application. In research contexts, the short form can be utilized for screening purposes and the use in

[1] Weiss et al. (2012) could show that gelotophobes showed deficits in handling their emotions, more self-reported aggressive behavior, and anger proneness.

large-scale studies. In the latter, the number of items for the assessment of constructs is often limited and the comparatively lower reliabilities can be compensated by larger sample sizes. In the applied context, the short form can serve as an economic instrument for the screening of the three dispositions toward laughter in large groups, for work place counseling, and the investigation of team processes (yet, the ultra-short form always needs to be complemented by the long form for individual counseling).

The aims of the current study were two-fold. First, we aimed to develop an ultra-short form of the standard self-report questionnaire on the three dispositions toward ridicule and laughter, the PhoPhiKat-45 (Ruch and Proyer, 2009). This newly developed questionnaire, labeled PhoPhiKat-9 was tested for its psychometric properties[2]. The development of the ultra-short form was motivated by the necessity to include a brief measure of the PhoPhiKat-9 in the project conducted by the Swiss National Centre of Competence in Research (LIVES—Overcoming vulnerability: Life course perspectives), which examines the effects of the post-industrial economy and society on the development of vulnerability (using a longitudinal and comparative approach in a representative sample of the Swiss work force). Second, we validated the short form by relating it to the performance in a ridicule and teasing scenario test. As it was shown previously that gelotophobes do not distinguish well between teasing and ridicule. We aimed to replicate this well-established finding in order to show the validity of the PhoPhiKat-9. Moreover, we established first relations of the three dispositions to relevant work place related variables (global life satisfaction, work satisfaction, work stress) in a large-scale representative sample of Swiss employees, to see whether the dispositions could help explaining vulnerabilities in the work place.

METHOD

Participants

Construction sample

The sample consisted of 157 German-speaking adults (34 males, 123 females). The age ranged between 18 and 59 years old (M = 28.l2, SD = 9.34).

Replication Sample 1

The sample consisted of 1774 German-speaking adults (443 males, 1331 females). The age ranged between 18 and 79 years old (M = 38.44, SD = 12.41).

[2]We followed the guidelines recommended by Smith et al. (2000). One requirement is that the original instrument has shown enough evidence of reliability and validity. For the PhoPhiKat-45, a variety of validation studies have shown its good psychometric characteristics and validity (see Ruch et al., 2014 for a review). A further requirement suggests that the development of the short form and the analysis of its psychometric properties should be conducted in two independent samples. We included data of two independent samples for the construction and replication, as well as two samples for the validation.

Validation Sample 1

The sample consisted of 246 German-speaking adults (204 females, 42 males). The age ranged between 19 and 72 years old (M = 42.54, SD = 12.66).

Validation Sample 2

The sample consisted of 1248 German-speaking adults (627 males, 627 females) from the NCCR- LIVES (data from the first wave of data collection in 2012). The age ranged between 26 and 56 years old (M = 42.73, SD = 8.73). The sample is representative for the Swiss working population.

Instruments

The *PhoPhiKat-45* (Ruch and Proyer, 2009) is a 45-item questionnaire for the assessment of gelotophobia (a sample item is "When they laugh in my presence I get suspicious"), gelotophilia ("When I am with other people, I enjoy making jokes at my own expense to make the others laugh"), and katagelasticism ("I enjoy exposing others and I am happy when they get laughed at"). Answers are given on a four-point answer format (1 = strongly disagree to 4 = strongly agree). Ruch and Proyer (2009) reported high reliability coefficients (all alphas $\geq$0.84) and high retest-reliabilities $\geq$0.77 and $\geq$0.73 for a 3 and 6-month time period, respectively.

The *Ridicule Teasing Scenario Questionnaire Revised* (RTSqr; Platt, 2008) contains nine scenarios that assess emotions toward predetermined ridicule and teasing social scenarios. Four teasing, four ridicule, and one ambiguous scenarios are presented with short stories where participants rate to which extent they would experience eight emotions (joy, sadness, anger, disgust, surprise, shame, and fear plus contempt in the revised version) on a nine point Likert scale (from 0 = lowest to 8 = highest experience of emotions). Eight total scores are computed for both ridicule and teasing by averaging across the four scenarios.

The *Satisfaction with Life Scale* (SWLS; Diener et al., 1985) assesses the participants' life satisfaction. Answers are given on a seven-point scale (1 = strongly disagree to 7 = strongly agree). A sample item is "The conditions of my life are excellent." In the current study (Validation Sample 2), the Cronbach's alpha was high (α = 0.89).

Global work satisfaction was assessed by one item ("In general, how satisfied are you with your work?") on a four-point scale (1 = not satisfied at all to 4 = very satisfied).

The *General Work Stress Scale* (GWSS; De Bruin, 2006) is a nine item questionnaire assessing individually perceived demands of the workplace (e.g., "Do you become so stressed at work that you forget to do important tasks"). A five-point answer format is used (1 = never to 5 = always) measuring work stress as a one-dimensional construct. Cronbach's alpha in the current study (Validation Sample 2) was .87 and thus comparable to earlier findings (see De Bruin, 2006).

Procedure

Participant Recruitment

Participants were recruited in four independent surveys, three online surveys, and one mixed-method survey. They were not paid, but were offered an individual feedback on their personality

scores (on demand) or could receive a gift voucher/make a donation in Validation Sample 2. All participants stayed anonymous at all times and they were free to withdraw from the study at any time. The studies fulfilled the ethical standards for research of the APA and approval from local ethic committees was granted.

Construction Sample

The study was announced on the website of the University of Zurich and in a free local newspaper distributed in the public transport of the Zurich area. Participants received a link to the online survey and filled in the questionnaires.

Replication Sample

Participants completed the survey on a website for research purposes hosted by the lab of the authors (http://www.charakterstaerken.org). The website was promoted by different means, such as press coverage (e.g., newspapers articles) and by contacting specific occupational groups, in order to ascertain heterogeneity of the sample.

Validation Sample 1

Individuals from the Replication Sample were contacted via email approximately 10-month after their initial participation and invited to take part in a new online survey. In this online survey, the participants completed the PhoPhiKat-9 short-form (plus one item) and the RTSqr.

Validation Sample 2

The data was collected within NCCR- LIVES (Swiss National Centre of Competence in Research LIVES—Overcoming vulnerability: Life course perspectives; data from the first wave of data collection in 2012). A representative sample of participants was drawn from the Swiss National Register of Inhabitants. In a mixed-method design, participants completed a first part of a questionnaire by phone or online (socio-demographic data and employment information), and the second part of the questionnaire online or paper-pencil (including the PhoPhiKat-9, SWLS, GWSS).

Ethics Statement

This study complies with the ethical standards of the Swiss Society for Psychology. Also, the study was approved by the Ethics Committee of the Institute of Psychology, University of Zurich. All participants gave consent to participate and were free to withdraw from the study at any time, and their anonymity was ensured. As incentive, they could receive a personalized feedback in the *Construction Sample*, the *Replication Sample*, and the *Validation Sample 1*. Additionally, for *Validation Sample 2*, the institute that conducted the data collection obtained informed consent, kept the personal information, and researchers received a dataset without any personal information, in which participants were assigned numerical codes. Participants were compensated for their participation with a gift for a value of 20 Swiss francs.

Construction of the Short Form PhoPhiKat-9

The items for the PhoPhiKat-9 were selected from the PhoPhiKat-45 (Ruch and Proyer, 2009) in the Construction Sample. For the gelotophobia scale, items were selected to represent three facets found by Platt et al. (2012); i.e., (a) coping with derision (i.e., "I avoid showing myself in public because I fear that people could become aware of my insecurity and could make fun of me"); (b) disproportionate negative responses to being laughed at ("It takes me very long to recover from having been laughed at"); and (c) paranoid sensitivity to anticipated ridicule ("When strangers laugh in my presence I often relate it to me personally"). The selected items had the highest factor loading on each facet respectively (Construction Sample; cf. Platt et al., 2012).

For selecting the items for gelotophilia and katagelasticism, a principal component analysis was computed with the 45 items of the PhoPhiKat-45. Three component were extracted and rotated according to the Oblimin criterion (delta = 0). The components represented the three traits and were labeled accordingly. The rationale for the selection of the items was: (a) highest factor loading on the intended factor (and low secondary loadings; the difference between secondary loadings should be $\geq$0.30), (b) high corrected item-total correlations, and (c) the content should not overlap too strongly with the items that were already selected. **Table 1** shows descriptive statistics for the nine item short form[3].

Next, we examined the factor structure of the nine item short form in a principal component analysis in the Replication Sample. Three components were extracted (Eigenvalues were 2.50, 2.09 and 0.92, respectively; explained variance = 61.19%) and rotated to the Oblimin criterion (delta = 0). Component 1 contained all gelotophobia items plus one item (with a high negative loading) that originally belonged to the gelotophilia scale (with loadings ranging from −0.59 to .79; see **Table 1**), component 2 constituted of the katagelasticism items (loadings ranging from 0.68 to 0.81), and component 3 of the remaining two gelotophilia-items (loadings were 0.76 and 0.88; see **Table 1**). Thus, eight items had their highest loadings on their target component, as theoretically expected, and had no high loadings on the other two components. However, one gelotophilia-item ("There is no difference for me whether people laugh at me or laugh with me") had its highest loading on the gelotophobia factor.

Investigating the nature of the short form, we computed a confirmatory factor analysis (CFA) for three different models (Replication Sample, N = 1,774). To evaluate the model fit, RMSEA and SRMR values lower than 0.10 were assumed to indicate acceptable fit (e.g., Browne and Cudeck, 1993). According to Bollen and Long (1993), a RMSEA of 0.09 SRMR of 0.06 would be around the limit of being a reasonable error. We further followed the recommendations of Schermelleh-Engel et al. (2003), additionally reporting CFI and TLI. For model 1, we assumed correlated factors and loadings of each item on one factor alone, without secondary loadings on another

[3]As different samples were utilized, internal consistencies are reported for all samples separately: In the replication sample, the Cronbach's alpha of the PhoPhiKat-9 were 0.69 for gelotophobia, 0.57 for gelotophilia, and 0.64 for katagelasticism. In the validation sample 1, the Cronbach's alpha of the PhoPhiKat-9 were 0.70 for gelotophobia, 0.69 for gelotophilia, and 0.38 for katagelasticism. In the validation sample 2, the Cronbach's alpha of the PhoPhiKat-9 were 0.64 for gelotophobia, 0.54 for gelotophilia, and 0.65 for katagelasticism.

TABLE 1 | Descriptive statistics and factor loadings for the nine items of the PhoPhiKat-9 short form in the replication sample.

Item description	Scale	*M*	*SD*	CITC	Loadings		
					Gelotophobia	Gelotophilia	Katagelasticism
"Public attention" (4)	Pho	2.22 (2.23)	0.92 (0.96)	0.53 (0.66)	**0.71 (0.63)**	−0.16 (−0.22)	0.12 (0.01)
"No difference" (8)	Phi	2.28 (2.11)	0.90 (0.95)	0.28 (0.48)	**0.59 (−0.20)**	0.18 (0.49)	0.11 (−0.05)
"Laughing at others (3)	Kat	1.36 (1.35)	0.63 (0.63)	0.52 (0.37)	−0.05 (−0.06)	−0.02 (−0.07)	**0.81 (0.49)**
"Self-focus" (7)	Pho	1.93 (2.11)	0.84 (0.90)	0.49 (0.62)	**0.79 (0.62)**	0.25 (−0.14)	0.12 (−0.14)
"Fun maker" (14)	Phi	2.10 (2.23)	0.91 (0.95)	0.40 (0.56)	−0.05 (0.07)	**0.88 (0.63)**	−0.02 (0.16)
"Causing fights" (6)	Kat	1.60 (1.43)	0.74 (0.66)	0.42 (0.44)	−0.11 (0.12)	0.06 (0.09)	**0.68 (0.50)**
"Long recovery" (25)	Pho	2.21 (2.21)	0.90 (1.11)	0.52 (0.40)	**0.79 (0.37)**	−0.03 (−0.23)	−0.12 (0.17)
"No shame" (26)	Phi	2.21 (2.07)	0.89 (0.93)	0.44 (0.70)	0.12 (0.02)	**0.76 (0.74)**	0.04 (0.07)
"Part of life" (27)	Kat	1.69 (1.77)	0.79 (0.89)	0.45 (0.56)	0.15 (−0.05)	−0.04 (−0.05)	**0.80 (0.68)**
Cronbach's α					0.70 (0.88)	0.56 (0.87)	0.66 (0.84)
M					2.12 (1.97)	2.20 (2.43)	1.55 (1.99)
SD					0.70 (0.54)	0.66 (0.55)	0.55 (0.46)

N = 1774. M, Mean; SD, Standard Deviation. Item descriptions refer to paraphrases. CITC, corrected item total correlation; Pho, gelotophobia. Phi, gelotophilia; Kat, katagelasticism. First column numbers in brackets are corresponding to position of the item on the PhoPhiKat-45. Bold values indicate high loadings.

factor. The null hypothesis of perfect fit for this model was rejected [$\chi^2_{(24)} = 496.54$, $p < 0.001$; CFI = 0.85, TLI = 0.77, RMSEA = 0.105 (0.097–0.114), SRMR = 0.08]. For model 2, the assumptions were the same as for model 1 except for the first gelotophilia-item, which was allowed to have a second loading on the gelotophobia-factor. This model yielded better results [$\chi^2_{(23)} = 301.99$, $p < 0.001$; CFI = 0.91, TLI = 0.87, RMSEA = 0.083 (0.075–0.091), SRMR = 0.06] with acceptable (but not high) model fit indices. In model 3, the gelotophilia-item was allowed to load only on the gelotophobia factor, with the loading on the gelotophilia factor restricted to zero, while the other model specifications remained the same. The model fit was acceptable [$\chi^2_{(24)} = 357.24$, $p < 0.001$; CFI = 0.89, TLI = 0.86, RMSEA = 0.088 (0.080–0.097), SRMR = 0.07]. Thus, model 2 and 3 yielded acceptable solutions, with one gelotophilia item also loading on the gelotophobia factor. As this item worked well in the earlier studies (see Ruch and Proyer, 2009) we therefore did not consider this a serious deviation.

TABLE 2 | Descriptive statistics of the PhoPhiKat-9 and PhoPhiKat-45 in the validation sample 1.

	M	*SD*	Alpha	CITC range	$t_{(224)}$
PhoPhiKat-9					
Gelotophobia	2.25	0.73	0.70	0.41–0.52	8.46***
Gelotophilia	2.23	0.71	0.69	0.50–0.53	−8.24***
Katagelasticism	1.52	0.46	0.38	0.15–0.28	−16.14***
PhoPhiKat-45					
Gelotophobia	1.99	0.56	0.89	0.25–0.69	
Gelotophilia	2.47	0.55	0.89	0.45–0.68	
Katagelasticism	1.97	0.43	0.84	0.27–0.63	

*N = 201–246. M, Mean; SD, Standard Deviation; Alpha, Cronbach's alpha; CITC, corrected item-total correlation; t-tests (df = 224) for mean level differences of the short and long form scales of gelotophobia, gelotophilia, and katagelasticism. ***p < 0.001.*

VALIDATION RESULTS

Characteristics of the PhoPhiKat–9 in the Validation Sample 1

First, the descriptive statistics of the PhoPhiKat-9 items in the Validation Sample 1 are reported in **Table 2**. Means, standard deviations, Cronbach's alpha and the corrected item-total correlations (CITCs) can be seen in **Table 2**.

The corrected item-total correlations (CITC) ranged between $r = 0.15$ and $r = 0.53$ for the short form. For the katagelasticism scale, the CITCs were remarkably lower and all below $r = 0.30$ (see **Table 2**). The Cronbach's alpha coefficients of gelotophilia and gelotophobia were acceptable (0.69 and 0.70; see **Table 2**), while the alpha of the katagelasticism scale was low (0.38). As expected, the Cronbach's alpha coefficients were smaller in the short form than in the PhoPhiKat-45 (see **Table 2**) due to the smaller number of items.

Second, we investigated mean level differences between the gelotophobia, gelotophilia, and katagelasticism scale, assessed by the short and the long form with *t*-tests for dependent samples. As shown in **Table 2**, the gelotophobia and gelotophilia means were higher in the short form, compared to the long form. The mean score of katagelasticism was lower in the short form assessment than the long form. Importantly, the results indicated that the cut-off for gelotophobia (>2.5 in the gelotophobia scale of the PhoPhiKat-45) could not be applied in the short form, as this would lead to an over-estimation of gelotophobes due to the increased mean in the short form. Therefore, we estimated the cut-off score equivalents for the short form by means of plotting the gelotophobia scores of the short and long form in a bivariate plot. The plot indicated that the equivalent of the 2.5 cut-off in the long form was reached by the approximate cut-off score of 2.67 in the short form. In both samples, the gelotophobia scores reached a cumulative percentage of 80.9%

at the values of 2.47 (long form) and 2.67 (short form). With this cut-off equivalent, that classification of gelotophobes was only minimally different between the PhoPhiKat-45 and the PhoPhiKat-9. Splitting the group according to the criterion of the long form (cut-off of 2.5) resulted in 40 individuals being classified as gelotophobes. Splitting the group according to the cut-off equivalent in the ultra-short form (>2.67) resulted in 43 individuals being classified as gelotophobes[4]. Third, we investigated the correlations of the short and the long form of the PhoPhiKat. The correlations between the respective traits of the short and long form were high (0.58–0.76, $p < 0.001$). As expected (see Ruch and Proyer, 2009), both gelotophobia scales were unrelated to the katagelasticism scales (−0.03 to −0.10, n.s.) and negatively related to the gelotophilia scales (−0.36 to −0.42, $p < 0.001$). The katagelasticism scales were positively related to gelotophilia (0.34–0.38, $p < 0.001$). Previously reported correlation patterns could be replicated for both forms of the PhoPhiKat and the inter-correlations between the short and long form indicated an acceptable content overlap[4].

Predicting Responses Toward Ridicule and Teasing Scenarios

To investigate the criterion validity of the PhoPhiKat-9, we utilized the RTSqr in the Validation Sample 1. Earlier research (e.g., Platt, 2008) showed that gelotophobes did not distinguish between ridicule and teasing when having to rate the emotions toward ridicule and teasing scenarios, assigning predominantly low joy, high fear, and high shame to both kinds of scenarios. Thus, in a first step, we investigated whether individuals above the cut-off point for gelotophobia would show similar response patterns of feeling high negative emotions and low joy when confronted with ridicule and teasing scenarios. We applied the cut-off equivalent for the short form (no gelotophobia ≤2.67, $n = 203$; gelotophobia >2.67, $n = 43$ individuals) for gelotophobia and computed two repeated measures ANOVAs (for the ridicule and teasing scenarios), with gelotophobia group (no gelotophobia vs. gelotophobia) as factor, the eight emotion ratings as repeated measures, and the intensity of emotion as dependent variable. For the ridicule scenarios, results showed that both main effects for type of emotion [$F_{(7, 1393)} = 65.61$, $p < 0.001$, $\eta_p^2 = 0.248$] and gelotophobia group [$F_{(1, 199)} = 16.73$, $p < 0.001$, $\eta_p^2 = 0.078$] were significant. Furthermore, the results were qualified by an interaction between gelotophobia group and type of emotion, $F_{(7, 1393)} = 16.86$, $p < 0.001$, $\eta_p^2 = 0.078$. **Figure 1** shows the means and confidence intervals (95%) of the eight emotion ratings in the two groups (gelotophobia vs. no gelotophobia) toward ridicule and teasing scenarios.

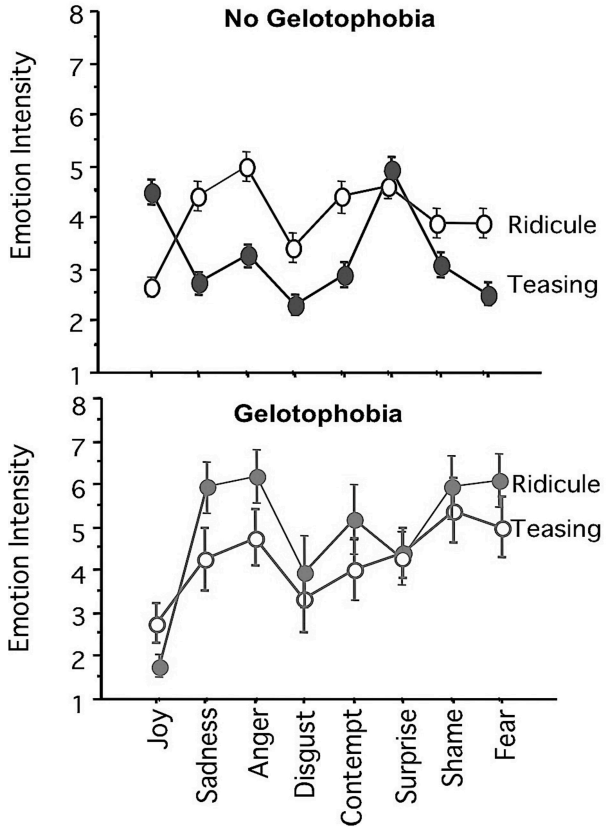


FIGURE 1 | Means and confidence intervals (95%) of the eight emotion ratings toward ridicule scenarios and teasing scenarios in gelotophobes and individuals with no fear of being laughed at (no gelotophobia).

Replicating the findings of Platt (2008), both groups of individuals assigned ridicule to negative feelings (mainly anger) and low joy. **Figure 1** shows that the gelotophobes had higher ratings of sadness, anger, disgust, contempt, shame, and fear compared to individuals without a fear of being laughed at (all $p < 0.05$, Bonferroni corrected). In line with the predictions, the level of gelotophobia predicted the disproportionate negative responses to being laughed at by eliciting more intense negative feelings toward ridicule scenarios.

Concerning the teasing scenarios, results showed that both main effects for type of emotion [$F_{(7, 1400)} = 27.67$, $p < 0.001$, $\eta_p^2 = 0.094$] and gelotophobia group [$F_{(1, 200)} = 27.40$, $p < 0.001$, $\eta_p^2 = 0.121$] were significant. Furthermore, the results were qualified by an interaction between gelotophobia group and type of emotion, $F_{(7, 1400)} = 33.61$, $p < 0.001$, $\eta_p^2 = 0.144$. As **Figure 1** indicates, gelotophobes were higher in anger, fear, disgust, contempt, shame, and lower in joy and surprise, compared to individuals with no fear of being laughed at (all $p < 0.05$,

[4]For the PhoPhiKat-45, age and gender differences were reported (see Ruch and Proyer, 2009). For replication purposes, we computed correlations between the three dispositions, age and gender in both forms. Concerning the relations to the participant's age, both katagelasticism scales were negatively related to age, as previously found; r_{short} (225) = −0.11, $p = 0.106$, r_{long} (246) = −0.14, $p = 0.031$. Also, the participant's age correlated negatively to gelotophobia [r_{short} (225) = −0.11, $p = 0.103$, r_{long} (246) = 0.18, $p = 0.005$], but was unrelated to gelotophilia [r_{short} (225) = −0.04, $p = 0.564$, r_{long} (246) = −0.01, $p = 0.878$]. Fourth, we computed a MANOVA with gender as factor and the scales of the PhoPhiKat long and short forms as dependent variables. The overall effect was significant, $F_{(6, 218)} = 2.21$, $p = 0.043$, $= 0.057$. In line with the expectations, *post-hoc* tests indicated that males scored higher on both katagelasticism scales ($p < 0.05$). Males and females did not differ in gelotophobia and gelotophilia (all n.s.). The results show that the short form and long form revealed the same patterns of relationships to the demographic variables, replicating former findings (see Ruch and Proyer, 2009).

Bonferroni corrected). Thus, gelotophobes did not evaluate the friendly teasing scenarios as such, but assigned them negative emotions (mostly shame, fear, and anger) and low joy, just as to the ridicule scenarios. This replicates former findings (see Platt, 2008) but also shows that this effect can be found for the short form of the PhoPhiKat-9 as well, validating its suitability for the assessment of gelotophobia. Furthermore, the results show the bias of gelotophobes toward social situations in which teasing occurs (i.e., banter at work, pro-social teasing among friends): Instead of seeing the joyful component, gelotophobes report that they would mainly feel anger, shame, and fear.

Next, we investigated the role of katagelasticism and gelotophilia in predicting responses to ridicule and teasing. As no cut-offs exists for these dispositions, we decided to compute four hierarchical multiple regression analyses (two for the teasing and ridicule scenarios) with gelotophilia and katagelasticism as criteria and the eight emotion ratings as predictor variables. A multiple regression model was estimated in which predictors were entered when they added to the prediction of the dependent variable substantially or removed, when they did no longer add substantially to the prediction due to the inclusion of another variable (STEPWISE-procedure). These predictors entered the analysis in a second block preceded by age and gender in a first block which entered simultaneously. First, the findings on gelotophilia are reported. In the ridicule scenarios, the regression led to a multiple correlation coefficient of $R = 0.36$, $F_{(3, 197)} = 9.62$, $p < 0.001$. Gelotophilia was solely predicted by the assigned joy to the scenarios ($\beta = 0.19$, $p < 0.001$), while neither age ($\beta = -0.003$, $p = 0.445$) nor gender ($\beta = -0.13$, $p = 0.282$) had a significant contribution. No other emotion rating entered in a further step. In the teasing scenarios, the multiple correlation was $R = 0.47$ [$F_{(3, 198)} = 18.95$, $p < 0.001$]. Again, gelotophilia was predicted by the joy rating entering the equation ($\beta = 0.36$, $p < 0.001$), while neither age ($\beta = -0.08$, $p = 0.239$) nor gender ($\beta = -0.06$, $p = 0.331$) contributed significantly. No other variable entered the equation. As expected, joy mainly predicted gelotophilia in both types of scenarios.

Concerning the prediction of katagelasticism in teasing scenarios, gender turned out to be significant predictor in the first step, $F_{(2, 198)} = 3.26$, $p = 0.037$, $R = 0.18$, $\beta = -0.20$, $p = 0.022$. Age did not predict the katagelasticism score, $\beta = -0.01$, $p = 0.162$. No further variable entered the equation, indicating that none of the emotion ratings toward teasing scenarios were good predictors of the joy of laughing at others. For the ridicule scenarios, the regression led to a multiple correlation coefficient of $R = 0.34$ [$F_{(4, 196)} = 6.25$, $p < 0.001$]. Gender (entering in the first step) had a significant contribution ($\beta = -0.18$, $p = 0.030$), but not age ($\beta = -0.01$, $p = 0.059$). Furthermore, there were unique contributions of the self-reported joy in ridicule ($\beta = 0.10$, $p < 0.001$) and contempt to the prediction of katagelasticism ($\beta = 0.04$, $p = 0.009$).

Gelotophobia, Gelotophilia, and Katagelasticism and Workplace Outcomes

Next, we investigated the relationship of the three dispositions toward ridicule and laughter to life and global work satisfaction, as well as work stress in a large and representative sample of Swiss employees (Validation Sample 2). This could give first indication of whether the three dispositions can help explaining workplace related vulnerabilities. Findings for gelotophobia are presented first. Here, the established cut-off score warrants the analysis of gelotophobes vs. non-gelotophobes. We utilized the adapted cut-offs for the PhoPhiKat-9. The means and standard deviations can be seen in **Table 3**; for individuals with (gelotophobia group; scores >2.67; $n = 115$) and without a fear of being laughed at (no gelotophobia; scores ≤2.67; $n = 1017$) separately[5].

Table 3 shows the means in life satisfaction, global work satisfaction, and work stress in individuals with or without a fear of being laughed at. Investigating group differences, we computed three ANOVAs with the gelotophobia group as the factor and life satisfaction, work satisfaction, and general work stress as dependent variables. Results indicated gelotophobes reported lower levels of life satisfaction and global work satisfaction, as well as higher perceived work stress (see **Table 3**) compared to individuals with no fear of being laughed at. Thus, in line with our hypotheses, gelotophobia was negatively related to indicators of satisfaction and went along with higher reported stress.

For the investigation of the relationship of gelotophilia and katagelasticism to life and work satisfaction and work stress, we computed hierarchical multiple regression analysis with gelotophilia and katagelasticism as predictors and life satisfaction, work satisfaction, and work stress respectively as criteria. The predictors entered the analysis simultaneously in a second block preceded by age and gender in a first block (both entering simultaneously as well). To predict life satisfaction, the regression led to a multiple correlation coefficient of $R = 0.12$, $F_{(4, 1246)} = 4.16$, $p = 0.002$. Life satisfaction was predicted by gelotophilia ($\beta = 0.07$, $p = 0.022$), and katagelasticism ($\beta = -0.10$, $p < 0.001$), while neither age ($\beta = 0.05$, $p = 0.059$) nor gender ($\beta = 0.007$, $p = 0.800$) had a significant contribution. For work satisfaction, the multiple correlation was $R = 0.06$ [$F_{(4, 1238)} = 0.96$, $p = 0.431$]. None of the predictors had a significant contribution (all $p > 0.200$). For work stress, the multiple correlation was $R = 0.14$ [$F_{(4, 1236)} = 6.03$, $p < 0.001$). Only katagelasticism predicted work stress ($\beta = 0.14$, $p < 0.001$), while neither gelotophilia ($\beta = -0.01$, $p = 0.708$), age ($\beta = 0.03$, $p = 0.370$) nor gender ($\beta = -0.01$, $p = 0.672$) contributed significantly.

TABLE 3 | Group differences for individuals with or without gelotophobia in life satisfaction, global work satisfaction, and work stress.

	No gelotophobia		Gelotophobia			
	M	*SD*	*M*	*SD*	$F_{(1, 1130)}$	η_p^2
Life satisfaction	5.40	1.04	4.56	1.26	68.04***	0.06
Work satisfaction	3.36	0.58	3.18	0.50	11.63***	0.01
Work stress	1.81	0.54	2.14	0.76	37.43***	0.03

*No Gelotophobia, Gelotophobia scores ≤2.67 on the PhoPhiKat-9. Gelotophobia, Gelotophobia scores >2.67 on the PhoPhiKat-9. ***p < 0.001.*

[5]Cut-off score equivalents for marked and extreme gelotophobia assessed with the short form are at 3.33 (marked) and 3.67 (extreme).

DISCUSSION

The aim of this study was two-fold. First, we adapted the PhoPhiKat-9 for the use in large-scale studies and as a screening tool in applied settings. Second, we established first relations to work-related outcome variables in a representative sample of the Swiss work force. In terms of construction of the PhoPhiKat-9, all three dispositions can be reliably assessed with this ultra-short form. The psychometric characteristics were satisfactory when considering that this ultra-short form should only be used in large samples. The relations to demographic variables were comparable to relations found for the standard PhoPhiKat-45. Two deviations from the original PhoPhiKat-45 occurred: First, the cut-off point for gelotophobia set at 2.5 on the original PhoPhiKat-45 could not be utilized with the ultra-short form, as the means were generally higher compared to those of the original scale. We therefore estimated cut-off score equivalents basing on the criterion for the sample that had filled in both forms (long and short form). The new cut-off was set at 2.67. Second, one item representing gelotophilia revealed high loadings on the gelotophobia component as well, which may need consideration in future studies (i.e., re-phrasing item).

We utilized two independent samples to validate the PhoPhiKat-9. In line with former studies (Platt, 2008; Platt et al., 2009), the present results replicated the misperception of teasing and ridicule by individuals with elevated scores in gelotophobia. In a work based context, gelotophobes are probably going to have problems distinguishing between the friendly smiling and banter between colleagues (see also Hofmann et al., 2015), taking it for bullying. There is a stable pattern of reporting being a victim of bullying and greater expressions in gelotophobia already starting from the age of six (self- and peer-reports; for an overview see Ruch et al., 2014). Gelotophobes are therefore more likely to feel bullied and discriminated in the workplace, leading to more perceived stress, and lower satisfaction with work and life (cf. Proyer et al., 2012b). This was substantiated by findings of the second validation, where gelotophobes described themselves as less satisfied with life and work, as well as more stressed at the work place, compared to those individuals without gelotophobia.

With respect to gelotophilia, the main finding was that higher ratings of gelotophilia went along with higher ratings of joy toward both, teasing and ridicule scenarios in the RTS-qr. Gelotophiles take humorous instances light-heartedly and will initiate them with pleasure. Surprisingly, no relations of gelotophilia to satisfaction and work stress were found, indicating that other factors might be more important in the prediction of those outcomes. Interestingly, katagelasticism was predicted by the joy and contempt assigned to ridicule scenarios. In line with the descriptions by Ruch and Proyer (2009), katagelasticists get pleasure from laughing at other and will also use this as a social corrective, or to take revenge on others (i.e., "an eye for an eye," see Ruch and Proyer, 2009). Already Tomkins (1969) stated that contempt toward another person might lead to laughter directed at this individual (see Hofmann et al., 2015): Katagelasticists might ridicule a person that is disliked or has overstepped a norm, and the ridicule goes along with laughter and humor targeted at the person (e.g., Tomkins, 1969; "the laugh becomes a vehicle of contempt," p. 367). Unexpectedly, the Cronbach's Alpha of the katagelasticism scale was lower in this sample than in the other three samples (0.38 compared to 0.64, 0.65, and 0.65 in the construction and validation samples respectively). Thus, the findings on the katagelasticism scale are best treated more cautiously in this sample, while the scale is stable in the other three samples. With respect to the second validation, negative relations of katagelasticism to life satisfaction and positive relations to work stress were found. One possible explanation might be that katagelasticists generally experience more conflicts with others (generally, as well as in the work place), as they overtly laugh at them. This potentially could lead to problems in the work place and consequently to increasing levels of stress. Alternatively, katagelasticism has been shown to positively relate to psychoticism and psychopathic traits (see Proyer et al., 2012a), as well as lower social desirability. Those higher order traits might be (partially) responsible for more conflicts that could lead lowered life satisfaction and higher work stress. Thus, future studies may investigate this hypothesized mechanism and also investigate the incremental validity of katagelasticism compared to higher order traits, such as psychoticism.

Two main limitations prevail: The factor structure of the PhoPhiKat-9 did not reveal a consistent pattern for the gelotophilia scale. The item "There is no difference for me whether people laugh at me or laugh with me" loaded higher on the gelotophobia scale than on the gelotophilia scale. It is hypothesized that this item was maybe interpreted differently to the initial meaning: If individuals fear being laughed at, it does not make a difference to them if people laugh with or at them, as both is negative. In the original sense, the item possessed a positive connotation: It does not make a difference whether people laugh at or with a gelotophile, as both is equally enjoyable. This item needs a clearer phrasing toward all laughter being good to fit on the gelotophilia factor. Moreover, the mechanisms between gelotophobia, and the lowered satisfaction and work stress need to be looked at in more detail, at best by studying phenomena longitudinally. Furthermore, future studies should aim at investigating the incremental validity of the three dispositions toward ridicule and laughter in the prediction of workplace related outcomes when controlling for broader personality traits (i.e., the "Big Five"). Moreover, future studies may opt for more balanced samples in terms of gender ratio.

APPLICATION

In light of work place behavior and career trajectories, all three dispositions relate to relevant behaviors and perceptions, such as work place bullying and perceived discrimination (e.g., Platt, 2008; Platt et al., 2009; Ruch and Proyer, 2009; Proyer and Ruch, 2010; Chen and Liu, 2012). The measurement of gelotophobia, gelotophilia, and katagelasticism in work place environments

can indicate important team processes relating to the popular topics of "good work practice" and "avoidance of incidents of work place bullying." Gelotophobia may link to unfavorable work outcomes, like feeling one is being bullied, misunderstanding any laughter and humor in teams, and maybe being more stressed and less satisfied with the work environment as a consequence. Understanding the (mis-) perception will assist in redressing the bias often placed toward the alleged victims. This is of concern not only to institutions, human resource units and those practicing workplace law but also to public and governmental bullying initiatives. Hence, intervention programs should aim at raising awareness about the role of laughter and laughing at the workplace in general, but also those with greater fear of being laughed at directly. There are no standardized programs addressing the fear of being laughed at, but learning about humor and laughter and how to deal with (perceived) ridicule may be beneficial for those with extreme expressions, i.e., formulating guidelines and offering advice for applied psychologists (see Platt et al., 2012). The ultra-short form is only utilized for screening larger samples, yet, the judgments on the three dispositions need to be consolidated by giving the PhoPhiKat-45 (or the short form PhoPhiKat-30) to individuals that potentially fear being laughed at or potentially are work place bullies. This potentially helps to improve team processes and relations among co-workers and customers.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of "Swiss Psychological Association"; and the Ethics Committee of the Department of Psychology, University of Zurich, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

Conception or design of the work: WR and JH. Data collection: JH, WR, RP, TP, FG. Data analysis and interpretation: JH, WR, RP, and FG. Drafting the article: JH. Critical revision of the article: WR, RP, FG, and TP. Final approval of the published version JH, WR, RP, FG, TP.

FUNDING

This publication benefited from the support of the Swiss National Centre of Competence in Research LIVES—Overcoming vulnerability: Life course perspectives, which is financed by the Swiss National Science Foundation (grant number: 51NF40-160590). The authors are grateful to the Swiss National Science Foundation for its financial assistance.

REFERENCES

Bollen, K. A., and Long, J. S. (1993). *Testing Structural Equation Models*. Newbury Park, CA: Sage.

Browne, M., and Cudeck, R. (1993). "Alternative ways of assessing model fit," in *Testing Structural Equation Models*, eds K. A. Bollen and J. S. Lang (Newbury Park, CA: Sage), 136–162.

Chen, G.-H., and Liu, Y. (2012). Gelotophobia and thinking styles in Sternberg's theory. *Psychol. Rep.* 110, 25–34. doi: 10.2466/04.09.20.PR0.110.1.25-34

De Bruin, G. P. (2006). The dimensionality of the general work stress scale: a hierarchical exploratory factor analysis. *South Afr. J. Ind. Psychol.* 32, 68–75. doi: 10.4102/sajip.v32i4.250

Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *J. Pers. Assess.* 49, 71–75. doi: 10.1207/s15327752jpa4901_13

Hofmann, J., Platt, T., Ruch, W., and Proyer, R. (2015). Individual differences in gelotophobia predict responses to joy and contempt. *Sage Open* 5, 1–12. doi: 10.1177/2158244015581191

Papousek, I., Aydin, N., Lackner, H. K., Weiss, E. M., Bühner, M., Schulter, G., et al. (2014). Laughter as a social rejection cue: gelotophobia and transient cardiac responses to other persons' laughter and insult. *Psychophysiology* 51, 1112–1121. doi: 10.1111/psyp.12259

Papousek, I., Ruch, W., Freudenthaler, H., Kogler, E., Lang, B., and Schulter, G. (2009). Gelotophobia, emotion-related skills and responses to the affective states of others. *Pers. Individ. Dif.* 47, 58–63. doi: 10.1016/j.paid.2009.01.047

Platt, T. (2008). Emotional responses to ridicule and teasing: should gelotophobes react differently? *Humor Int. J. Humor Res.* 21, 105–128. doi: 10.1515/HUMOR.2008.005

Platt, T., Proyer, R. T., and Ruch, W. (2009). Gelotophobia and bullying: the assessment of the fear of being laughed at and its application among bullying victims. *Psychol. Sci. Q.* 51, 135–147. doi: 10.5167/uzh-19396

Platt, T., Ruch, W., Hofmann, J., and Proyer, R. T. (2012). Extreme fear of being laughed at: components of gelotophobia. *Israeli J. Humor Res.* 1, 86–106. doi: 10.5167/uzh-51808

Proyer, R. T., Flisch, R., Tschupp, S., Platt, T., and Ruch, W. (2012a). How does psychopathy relate to humor and laughter? Dispositions towards ridicule and being laughed at, the sense of humor, and psychopathic personality traits. *Int. J. Law Psychiatry* 35, 263–268. doi: 10.1016/j.ijlp.2012.04.007

Proyer, R. T., and Ruch, W. (2010). Dispositions towards ridicule and being laughed at: current research on gelotophobia, gelotophilia, and katagelasticism. Introduction to the special issue. *Psychol. Sci. Q.* 52, 49–59.

Proyer, R. T., Ruch, W., and Chen, G. H. (2012b). Gelotophobia: life satisfaction and happiness across cultures. *Humor Int. J. Humor Res.* 25, 23–40. doi: 10.1515/humor-2012-0002

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. T. (2014). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor Int. J. Humor Res.* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., and Proyer, R. T. (2008a). The fear of being laughed at: individual and group differences in Gelotophobia. *Humor Int. J. Humor Res.* 21, 47–67. doi: 10.1515/HUMOR.2008.002

Ruch, W., and Proyer, R. T. (2008b). Assessment criteria for the fear of being laughed at. *Swiss J. Psychol.* 67, 19–27. doi: 10.1024/1421-0185.67.1.19

Ruch, W., and Proyer, R. T. (2009). Extending the study of gelotophobia: on gelotophiles and katagelasticists. *Humor Int. J. Humor Res.* 22, 183–212. doi: 10.1515/humr.2009.009

Samson, A. C., and Meyer, Y. (2010). Perception of aggressive humor in relation to gelotophobia, gelotophilia, and katagelasticism. *Psychol. Test Assess. Model.* 52, 217–230.

Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and

descriptive goodness-of-fit measures. *Methods Psychol. Res. Online* 8, 23–74.

Smith, G. T., McCarthy, D. M., and Anderson, K. G. (2000). On the sins of short-form development. *Psychol. Assess.* 12, 102–111. doi: 10.1037/1040-3590.12.1.102

Tomkins, S. S. (1969). *Affect, Imagery, Consciousness*. New York, NY: Springer.

Weiss, E. M., Schulter, G., Freudenthaler, H., Hofer, E., Pichler, N., and Papousek, I. (2012). Potential markers of aggressive behavior: the fear of other persons' laughter and its overlaps with mental disorders. *PLoS ONE* 7:e38088. doi: 10.1371/journal.pone.0038088

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Hofmann, Ruch, Proyer, Platt and Gander. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 November 2017
doi: 10.3389/fpsyg.2017.02043

Evaluation of a Picture-Based Test for the Assessment of Gelotophobia

Willibald Ruch[1], Tracey Platt[2], Richard Bruntsch[1] and Róbert Ďurka[3]*

[1] Department of Psychology, University of Zurich, Zürich, Switzerland, [2] Department of Psychology, University of Wolverhampton, Wolverhampton, United Kingdom, [3] Department of Psychology, Catholic University in Ružomberok, Ružomberok, Slovakia

This study examines whether coding open answers in a picture-based test, as to the extent they reflect the fear of being laughed at (i.e., gelotophobia), demonstrates sufficient validity to construct a semi-projective test for the assessment of gelotophobia. Previous findings indicate that cartoon stimuli depicting laughter situations (i.e., in the pilot version of the Picture-Geloph; Ruch et al., 2009) on average elicit fear-typical responses in gelotophobes stronger than in non-gelotophobes. The present study aims to (a) develop a standardized scoring procedure based on a coding scheme, and (b) examine the properties of the pilot version of the Picture-Geloph in order to select the most acceptable items for a standard form of the test. For Study 1, a sample of $N = 126$ adults, with scores evenly distributed across the gelotophobia spectrum, completed the pilot version of the Picture-Geloph by noting down what they assumed the protagonist in each of 20 cartoons would say or think. Furthermore, participants answered the GELOPH<15> (Ruch and Proyer, 2008), the established questionnaire for the subjective assessment of the fear of being laughed at. Agreement between two independent raters indicated that the developed coding scheme allows for objective and reliable scoring of the Picture-Geloph (mean of intraclass correlations = 0.66). Nine items met the criteria employed to identify the psychometrically most reliable and valid items. These items were unidimensional and internally consistent (Cronbach's alpha = 0.78). The total score of this selection (i.e., the Picture-Geloph<9>) discriminated significantly between non-fearful, slightly, markedly, and extremely fearful individuals; furthermore, it correlated sufficiently high ($r = 0.66$; $r_c = 0.79$ when corrected for reliability of both measures) with the GELOPH<15>. Cronbach's alpha (0.73) was largely comparable whereas the estimate of convergent validity was found to be lower in one ($r = 0.50$; $r_c = 0.61$; $N = 103$) of the two samples in Study 2. Combining all three samples ($N = 313$) yielded a linear relationship between the self-report and the Picture-Geloph. With the Picture-Geloph<9> and the developed coding scheme, an unobtrusive and valid alternative instrument for the assessment of gelotophobia is provided. Possible applications are discussed.

Keywords: assessment, cartoons, gelotophobia, fear of being laughed at, perception of laughter, personality

OPEN ACCESS

Edited by:
John F. Rauthmann,
Wake Forest University, United States

Reviewed by:
Ursula Beermann,
University of Innsbruck, Austria
Hugo Carretero-Dios,
University of Granada, Spain

****Correspondence:***
Willibald Ruch
w.ruch@psychologie.uzh.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *17 September 2017*
Accepted: *08 November 2017*
Published: *21 November 2017*

Citation:
Ruch W, Platt T, Bruntsch R and Ďurka R (2017) Evaluation of a Picture-Based Test for the Assessment of Gelotophobia. Front. Psychol. 8:2043. doi: 10.3389/fpsyg.2017.02043

INTRODUCTION

The concept of the fear of being laughed at (gelotophobia) was first developed in a clinical setting to describe and explain the negative perception of laughter by certain individuals (Titze, 2009). A core feature is that gelotophobes are deeply convinced that something essential is wrong with them and therefore it is inevitable that they make a funny impression on others (Ruch and Proyer, 2008). Based on this belief, they are—with a somewhat paranoid tendency—likely to feel that laughter is directed at them and thereupon misperceive it as ridicule (Platt, 2008; Titze, 2009).

When empirical research began, it became apparent that there are gradual interindividual differences in gelotophobia and a unidimensional approach (with the absence of the fear at the lowest end and extreme fear of being laughed at the upper end of a continuum) was established (cf. Ruch et al., 2014a). Since then, a number of studies have empirically validated the usefulness of the gelotophobia concept (Samson et al., 2011; Ivanova et al., 2012; Platt et al., 2013; Papousek et al., 2014; Durka and Ruch, 2015).

Fear and shame are the dominant negative emotions reported by gelotophobes (Platt and Ruch, 2009). When confronted with scenarios of either playful teasing or mean-spirited ridicule, extreme gelotophobes were found to respond with the same high amount of fear and shame in both scenarios. In contrast, non-gelotophobes showed a distinct, and less extreme, negative emotional response only to mean-spirited ridicule (Platt, 2008). There is also evidence that they blend the expressions of joy and contempt when decoding other's facial displays of emotions (Hofmann et al., 2015; Ruch et al., 2015); i.e., for them a joyful face may hide an evil mind. On the level of physiological responses, it was found that cheerful and benign auditory laughter stimuli evoked a pronounced and more sustained decrease of the heart rate in gelotophobes (as compared to non-gelotophobes), which is regarded as indicating that gelotophobes perceive harmless laughter as a social rejection cue (Papousek et al., 2014). It was found that gelotophobes' atypical responses to laughter could be triggered by different modalities of laughter, namely, by peculiarities in facial expression of smiles and laughter, the sound of laughter, and laughter-related body movement (Ruch et al., 2014b). As regards positive emotions, gelotophobes rated having a low inclination to joy (Platt and Ruch, 2009), and these lower levels of joy were also objectively measurable in facial expressions (Platt et al., 2013).

Taken together, gelotophobes are susceptible to "false alarms"; i.e., misinterpreting friendly and innocent laughter as malicious and threatening. They have a high propensity to fear and shame when confronted with laughter situations, and they have a low inclination to feeling and expressing joy. Therefore, as gelotophobia scores increase, individuals may detect real attempts of ridicule more readily, but this sensitivity comes at the price of being systematically misled by a biased perception when facing harmless or friendly social situations. When it comes to the assessment of gelotophobia, the outlined fear-typical tendencies can be targeted when aiming to measure the fear of being laughed at.

The fear of being laughed at was identified as a trait associated with a considerable range of psychological outcomes (e.g., social withdrawal, relationship status, experience of positive affect, life satisfaction, mental health; cf. Platt and Forabosco, 2012). In psychological humor research, gelotophobia was found to moderate the experience of humor situations: as gelotophobia scores increase, the valence of the response to smiling and laughter is inverted from a positive emotional response (e.g., amusement) to a negative response (e.g., fear and shame; cf. Ruch et al., 2014a). Consequently, the fear of being laughed at is worth considering when conducting experiments that involve the processing of humorous stimuli (cf. Fink et al., 2011) and should also be considered in clinical practice when dealing with patients suffering from social withdrawal due to disproportionate feeling of fear and shame in laughter situations (Platt et al., 2016).

Assessing Gelotophobia

The standard self-report instrument is the GELOPH<15> (Ruch and Proyer, 2008), a 15-item self-report instrument utilizing a 4-point answer format. Cut-off points for slight, marked, and extreme fear of being laughed at were defined (Ruch and Proyer, 2008). While in non-clinical samples across the world typically the rate of slight fear is low (between 1.2 and 10%) and never exceeds 1% of the population (see overview in Platt and Forabosco, 2012), in clinical samples rates of 40% for slight fear and 10% for extreme fear were reported (Forabosco et al., 2009; Samson et al., 2011).

For the assessment of gelotophobia a multi-method approach was seen as desirable and therefore work on alternative methods of assessment, the structured interview (Platt et al., 2012) and the *Picture-Geloph*, a test with an open-ended answer format (Ruch et al., 2009) have been initiated. The advantage the structured interview is that it gives insight into the etiology of the problem and, in contrast to the questionnaire, does not impose on the participant the preconceived characteristics of gelotophobia. The Picture-Geloph uses 20 cartoons depicting fear-relevant situations, i.e., ambiguous social interactions showing people who were possibly being laughed at or could be seen as ridiculous. The test-taker is asked to fill in the empty thought or speech balloon and to write down what this person might be thinking or saying. The answers are then coded on a 5-point scale, ranging from −2 (i.e., answer reflects enjoyment of the situation) to +2 (i.e., answer reflects a fear of ridicule). Some of these cartoons include a laughing person (depicted by laugh utterances, or body movement) while others do not. Yet both are seen to be conducive to fearful answers. This is in line with the findings that gelotophobes also respond with increased fear and shame to harmless social situations (i.e., playful teasing; Platt, 2008) and that gelotophobes' negative responses to laughter can be triggered by interpretation of visual and acoustic modalities (e.g., laugh sounds, facial expression, and body movement; Ruch et al., 2014b). Inasmuch as it (a) necessitates the attribution of one's own experience of an ambiguous situation to another person, and (b) restricts the interpretation of the stimuli to the perspective of the protagonist and specifies the response by a thought or speech balloon (i.e., the task is not to associate freely), the Picture-Geloph may be classified as a semi-projective

test[1] (Greenstein and Tarrow, 1970; Gregory, 2004) comparable to the Rosenzweig Picture Frustration test (Rosenzweig, 1978). Compared to questionnaires semi-projective tests do have a lower face validity (and accordingly the measurement intention is not easily guessed) but unlike projective tests they can have good reliability (e.g., Sokolowski et al., 2000; Proyer, 2007).

A pilot study with the pilot version of the Picture-Geloph confirmed that gelotophobes are inclined to see mockery and laughing-at interactions in a variety of the social situations depicted by the cartoon stimuli whereas non-gelotophobes were inclined to respond with positive emotions instead (Ruch et al., 2009). While the results of the pilot study indicate that the rationale of the Picture-Geloph may be valid and promising, several steps are required before it could be used as a routine method for the assessment of the fear of being laughed at. The authors gave three recommendations to improve the test for further use: first, a larger pool of representative statements for the five steps of the rating scale needs to be developed to facilitate the coding process and to further enhance objectivity. Secondly, the importance of prior training of the coders is pointed out. In their study, the correlation between the total score of the Picture-Geloph and the Geloph<46> (i.e., the initial version of the GELOPH<15>; cf. Ruch and Proyer, 2008) was 0.72 for the trained coder but only 0.34 for the person less familiar with the concept. Thirdly, weaker items need to be identified and eliminated to eventually develop a reliable shorter standard form. In the pilot study a Cronbach's alpha of 0.68 (for all 20 items) was reported, which increased to 0.74 after tentatively eliminating the eight items with corrected item total correlations of <0.25.

CONSTRUCTION OF THE PICTURE-GELOPH

The Picture-Geloph should be applicable for normal and clinical samples. Thus, for the construction of the instrument ideally a sample is needed that covers all levels of gelotophobia. Given that typically 90% in a sample are non-gelotophobes an oversampling of individuals from the higher end of the spectrum is needed, to have a sufficient size (of slight, marked, and extreme gelotophobes) to represent the entire spectrum and to allow for reliable group comparisons.

Using such a sample the construction project involves five steps. In the first step a coding scheme is developed and appraised to be further on used in the scoring of the test. A catalog of responses will be developed from a large pool of answers and they will be assigned a score (from −2 to +2) according. The coding scheme used by Ruch et al. (2009) will be taken as a basis but modified based on the responses of the present sample that will include more high-scorers. Also the theoretical rationale for the five stets of the answer scale will be improved. It will need to be verified that different trained coders converge in assigning the scores.

In the second step, the most fitting items will be identified and selected for the standard form of the Picture-Geloph. This is accomplished by engaging in two steps of analyses: (1) identifying the items where the coded answers match the GELOPH results both in terms of discriminatory power and the hedonic level of the answers, and (2) examining the psychometric properties of these items. Regarding the former, an item was considered ideal, if an item discriminates strongly among the five groups of people defined by no, slight, marked, and extreme fear of being laughed at (as verified by a significant linear trend in an ANOVA with *post hoc* tests yielding significant differences between adjacent groups), and where the no fear group (in the GELOPH) indeed on average yields affectively positive answers (e.g., < -0.5) and the average answers of the marked and extreme groups indicates a fearful answer (e.g., >1.0). Thus, an item is not considered optimal if it correlates highly with the GELOPH, but even marked gelotophobes interpret the situation as joyful, or if even the non-gelotophobes gives give answers to be coded as gelotophobic (e.g., when there is overt laughter and respondent acknowledges the fact that laughter is directed at him or her). As responses are rated to the degree to which they are fear-typical with absolute category labels (e.g., "*Explicitly fearing laughter*" or "*Neutral*"; see **Table 2**) it is reasoned that only such stimuli can be seen as conceptually valid which on average (a) elicit fear-typical responses (as identified by the coding scheme used) in gelotophobes but fear-atypical responses in non-gelotophobes, and (b) elicit more fear-typical responses as gelotophobia scores increase in different groups of gelotophobic participants. That is, even if there are *relative* differences between groups of individuals with different degrees of the fear, the item scores derived with the unified coding scheme are desired to reflect the *absolute* presence or absence of the fear. In the second step, the internal psychometric and structural properties of the items of the pilot version Picture-Geloph will be considered to refine the selection for the standard form by selecting items with the most acceptable loadings on the first unrotated principal component and corrected item-total correlations.

The third step is to determine the reliability and convergent validity of the newly developed standard version of the test (a) in terms of its correlation with the GELOPH<15>, and (b) in terms of whether it conceptually represent the variance of gelotophobia across *all defined levels* of the fear; i.e., whether the interpretation of the answers accurately represents the levels of gelotophobia as defined by the GELOPH (e.g., non-gelotophobes give non-fearful answers also at the level of the Picture-Geloph total score).

The fourth step will derive cut-off values for the score of the standard form of the test to enable the classification of subjects into non-fearful, slightly fearful, markedly fearful, and extremely fearful groups. The steps 1–4 will be undertaken in Study 1.

The fifth and final step in the construction is to find out whether the results from third and fourth step can be replicated in a further sample (Study 2).

[1]This approach is different from semi-projective grid techniques, which typically require the test-taker to rate predefined statements concerning ambiguous stimuli (mostly pictures; cf. Ziegler et al., 2007).

STUDY 1

Study 1 was designed to conduct the first four steps described above. Thus it pursues four aims, namely (a) to elaborate a coding scheme for the open answers required by the Picture-Geloph, (b) to select the most acceptable items from the pilot version of the Picture-Geloph in order to propose a standard form of the test, (c) to determine estimates of the reliability and convergent validity of the standard form, and (d) to suggest guidelines for its practical use in terms of cut-off values for the interpretation of the scores. In the sample, participants were included who provided a meaningful answer to each of the 20 items of the pilot version of the Picture-Geloph.

Method

Participants

The sample was recruited worldwide over the Internet on a gelotophobia-dedicated website. It consisted of 126 adults, 50% male and 50% female; ages ranged from 18 years to 64 years ($M = 28.5$; $Md. = 24$; $SD = 11.6$). The sample consisted of 80.2% single, 6.3% cohabiting, 10.3% married, 1.6% divorced, and 1.6% widowed individuals.

Overall, an inspection of the averaged GELOPH<15> total scores confirmed that the recruitment strategy was successful. Participants' gelotophobia scores ranged from 1.27 to 4.0 ($M = 3.17$, $SD = 0.58$). The cut-off points for gelotophobia (i.e., 2.5 for slight, 3.0 for marked, and 3.5 for extreme fear; Ruch and Proyer, 2008) were applied and yielded 11.9% ($n = 15$) individuals with no fear, and 17.46% ($n = 22$) with slight fear, 36.51% ($n = 46$) with marked fear, and 34.13% ($n = 43$) with extreme fear of being laughed at.

Instruments

The GELOPH<15> (Ruch and Proyer, 2008) is a questionnaire assessing the level of the fear of being laughed at (i.e., gelotophobia). It consists of 15 items in a 4-point answer format (1 = *strongly disagree* to 4 = *strongly agree*). A sample item is "When others laugh in my presence I get suspicious." Cronbach's alpha was 0.89 in the present sample, which is comparable to the English norm sample ($\alpha = 0.90$; Platt et al., 2009). The GELOPH<15> has been adapted to a variety of languages across different cultural contexts (e.g., Proyer et al., 2009; Chen et al., 2011, 2013; Stefanenko et al., 2011).

The Picture-Geloph (Ruch et al., 2009), in its pilot form, is a 20-item semi-projective test assessing the fear of being laughed at. Item scores are derived by coding the degree of the positive (i.e., joyful) vs. negative (i.e., laughing at) valence of participants' written responses to cartoons. Cartoons are depicting social situations relevant to the fear of being laughed at with differing degrees of ambiguity. The 20 situations cover the following themes: (a) two persons might be laughing at or mocking a third one (five pictures), (b) a person is called to a situation in which he or she might make a fool of him- or herself (four pictures), (c) a person is in an unpleasant situation and/or might be laughed at or mocked by another person (seven pictures), (d) a person is criticized by another person (four pictures), and (e) a person is envious of others because they amuse themselves and he or she is not taking part (four pictures). The situations that are shown by the pictures are listed in **Table 1**.

As **Table 1** shows, the cartoons depict either one person obviously interacting with a protagonist (designated by a thought or speech balloon), or ambiguous situations with one or more additional persons, in which the protagonist may—but may as well not—be concerned, or group situations that require the protagonist to do something (and one situation in which the protagonist is watching two persons interacting on TV).

Procedure

Data collection was administered via a website, specifically designed to collect the data and took place over the period of 6 years. Information websites such as Wikipedia, as well as media coverage of feature stories on gelotophobia were utilized to elicit participants by providing a URL that directed interested people to the website. In accordance with the University of Zurich's code of ethics, assessment was conducted anonymously. Moreover, participants were able to quit at any time and were able to request to have any data removed from the database without any consequences or drawbacks. No personal identification information was taken but participants were offered a more in-depth assessment if they left a contact email address for where the participant's gelotophobia score would be discussed in more general terms to help them gain insight into their own gelotophobia. After logging in on the website with a made up user name and a password, the participants first filled out the GELOPH<15>. After completing they filled out the pilot version of the Picture-Geloph. The study was conducted following the ethical guidance of the University of Zurich ethics checklist. Full disclosure and informed consent was provided prior to participation in the study by clicking on an "accept and continue"-link on the website. No participant had access to the study without agreeing. Altogether 403 participants visited the website and left data. For Study 1 only participants were used that had no missing data. As extreme gelotophobia is rare it was decided to retain the remaining 277 participants for potential use in Study 2.

A coding scheme for the appraisal of the valence of responses was developed (see **Table 2**). As **Table 2** shows, responses explicitly expressing the feeling or anticipation of being laughed at, mocked, made fun of, etc., were assigned to the most extreme scale value +2 (i.e., "explicitly fearing laughter"). Responses expressing negative emotion such as shame and fear, the impulse of withdrawal from the situation, the wish that the other person(s) would stop laughing, or feeling paranoid were rated as a "negative response, but not explicitly fearing laughter" (+1) to account for their more implicit indication of the fear of being laughed at. "Neutral" (0) values were assigned to responses that were not indicative of gelotophobic symptoms but, in turn, did not exhibit positive valence as well. A value of −1 ("slight enjoyment/engagement") was assigned to responses that were not indicative of gelotophobic symptoms but expressing positive attributes, motives or emotions, and engagement in situations bearing the risk of being laughed at for the individual. Responses coded with a value of −2 ("full enjoyment/engagement") met the criteria for −1 to a higher degree, that is, responses that reflected

TABLE 1 | Descriptions of the cartoon pictures.

Item	Content	Type
1	A woman and a man are laughing ("Ha, Ha, Ha!"), a man with an empty thought balloon is watching.	a/e
2	A man on the phone is saying: "Can you speak more clearly? I cannot understand you" and a person in a phone booth with an empty speech balloon.	c/d
3	A man smiling at a woman with an empty speech balloon is saying: "Would you like to go out dancing with me? Heh-heh!"	b
4	A woman with an empty thought balloon is standing in front of a crowd of four sitting persons holding a paper in her hand.	c
5	A woman with a paper in her hand is pointing at a boy with an empty thought balloon saying "Heh! You can't even solve such an easy task!"	d
6	A man is saying to another man with an empty speech balloon: "Will you join the Karaoke? It doesn't matter if you can't sing."	b
7	Two boys are looking at each other laughing ("Hee hee!"), a girl with an empty thought balloon is looking away.	a/e
8	A man with an empty thought balloon is watching two persons on TV saying: "That's really a very funny party!"	e
9	Two persons in an audience are laughing ("Hee Hee!"), a woman with a hat is sitting in the row in front of them.	a
10	A man is saying to another man with an empty speech balloon: "Relax and don't always stand around in such an awkward manner."	d
11	Four persons are sitting around a table, among them one with an empty thought balloon and one who is saying: "Let's introduce ourselves. I will start."	c
12	A man is smiling and looking at another man who obviously has tripped and fallen backward on the ground with an empty thought balloon.	c
13	Two women are standing close to each other, one of them saying "That is really interesting what you're saying. I think that's very amusing hee-hee!" A woman with an empty thought balloon is walking away.	a
14	A man is saying to another man with an empty speech balloon: "We want to perform a comedy onstage. Want to join?"	b
15	A woman is pointing at a girl sitting on a pupil's desk with an empty thought balloon saying: "You did that really well."	c
16	A boy is running away with a book from an old man with an empty thought balloon saying: "Heh, Heh!"	c
17	A man is saying to another man with an empty speech balloon: "Heh, Heh! Hello neighbor!"	c
18	A boy is saying to a man with an empty speech balloon wearing a hat, a beard and a necktie: "You really look funny!"	d
19	A woman is saying to a man with an empty speech balloon: "Come and join us for the party, it will be a good laugh!"	b
20	A man and a child are sitting at a table laughing ("HA! HA! HA!"), a man with an empty thought balloon and a glowing nose is watching.	a/e

Type—type of situation: (a) two persons might be laughing at or mocking a third one (five pictures); (b) a person is called to a situation in which he or she might make a fool of him- or herself (four pictures); (c) a person is in an unpleasant situation and/or might be laughed at or mocked by another person (seven pictures); (d) a person is criticized by another person (four pictures), and (e) a person is envious of others because they amuse themselves and he or she is not taking part (four pictures).

TABLE 2 | Scoring key for the coding of responses.

Score	General definition	Example responses for Item 1
+2	"*Explicitly fearing laughter*": Responses that indicate fear of being laughed at by explicitly expressing the feeling or anticipation of being laughed at, ridiculed, made fun of, mocked, being seen as ridiculous, etc.	"Are they laughing at me?" "I need to get out of here they are laughing at me," "They are probably mocking me," "Why does everyone laugh at me?" "might be making fun of me."
+1	"*Negative response, but not explicitly fearing laughter*": Responses that are indicative of gelotophobic symptoms but do not explicitly express the feeling or anticipation of being laughed at (or ridiculed, made fun of, etc.) or might be as well indicative of a condition other than gelotophobia (such as social phobia), for example, expressing shame and fear, body symptoms related to fear and shame, the impulse of withdrawal from the situation, the wish that the other person(s) stop laughing, or feeling paranoid	"What are you laughing about?" "There something wrong with me (my clothes, hair,. . .)," "Am I that strange looking?" "I look ridiculous," "Jerks," "idiots," "What's so funny?" "Just ignore them," "What did I do?" "Did I do something stupid?" "Confusion," "paranoid," "Look the other way and walk away. Walk, walk, walk," "Please stop," "shut up," "They think I have a big nose," "frigid numbness, insecurity," "panic."
0	"*Neutral*": Neutral responses that are not indicative of gelotophobic symptoms but also have no positive valence in the sense of responses scored "−1" or "−2," but also nonsense and humorous answers	"What are they talking about?" "Feel they talk about me," "what is happening there?" "Hi, why are they wearing hats?" "Probably these people are sharing a funny story or joke," "I'd like to join in on the laughter but I don't know what to say."
−1	"*Slight enjoyment/engagement*": Responses that are not indicative of gelotophobic symptoms but expressing positive attributes, motives or emotions, but for some pictures also answers that reflect behavior that is contrary to a gelotophobic reaction (for example, joining a comedy performance group instead of avoiding it because of fear of being laughed at), or "healthy" reactions to a situation, less than a "−2"	"They look like their having fun," "they are enjoying themselves. Good for them," "I wish I knew that joke, seems funny."
−2	"*Full enjoyment/engagement*": Responses that match the definition of "+1" responses but in a higher degree (for example, joining a comedy performance group *enthusiastically*)	"They are having a good time as friends enjoying each others company," "she likes him and he said something funny."

General definitions apply to all items. The assignment of example responses to score values varies between items. For example, an anger response will be assigned to a gelotophobic "+1" value if there is no obvious reason to take offense (for example, in item 1) but will be scored as a "−1" if the protagonist is insulted or explicitly bothered (for example, in item 10 or 18).

enthusiasm on top of positive attributes, motives or emotions, and indicators of behavioral engagement (vs. withdrawal)[2].

A detailed definition of gelotophobia and a general characterization of gelotophobic persons, which was based on the state-of the art of the current findings of gelotophobia research, were utilized to train two coders. Furthermore, they were provided with the general definitions of rating scale steps shown in **Table 2**. Coders were blind to participants' gelotophobia scores as obtained by the GELOPH<15>. One coder derived categories of responses item-wise and for every step of the rating scale and compiled them in a catalog. The analyses were based on the ratings of this coder. The other coder was used for estimating the level of convergence.

Analysis and Results

The total scores derived by averaging each of the coders' ratings over all 20 items of the pilot version of the Picture-Geloph were highly intercorrelated ($r = 0.91$) and both coders had a perfect blind agreement in 58% of responses rated. To attain an estimate of the objectivity and reliability of the coding procedure in terms of the degree of absolute agreement among measurements, intraclass correlations (ICCs) were computed between the two coders' rating scores for the 20 items of the pilot version separately [by use of a two-way model (as the same two coders rated all responses), random effects (i.e., assuming that raters are replaceable), single measurements (i.e., analyzing individual item scores), and an agreement criterion (i.e., not adjusting the agreement for possible mean differences between the two coders to inform on the absolute objectivity of the rating procedure); cf. McGraw and Wong, 1996]. The results are given in **Table 3**. As **Table 3** shows, ICCs in the pilot version of the Picture-Geloph ranged from 0.39 to 0.83 with a median of 0.73 (and a mean of 0.67), showing that most of the variance in the ratings could be attributed to the participants (i.e., indicating an overall acceptable interrater agreement).

To ensure conceptual validity for the standard version of the Picture-Geloph, it was desired to arrive at a set of stimuli that gelotophobes respond to in a fear-typical way whereas non-gelotophobes respond without an indication of the fear of being laughed at or even in a positive way. As a second criterion, stimuli were defined as conceptually valid if responses in the groups of slight, marked, and extreme gelotophobes (as assessed with the GELOPH<15>) differed from each other in terms of different group means of scores within the Picture-Geloph[3]. Accordingly, we used participants' gelotophobia scores as assessed with the GELOPH<15> to generate five groups with different degrees of gelotophobia in order to analyze which of the stimuli elicit responses that match the outlined criteria: (a) gelotophobes' responses on average lie beyond a "neutral" threshold in terms of fear-typical responses whereas non-gelotophobes responses reflect absence of the fear in terms of positive responses, (b) group means of scores (i.e., codings of responses) show a linear increase along with the fear of being laughed at, and (c) among the group of gelotophobes, slight, marked, and extreme fear of being laughed at is reflected in higher scores among extreme gelotophobes than in the other two groups and higher scores among marked gelotophobes than in the group categorized as having a slight fear of being laughed at (according to the self-report measure).

Accordingly, Picture-Geloph single item score means were examined between the five groups with increasing gelotophobia scores separately (no fear, slight, marked, and extreme fear). Such items were selected (a) to which no-fear individuals on average responded to in a fear-atypical way (as indicated by negative group means), (b) plus to which marked fear individuals on average responded to in a fear-typical way (as indicated by positive group means), and (c) plus for which there was a constant increase in the item scores along with the gelotophobia level of the groups, i.e., for which there was no significant deviation from a linear trend (as tested using consecutive one-way ANOVAs with the GELOPH<15>, while employing gelotophobia level as the group factor and the Picture-Geloph items score as the dependent variable). These criteria led to a selection of nine items (i.e., items 1, 3, 4, 6, 7, 9, 11, 14, 19; see **Table 1** for content).

To inspect the internal psychometric properties of these items within the full pilot version of the scale, the corrected item-total correlations of the ratings of responses to the pictures were computed. Furthermore, a principal component factor analysis was performed on the intercorrelations of the ratings of responses to the 20 items in order to compute their loadings on their first unrotated principal component. There were six factors with eigenvalues exceeding unity (eigenvalues were 5.19, 1.50, 1.28, 1.22, 1.10, and 1.02). The first factor alone explained 25.95% of variance. To further inspect the properties of the single items, the means, standard deviations, and the frequencies of the different types of responses to every picture were computed. The results are given in **Table 3**.

Table 3 shows that the most frequent coding was "negative response, but not explicitly fearing laughter" (+1), and more than 50% of the answers were yielded by this and the "explicitly fearing laughter" (+2) answer categories together. All of the selected nine items had acceptable loadings on the first unrotated principal component (>0.50) and acceptable corrected item-total correlations (>0.40) within the 20-item scale. These were taken to generate the standard form of the test, which will be labeled as the *Picture-Geloph<9>* in the remainder of this report.

Evaluation of the Standard form (Picture-Geloph<9>)

ICCs in the Picture-Geloph<9> ranged from 0.56 to 0.83 with a mean of 0.66, indicating that the overall interrater agreement was as acceptable as in the pilot version. A principal component factor analysis was performed on the intercorrelations of the ratings of responses to the nine items. There were two factors with eigenvalues exceeding unity (the first factor alone explained 37.56% of variance). The inspection of the scree plot (eigenvalues

[2]Materials and data are available upon request from the first author.

[3]There is no doubt that an item which is not meeting these criteria can still be found as a good indicator of the fear of being laughed at. However, the conceptual validity can be seen as highest among those stimuli, which also according to the true meanings of the categories of the coding scheme elicit fearful responses mostly in gelotophobes but not in fear-free individuals (and fear-atypical responses mostly in fear-free individuals but not in gelotophobes). Nevertheless, items that are to difficult or to easy might average out and be useful if included together.

of the first two factors were 3.38 and 1.08) suggested that the items were unidimensional, which was substantiated by the results of a parallel analysis (Horn, 1965)[4]. Cronbach's alpha for the nine-item scale was 0.78, indicating good internal consistency.

The mean total score of ratings of responses to the items of the Picture-Geloph<9> correlated moderately strong with the subjective self-report measure (GELOPH<15>) with $r = 0.66$, $p < 0.001$ ($r_c = 0.79$, when corrected for reliability of both measures as an estimate for the correlation of the true scores). The total score of the Picture-Geloph<9> was not correlated to participants' age, $r = -0.11$, $p = 0.243$, but there was a trend for a correlation with gender, $r = 0.17$, $p = 0.061$ (with females tending to have higher scores than males).

To test whether individuals with higher degrees of self-reported gelotophobia would give more gelotophobic responses in the Picture-Geloph<9> than groups with lower degrees of subjective fear, Picture-Geloph<9> test scores were compared between four groups with different levels of gelotophobia (i.e., non-fearful individuals, individuals with slight, marked, and extreme fear of being laughed at). One-way ANOVAs with subsequent *post hoc* tests was conducted (Fisher's least significant difference, LSD; effects with $p < 0.05$ are reported), with the Picture-Geloph<9> sum score as the dependent variable and the level of self-reported gelotophobia (as defined by the cut-off points of the GELOPH<15>) as a group factor. The results are given in **Table 4**.

As **Table 4** shows, Picture-Geloph<9> sum scores differed significantly as a function of the self-reported fear with a large effect size. *Post hoc* tests revealed that means of Picture-Geloph<9> scores differed among all groups, i.e., there were score differences in the full spectrum of self-reported gelotophobia. The effect sizes of *post hoc* comparisons were large and ranged between $d = 0.68$ [95% confidence interval (CI) = (0.009; 1.36); i.e., for the comparison between the fear-free group and the group with slight gelotophobia] and $d = 2.43$ [95% CI = (1.70; 3.17); i.e., for the comparison between the fear-free group and the group with extreme gelotophobia scores].

To gain a deeper insight into the kind of responses made in the Picture-Geloph<9> by the different gelotophobia groups, the number of answers in every step of the coding scheme was counted for every person. A one-way ANOVAs with subsequent *post hoc* tests (LSD; effects with $p < 0.05$ are reported) was computed with the frequency of responses as the dependent variable for each of the answer categories separately (e.g., "negative response, but not explicitly fearing laughter," +1) and the gelotophobia level (as defined by the GELOPH<15> score) as a group factor. The results are also given in **Table 4**. As **Table 4**

[4]In the parallel analysis the eigenvalues of the factors were compared to the means of eigenvalues originating from principal components analyses of 100 datasets with random data generated by permutations of the raw data set. The eigenvalue of the first, but not the second factor, met the retention criterion, as the eigenvalue of the second factor did not exceed the mean ($M = 1.28$) and consequently also did not exceed the upper 95th percentile (1.37) of the distribution of eigenvalues of second factors retrieved from the random data sets.

TABLE 3 | Descriptive statistics, psychometric properties, interrater agreement, and the frequency distribution of the ratings of responses for the pilot version of the Picture-Geloph (Study 1).

Item	*M*	*SD*	−2 (%)	−1 (%)	0 (%)	+1 (%)	+2 (%)	ICC	CITC	FUPC
1	1.15	0.83	2.4	0.8	11.1	50.8	34.9	0.83	0.51	0.59
2	−0.07	0.94	0.0	46.8	14.3	38.1	0.8	0.56	0.42	0.50
3	0.49	0.99	5.6	11.1	19.8	55.6	7.9	0.59	0.43	0.51
4	0.46	0.81	0.8	8.7	42.9	38.9	8.7	0.59	0.53	0.61
5	0.43	0.67	0.0	7.1	46.0	43.7	3.2	0.70	0.29	0.36
6	0.34	1.16	12.7	7.9	21.4	48.4	9.5	0.78	0.42	0.51
7	0.87	0.87	1.6	2.4	27.8	43.7	24.6	0.76	0.45	0.53
8	−0.04	0.72	3.2	15.9	65.1	13.5	2.4	0.59	0.20	0.24
9	1.01	0.84	0.0	7.9	11.1	53.2	27.8	0.56	0.57	0.65
10	0.29	0.79	1.6	15.1	37.3	45.2	0.8	0.42	0.31	0.39
11	0.38	0.77	0.0	11.1	46.0	36.5	6.3	0.63	0.46	0.54
12	0.75	0.95	0.0	12.7	22.2	42.1	23.0	0.69	0.52	0.61
13	0.66	0.68	0.0	3.2	36.5	51.6	8.7	0.65	0.34	0.42
14	0.47	1.13	11.1	8.7	11.1	60.3	8.7	0.65	0.49	0.57
15	−0.07	1.03	11.9	18.3	37.3	30.2	2.4	0.83	0.37	0.44
16	0.77	0.87	0.8	5.6	30.2	42.9	20.6	0.73	0.41	0.48
17	0.03	1.03	10.3	14.3	42.9	27.0	5.6	0.74	0.37	0.44
18	−0.21	1.07	20.6	7.1	46.0	25.4	0.8	0.39	0.40	0.49
19	0.37	1.12	11.1	6.3	28.6	42.9	11.1	0.76	0.54	0.61
20	0.83	0.86	3.2	4.8	13.5	62.7	15.9	0.75	0.40	0.49
Md.	0.45	0.87	2.0	8.3	29.4	43.3	8.7	0.73	0.42	0.51

*N = 126. M, mean; SD, standard deviation; ICC, intraclass correlation between ratings of two different coders; CITC, corrected item-total correlation; FUPC, loading on the first unrotated principal component; r_{rater}, intercorrelations between the raters; Items, ratings of responses to pictures 1–20; Md., median of columns. Frequencies of responses: "+2" = "explicitly fearing laughter," "+1" = "negative response, but not explicitly fearing laughter," "0" = "neutral," "−1" = "slight enjoyment/engagement," "−2" = "full enjoyment/engagement" (see **Table 2** for details).*

TABLE 4 | Frequency of types of responses and the sum score of the Picture-Geloph<9> as a function of level of gelotophobia (Study 1).

	Gelotophobia level (GELOPH<15>)										
	No		Slight		Marked		Extreme		Group comparisons		
Picture-Geloph<9>	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*(3,122)	*p*	η^2
Frequency of responses											
"+2"	0.27_{a}	0.59	0.50_{a}	0.74	1.39	1.42	2.26	1.92	10.66	<0.001	0.21
"+1"	3.00_{a}	1.81	$3.91_{a,b}$	1.63	$4.46_{b,c}$	1.72	$4.79_{b,c}$	1.68	4.61	0.004	0.10
"0"	2.87_{a}	0.92	$2.82_{a,b}$	1.71	$2.20_{a,b,c}$	1.34	1.65_{c}	1.45	4.71	0.004	0.10
"−1"	1.53	1.19	0.95_{a}	0.79	0.63_{a}	0.80	0.21	0.47	12.75	<0.001	0.24
"−2"	1.33_{a}	1.18	0.82_{a}	1.05	0.33_{b}	0.73	0.09_{b}	0.48	11.12	<0.001	0.22
Sum score of ratings	−0.67	4.56	2.32	4.24	5.96	4.44	8.91	3.70	24.86	<0.001	0.38

N = 126. M, mean; SD, standard deviation; η^2, partial eta squared. $_{a,b,c}$*Means of one row sharing a subscript do not differ significantly from each other. Groups were defined by individuals' GELOPH<15> scores: no gelotophobia <2.5, slight gelotophobia <3.0, marked gelotophobia <3.5, and extreme gelotophobia ≥3.5 (scale maximum = 4.0). "+2" = "explicitly fearing laughter," "+1" = "negative response, but not explicitly fearing laughter," "0" = "neutral," "−1" = "slight enjoyment/engagement," "−2" = "full enjoyment/engagement" (see* ***Table 2*** *for details).*

shows, the frequencies of the five answer categories differed significantly as a function of the GELOPH<15>-defined levels of gelotophobia (no, slight, marked, extreme), with medium to large effect sizes. On a descriptive level, as gelotophobia increased across the four groups, the prevalence of fear-atypical (i.e., "−1" and "−2") answers and neutral responses decreased, whereas the prevalence of fear-typical (i.e., "+1" and "+2") answers increased. However, although differences between the groups were in the expected direction, *post hoc* tests revealed that the means of frequencies of the different answer types did not differ significantly between all pairs of adjacent groups (cf. **Table 4**).

Deriving Cut-Offs for the Practical Use of the Picture-Geloph<9>

For the use of the newly designed standard version of the test, for example, for individual testing, the correspondence between the GELOPH<15> scores and the Picture-Geloph<9> scores were examined to derive cut-off points defining different levels of gelotophobia. The sum score of ratings (see **Table 4**; computed by adding the nine single item ratings) has a theoretical span of −18 to +18 and the actual values vary from a minimum of −9 to the maximum of 16 ($M = 5.54$, $SD = 5.24$). In the GELOPH agreeing to half of the items and disagreeing to the other half yields 2.5, and this is the cut-off score for where slight gelotophobia begins. A score of 0 has the same substance in the Picture-Geloph as this means there are as many gelotophobic interpretations as non-gelotophobic ones. A score lower than 0 indicates that at least most of the answers were neutral or fear-atypical. While 0–4 defines slight fear, 5–8 stands for marked fear, and 9 and more stands for extreme fear. This yields 19.8% with slight, 36.5% with marked, and 31% with an extreme expression of the rated fear, respectively (and 12.7% with no fear, i.e., <0). These group sizes largely corresponded to the groups defined by the established GELOPH<15> cut-off points (see section "Method").

These scores also reflect differences in the GELOPH. A one-way ANOVAs with subsequent *post hoc* tests (LSD, effects with $p < 0.05$ are reported) with the GELOPH<15> mean as the dependent variable and the level of rated gelotophobia (as defined by the cut-off points of the sum score) as a group factor was conducted. Self-reported gelotophobia differed significantly as a function of the group factor as generated by the mentioned cut-off values [$F(3,122) = 31.16$, $p < 0.001$, $\eta^2 = 0.43$). *Post hoc* tests revealed that means of self-reported gelotophobia differed among all groups as defined by the cut-off values (non-gelotophobes: $M = 2.39$, $SD = 0.70$, $n = 16$; slight fear group: $M = 2.90$, $SD = 0.50$, $n = 25$; marked fear group: $M = 3.25$, $SD = 0.41$, $n = 46$, and extreme fear of being laughed group: $M = 3.57$, $SD = 0.26$, $n = 39$). The effect sizes of *post hoc* comparisons were medium to large and ranged between $d = 0.34$ [95% CI = (−0.18; 0.85); i.e., for the comparison between the group with slight gelotophobia and the group with marked gelotophobia] and $d = 1.92$ [95% CI = (1.18; 2.65); i.e., for the comparison between the fear-free group and the group with extreme gelotophobia scores]. To account for the standard error of measurement, the CI was computed for the sum score of the Picture-Geloph<9> (accepting an alpha error at the 5% level) with a margin of error of 2.10. Consequently, as a heuristic (i.e., as slightly liberal) guideline, a CI of ±2 may be suggested when using the Picture-Geloph<9> for individual testing[5]. As the theoretical range of the scale is from −18 to +18 a CI of ±2 is acceptable.

STUDY 2

As selection and validation of the Picture-Geloph<9> necessarily has been subject to the idiosyncrasy of the sample used in Study 1, an independent sample was used to cross-validate the findings. Accordingly, using an additional sample, Study 2 was designed to pursue the fifth aim of this paper: (a) to determine whether

[5]To illustrate: If a person's test score is "1," this would indicate a slight fear of being laughed at. However, taking into account the imperfect reliability of the scale, this person's true score may as well indicate the absence of the fear of being laughed at (as defined by the suggested cut-off points), i.e., the true score might as well be "−1" (i.e., "1" minus the CI of "2").

estimates of the reliability and convergent validity of the Picture-Geloph<9> are comparable to the ones found in Study 1, and (b) to find out whether the suggested guidelines for the practical use of the Picture-Geloph<9> (i.e., in terms of cut-off values for the interpretation of the scores) are useful also in this sample. Participants were included who provided a meaningful answer to each of the nine items of the newly developed standard form of the Picture-Geloph (i.e., the Picture-Geloph<9>).

Method

Participants

Sample 2 consisted of 103 adults, 44.7% male and 55.3% female; ages ranged from 18 years to 60 years ($M = 26.2$; $SD = 10.9$). The sample consisted of 70.9% single, 12.6% cohabiting, 13.6% married, 2.9% divorced, and no widowed individuals.

Overall, an inspection of the averaged GELOPH<15> total scores confirmed that the recruitment strategy again was successful. Participants' gelotophobia scores in Sample 2 ranged from 1.53 to 3.93 ($M = 3.00$, $SD = 0.58$). The cut-off points for gelotophobia (i.e., 2.5 for slight, 3.0 for marked, and 3.5 for extreme fear; Ruch and Proyer, 2008) were applied and yielded 24.3% ($n = 25$) individuals with no fear, and 75.7% ($n = 78$) gelotophobes. Among the latter there were 15.5% ($n = 16$) with slight fear, 39.8% ($n = 41$) with marked fear, and 20.4% ($n = 21$) with extreme fear of being laughed at. While the demographic characteristics of the sample were comparable to the sample used in Study 1, gelotophobia scores are lower in the present sample with more fear-free individuals and fewer extreme gelotophobes; i.e., variability was reduced.

Furthermore, a third sample (Sample 3) was used consisting of 84 adults (35% males; age: $M = 23.7$, $SD = 0.9.7$). Their GELOPH<15> scores were high on average ($M = 2.95$, $SD = 0.65$). There were 21.4% ($n = 18$) individuals each with no fear, slight fear, and extreme fear of being laughed at, respectively, while 35.7% ($n = 30$) had a marked fear of being laughed at.

Procedure

The procedure of Study 2 was identical to Study 1, except that this time the sample (Sample 2) was composed of individuals that had some missing data but answered all of the items of the newly developed 9-item standard form, i.e., the Picture-Geloph<9> (i.e., items 1, 3, 4, 6, 7, 9, 11, 14, 19, see **Table 1** for content). Cronbach's alpha of the GELOPH<15> was 0.87 in Sample 2. Sample 3 answered to at least six of the nine items and Cronbach's alpha of the GELOPH<15> was 0.89.

Analysis and Results

ICCs were computed between the two coders' rating scores for the nine items of the standard form separately in Sample 2 (again by use of a two-way model, assuming random effects, including single measurements, and an agreement criterion; cf. McGraw and Wong, 1996). ICCs in the Picture-Geloph<9> ranged from 0.48 to 0.75 with a mean of 0.61, indicating that the overall interrater agreement was somewhat lower than the one found in Study 1. To test for unidimensionality in Sample 2, a principal component factor analysis was performed on the intercorrelations of the ratings of responses to the nine items. The inspection of the scree plot (eigenvalues exceeding unity were 2.94, 1.29, and 1.10) suggested that the items were unidimensional, which was substantiated by the results of a parallel analysis (Horn, 1965)[6]. Cronbach's alpha for the 9-item scale was 0.73, indicating that internal consistency was somewhat lower than in Study 1 (0.78).

The results for the individual items (descriptive statistics, frequency distribution of the ratings of responses, interrater agreement, factor loadings, corrected item total correlations, and correlations with the GELOPH<15>) for the final version of the test (i.e., *Picture-Geloph<9>*) in Sample 2 are computed and presented in **Table 5**.

Table 5 shows that again the most frequent coding was "negative response, but not explicitly fearing laughter" (+1), and more than 50% of the answers were yielded by this and the "explicitly fearing laughter" (+2) answer categories together. All of the selected nine items loaded positively on the first unrotated principal component (median of loadings >0.50) and positive corrected item-total correlations (median >0.40) within the 20-item scale. The items correlated significantly with the subjective assessment of the fear of being laughed at confirming that these items are suited to measure gelotophobia.

At the scale level, in Sample 2 the sum score of ratings of responses to the items of the Picture-Geloph<9> ($M = 4.46$, $SD = 4.88$) correlated moderately strong with the subjective self-report measure (GELOPH<15>) measure, $r = 0.50$, $p < 0.001$ ($r_c = 0.61$, when corrected for attenuation due to imperfect reliability of both measures). The total score of the Picture-Geloph<9> was correlated to participants' age, $r = -0.21$, $p = 0.034$, and there were no gender differences, $p = 0.275$. In Sample 3, the Picture-Geloph<9> ($M = 3.29$, $SD = 5.42$) correlated highly with the GELOPH<15> measure, $r = 0.65$, $p < 0.001$, and there were no correlations with age ($r = -0.08$) or gender ($r = 0.07$). Thus, while the results will be better in the sample the items are selected than in replication samples, the high correlation in Sample 3 (similar to Sample 1) suggests that Sample 2 is the anomalous one (due to a lower variability of scores), and the results of Sample 1 can be trusted.

Again, one-way ANOVAs with subsequent *post hoc* tests was conducted (Fisher's LSD; effects with $p < 0.05$ are reported) for combined Sample 2 and Sample 3, with the Picture-Geloph<9> sum score as the dependent variable and the level of self-reported gelotophobia (as defined by the cut-off points of the GELOPH<15>) as a group factor. The results are given in **Table 6**.

As **Table 6** shows, Picture-Geloph<9> sum scores differed significantly as a function of the self-reported fear with a large effect size. *Post hoc* tests revealed that means of Picture-Geloph<9> scores differed among all groups, except the last two (i.e., extreme and marked gelotophobia), which were in the right direction, however. Furthermore, the number of answers in every step of the coding scheme was counted for every person

[6]The eigenvalue of the second factor (and consequently also the eigenvalue of the third factor) did not exceed the mean ($M = 1.31$) and consequently also did not exceed the upper 95th percentile (1.42) of the distribution of eigenvalues of second factors retrieved from the random data sets (see Study 1 for the details of the procedure).

and subjected to one-way ANOVAs again. As **Table 6** shows, the frequencies of the five answer categories differed significantly as a function of the GELOPH<15>-defined levels of gelotophobia (no, slight, marked, extreme), with medium to large effect sizes. Again, as gelotophobia increased across the four groups, the prevalence of fear-atypical (i.e., "−1" and "−2") answers and neutral responses decreased, whereas the prevalence of fear-typical (i.e., "+1" and "+2") answers increased. Again, although differences between the groups were in the expected direction, *post hoc* tests revealed that the means of frequencies of the different answer types did not differ significantly between all pairs of adjacent groups (cf. **Table 6**).

As the three samples were recruited the same way a final analysis used all of them for a comparison. Studying all three subsamples together allowed for the most reliable inquiry of the form of the function linking the Picture-Geloph to the GELOPH<15>. As there were enough participants in terms of cell sizes, two groups of non-gelotophobes were distinguished, namely borderline and no fear. The 3 (samples) × 5 (level of fear of being laughed at) ANOVA yielded a main effect for level of gelotophobia, $F(4,298) = 36.73$, $p < 0.001$, $\eta^2 = 0.33$, with no main effect of sample, $F(2,298) = 1.89$, $p = 0.15$, $\eta^2 = 0.01$, and no sample × level of gelotophobia interaction, $F(8,298) = 1.25$, $p = 0.27$, $\eta^2 = 0.03$. *Post hoc* tests revealed that all adjacent means were significantly different ($p < 0.001$), with non-gelotophobes ($n = 19$) scoring on the non-fearful side ($M = -3.16$; $SD = 4.03$) and borderline ($n = 39$) scoring in the indifference region ($M = -0.19$; $SD = 3.50$). Gelotophobes tend to give fearful answers, with the ones from slight ($n = 56$) gelotophobes being above the scale midpoint but reaching into the indifference region ($M = 2.85$; $SD = 3.98$), marked ($n = 117$) gelotophobes scoring clearly above the midpoint ($M = 5.89$; $SD = 4.72$), and extreme gelotophobes ($n = 82$) being highest with two standard deviations above the scale midpoint ($M = 7.78$; $SD = 3.92$). Except between the last two groups there is always an interval of three points between adjacent groups; i.e., there is a linear increase.

TABLE 5 | Descriptive statistics, the frequency distribution of the ratings of responses, interrater agreement, psychometric properties, and correlations with GELOPH<15> for the Picture-Geloph> (Sample 2).

Item	*M*	*SD*	−2 (%)	−1 (%)	0 (%)	+1 (%)	+2 (%)	ICC	CITC	FUPC	r_{GELOPH}
1	1.03	0.79	1	2.9	14.6	55.3	26.2	0.72	0.47	0.61	0.41**
3	0.54	0.87	2.9	9.7	24.3	56.3	6.8	0.49	0.20	0.30	0.23*
4	0.27	0.81	1.9	9.7	54.4	27.2	6.8	0.62	0.33	0.49	0.19*
6	0.38	1.16	12.6	8.7	15.5	54.4	8.7	0.64	0.57	0.73	0.31**
7	0.76	0.91	1.9	2.9	35.9	35.9	23.3	0.75	0.40	0.58	0.28**
9	0.69	0.90	0	12.6	22.3	48.5	16.5	0.58	0.43	0.60	0.21*
11	0.17	0.85	4.9	10.7	49.5	32	2.9	0.48	0.40	0.53	0.28**
14	0.42	1.18	14.6	5.8	9.7	63.1	6.8	0.58	0.52	0.68	0.32**
19	0.19	1.09	9.7	16.5	23.3	45.6	4.9	0.62	0.36	0.51	0.29**
Md.	0.42	0.90	2.9	9.7	23.3	48.5	6.8	0.62	0.40	0.58	0.28**

N = 103. M, mean; SD; standard deviation; ICC, intraclass correlation between ratings of two different coders; CITC, corrected item-total correlation; FUPC, loading on the first unrotated principal component; r_{GELOPH}, correlations between item and the GELOPH<15>; Item, item number in Study 1; Md., median of columns. Frequencies of responses: "+2" = "explicitly fearing laughter," "+1" = "negative response, but not explicitly fearing laughter," "0" = "neutral," "−1" = "slight enjoyment/engagement," "−2" = "full enjoyment/engagement" (see ***Table 2*** *for details). *p < 0.05 (two-tailed). **p < 0.01 (two-tailed).*

TABLE 6 | Frequency of types of responses and the sum score of the Picture-Geloph<9> as a function of level of gelotophobia (Study 2; Sample 2 and Sample 3 combined).

	Gelotophobia level (GELOPH<15>)										
	No		Slight		Marked		Extreme		Group comparisons		
Picture-Geloph<9>	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*(3,183)	*p*	η^2
Frequency of responses											
"+2"	0.16_a	0.48	0.53_a	0.78	1.31_b	1.79	1.20_b	1.17	8.70	<0.001	0.13
"+1"	2.59	1.46	4.34_a	1.79	4.58_a	1.81	4.91_a	2.12	14.46	<0.001	0.19
"0"	3.42	1.64	2.65_a	1.18	2.09_a	1.38	2.33_a	1.80	7.33	<0.001	0.11
"−1"	1.74	1.17	0.77_a	0.96	0.66_a	0.89	0.38_a	0.70	17.28	<0.001	0.22
"−2"	1.09_a	1.20	0.72_a	1.02	0.36_b	0.78	0.19_a	0.48	8.70	<0.001	0.13
Sum score of ratings	−0.99	3.80	3.19	3.84	5.84_a	4.92	6.54_a	3.81	28.94	<0.001	0.32

N = 187. M, mean; SD, standard deviation; η^2, partial eta squared. $_{a,b,c}$Means of one row sharing a subscript do not differ significantly from each other. Groups were defined by individuals' GELOPH<15> scores: no gelotophobia <2.5, slight gelotophobia <3.0, marked gelotophobia <3.5, and extreme gelotophobia ≥3.5 (scale maximum = 4.0). "+2" = "explicitly fearing laughter," "+1" = "negative response, but not explicitly fearing laughter," "0" = "neutral," "−1" = "slight enjoyment/engagement," "−2" = "full enjoyment/engagement" (see ***Table 2*** *for details).*

In the total sample with 313 adults the contingency between the coded levels of fear of being laughed at in the questionnaire and the picture test can be estimated. The cut-off values are applied and the cross-tabulation of scores (see **Table 7**) yielded a significant effect [$\chi^2_{(16)} = 157.94$, $p \leq 0.001$] amounting to a correlation of 0.62.

Table 7 shows that the coded level of fear tends to correspond to each other. The scores in the diagonal are highest in both row and columns. The next highest frequencies can typically be found in the two adjacent cells. Interestingly, the gelotophobes with a marked fear of being laughed at have the largest variance in their scores, including having no fear at all in the Picture-Geloph. Thus, while there is no perfect overlap, the correspondence is striking and future studies need to see whether the Picture-Geloph has incremental validity over the GELOPH<15>.

Discussion

For the present paper, two samples with a good distribution of slight, marked, and extreme gelotophobes were recruited, allowing for the construction, psychometric evaluation, and determining the validity of a standard form of the Picture-Geloph throughout the full spectrum of the fear of being laughed at. A coding scheme for the scoring of the test was developed and sufficient interrater agreement was found as an indicator of the objectivity and reliability of the standardized coding procedure. The compiled catalog of response-categories assigned to the respective scale values can be used as a reference in future studies as well as for supplementary information in individual testing in addition to the GELOPH (see **Table 2** for examples of reference answers to Item 1). Retest reliability estimates are still missing and hence it is not clear how much scores might fluctuate and depend on testing condition. However, internal consistency is high but not justifying sole administration. Hence further validation studies need to be conducted before routine use in individual testing is warranted.

In Study 1, the conceptually most valid items were selected and yielded acceptable psychometrical properties within the full scale (in terms of their loadings on the first unrotated principal component and their corrected item-total correlations). Thereby, it was ensured that the absolute meanings of the ratings of responses on average corresponded to individuals' gelotophobia scores (i.e., as assessed with the established questionnaire, GELOPH<15>). Nine items were found which (a) elicit fear-typical responses (as identified by the coding scheme used) in gelotophobes but fear-atypical responses in non-gelotophobes, and (b) elicit more fear-typical responses as gelotophobia scores increase in different groups of gelotophobic participants. These items were used to generate a standard form of the test (i.e., the Picture-Geloph<9>). The items of the Picture-Geloph<9> were unidimensional and the reliability, as estimated by the internal consistency (Cronbach's alpha = 0.78), was higher than compared to the initial study by Ruch et al. (2009), who reported a Cronbach's alpha of 0.68 for the 20-item version and a coefficient of 0.74 for their 12-item proposal of a short form.

It was found that the Picture-Geloph was suitable to assess differences in the full spectrum of gelotophobia. As expected, individuals with higher degrees of self-reported gelotophobia gave more fear-typical responses than individuals with lower degrees of subjective fear, which was not only indicated by a substantial correlation between the scores of the Picture-Geloph<9> and the GELOPH<15>, but also by a comparison of Picture-Geloph<9> scores between groups with different levels of self-reported gelotophobia. Hence, the Picture-Geloph<9> can be regarded as suitable to validly assess gelotophobia in its full spectrum. Cut-off values for the sum score of the Picture-Geloph<9> (for the classification of subjects as non-fearful, slightly fearful, markedly fearful, or extremely fearful) were derived and found to separate the sample into four groups with differing GELOPH<15> score means.

In Study 2, the results of Study 1 were generally replicable, with some exceptions: (1) internal consistency and the correlation between the scores of the Picture-Geloph<9> and the GELOPH<15> (as an estimate of convergent validity) were numerically lower in Study 2 (for Sample 2 but not Sample 3), (2) the total score of the Picture-Geloph<9> was slightly correlated to participants' age in Study 2 (only Sample 2), (3)

TABLE 7 | Crosstab of GELOPH<15> and Picture-Geloph<9> data, segmented into no fear, borderline, slight, marked, and extreme fear of being laughed at (Samples 1–3 combined).

Coded groups	No fear (1.0–2.0)	Borderline (2.1–2.5)	Slight (2.6–3.0)	Marked (3.1–3.5)	Extreme (3.6–4.0)	Total
No fear (≤ −4.0)	8	6	5	6	0	25
	42%	15%	9%	5%	–	8%
Borderline (−3.9 to 0.0)	8	16	9	6	5	44
	42%	41%	16%	5%	6%	14%
Slight (0.1–4.0)	3	14	21	22	9	69
	16%	36%	38%	19%	11%	20%
Marked (4.1–8.0)	0	3	18	51	27	99
	–	8%	32%	44%	33%	32%
Extreme (>8.0)	0	0	3	32	41	76
	–	–	5%	27%	50%	24%
Total	19	39	56	117	82	313
	100%	100%	100%	100%	100%	100%

N = 313.

the groups generated by the cut-off values for the sum score of the Picture-Geloph<9> that were derived in Study 1 did only in part differ from each other as to their mean level of self-reported gelotophobia in Study 2, indicating that the Picture-Geloph<9> was mainly suitable to discriminate between non-fearful and fearful individuals in the sample of Study 2. These deviations between the results of Study 1 and Study 2 may, partially, be attributed to the characteristics of one of the sample used in Study 2. Considering Sample 2 alone (a) the overall sample size was smaller, decreasing the power of statistical tests, (b) there was a smaller proportion of gelotophobes, and (c) especially extreme gelotophobes were less represented (as compared to Study 1), overall leading to a reduced variance in gelotophobia scores. The reduced correlation between the scores of the Picture-Geloph<9> and the GELOPH<15> in Study 2 (as compared to Study 1) may, in part, also be explained by the selection procedure employed to generate the Picture-Geloph<9> in Study 1: the criteria used to identify the conceptually most valid items may have led to selecting foremost items with a large linkage to self-reported fear. That is, preferring such items that (a) elicited scores with a lower "starting point" (i.e., non-fearful individuals, as defined by the GELOPH<15>, on average had negative scores in the selection of items), and (b) with a linear increase across the different groups of self-reported gelotophobia, may have increased both the variance of the Picture-Geloph<9> score as well as the covariance between the total scores of the Picture-Geloph<9> and the GELOPH<15> in this sample. Hence, because of the idiosyncrasies of the different samples, a lower estimate of convergent validity should have been expected when cross-validating the Picture-Geloph<9> with Sample 2. Still, there was a substantial correlation between the two measures of gelotophobia in Study 2, and adding Sample 3 yielded stronger results (despite the fact that the total score was based on six to eight items only). Taking into account that the GELOPH<15> and the Picture-Geloph<9> are different types of methods for the assessment of gelotophobia (i.e., a self-report vs. a semi-projective test), the coefficient found for the estimation of convergent validity in Sample 2 still can be seen as sufficiently high (i.e., due to a common-method effect, the correlation between two questionnaires can be expected to be higher than the correlation between a questionnaire and a different method of assessment, such as an objective or semi-projective test, even if all have the same validity).

The findings of our studies indicate, one more time, that the assumptions of Ruch et al. (2009) were substantial: gelotophobes tend to respond differently than the normal population when faced with situations in which they potentially could be laughed at, ridiculed or otherwise be evaluated as deficient or ridiculous. As a basic extension of their findings, the present study reveals that with an increasing level of the fear, this bias becomes more evident. At the same time, these results demonstrate that the Picture-Geloph<9> can be instrumental in the assessment of the varying levels of gelotophobia.

Limitations

The present study demonstrated that the cartoons that are used by the Picture-Geloph<9> are suitable to evoke valid responses in gelotophobes. However, the sampling of the stimuli from which they were selected (i.e., the pilot version of the Picture-Geloph) may neglect important aspects of the fear of being laughed at. As the situations involving laughter are highly ambiguous, they do not provide explicit evidence that the protagonist is actually being addressed by the laughter. The feeling of being laughed at, therefore, is the result of a paranoid tendency to relate laughter to oneself (cf. Platt et al., 2012). It would be interesting to also capture gelotophobes' disproportionately negative reactions in situations in which the normal population would also feel that they were being laughed at.

As a further limitation, the rating of responses was based on plausible but yet untested theoretical assumptions and therefore there were several disputable decisions made in the assignment of responses to the rating scale values by the two raters. For example, depending of the situation depicted by the item, responses reflecting *anger* were either assigned to the "negative response, but not explicitly fearing laughter" (+1) or the "slight enjoyment/engagement" (−1) rating scale value. In pictures where there was no evidence for the target person being addressed by the laughing persons, anger was interpreted as a possible sign of paranoid sensitivity to ridicule and hence rated as a possible indicator of gelotophobia, whereas angry responses to pictures in which the target person was obviously addressed with criticism or an insult were considered as a "healthy" reaction (as opposed to internalizing, i.e., thinking one deserves being criticized for one's funny looks or awkward posture) and rated with the "slight enjoyment/engagement" (−1) scale value. Such discrepancies were not observed for more extreme answers.

The rating of responses reflecting embarrassment and shame needs reconsideration too. Being embarrassed or ashamed as a consequence of "justified" ridicule can be considered a usual response. However, responses reflecting shame and embarrassment were construed as possible indicators of gelotophobia in the reference-coding catalog. Shott (1979) points out that embarrassment is originated by deficiencies of the self-*presentation*, whereas shame occurs when others view one's self—*per se*—as deficient. In line with this distinction, Tangney et al. (1996) suggest "shame is associated with more global and enduring negative attributions about oneself, whereas embarrassment is tied to more transient, situation-specific failures and pratfalls" (p. 1258). That would explain why gelotophobes are prone to shame: they misinterpret the criticism conveyed by ridicule (or rather anything they misperceive as ridicule) in line with their global belief that something essential is inherently wrong with them. But why and when are gelotophobes supposed to be embarrassed then? Shott (1979) reasons that, when the self is believed to be deficient, this also leads to the subjective impression of an inadequate self-*presentation*. As the crucial point, the situations depicted by the Picture-Geloph do not provide explicit evidence that the protagonist was transgressing relevant social norms prior to occurring laughter or criticism (or even is being laughed at), which may most likely be the reason why non-gelotophobes tend to respond neutrally or with enjoyment and the expression of positive attributes (or otherwise fear-atypically) to the stimuli. For the given reasons, responses expressing shame and embarrassment were construed

as indicators of a biased conviction to be or to appear deficient, inadequate or ridiculous and were rated equally as negative ("negative response, but not explicitly fearing laughter," i.e., +1) responses.

More research is needed on the optimal administration of the test. The setting in the present study might have primed some of the responses. The participants came to the data collection website pre-informed about the fear of being laughed at, and they also first filled in the GELOPH which might have led to different answers compared to randomly recruited participants who filled in the Picture Geloph as the first instrument. Systematic studies are needed to estimate the effects of the priming; e.g., whether level of face value is affected. Likewise, more research is needed examining when the validity of the test is maximal; such studies might vary the drawing style, but it is also of interest to add filler items to lower the face value of the test. Validity information needs to be accumulated to allow deciding whether the test can also be used in individual testing. Finally, in the present study two coders were employed (albeit only one was used further on). A study is needed to find out the optimal number of coders and the optimal level of training to be able to give profound advice on the optimal scoring circumstances.

Recommendations for Future Studies

As a recommendation for future studies, it would be desirable to further extend the range of methods for the assessment of gelotophobia, for example, by means of an objective test. An objective assessment of gelotophobia could be based on the empirical findings of the studies conducted so far. For example, it has been shown that gelotophobes respond differently in a variety of modalities (such as behavioral, physiological, emotional) when encountering different stimuli (such as hearing laughter, being teased, or judging faces; Platt et al., 2013; Papousek et al., 2014; Ruch et al., 2014b, 2015). An objective test does not have face validity, which may complement the assessment of gelotophobia with an instrument in which scores are not easily influenced by response bias. Furthermore, the responses were only rated in a quantitative manner in the present study. A qualitative analysis might allow for a deeper understanding of gelotophobia and could be applied to test theoretical assumptions.

Possible Applications

The interrater agreement indicates that, with some training, the Picture-Geloph<9> can be adopted by everyone. The proposed cut-off points for the mean total score of the ratings of responses for the 9-item standard version of the test are suitable to classify subjects into the categories non-fearful, slightly, markedly and extremely fearful. Hence, the Picture-Geloph<9> can be suggested for both the use in larger investigations and as supplementary information in individual testing. The Picture-Geloph<9> may be preferred to the GELOPH<15>, especially when it is desired not to impose the preconceived characteristics of gelotophobia on the test taker in the way that a questionnaire does. Furthermore, in individual testing, individuals' responses to the Picture-Geloph<9> might be subsequently used as a starting point for a diagnostic interview. For example, in a therapeutic context a clinician may first administer the Picture-Geloph<9> and subsequently go through the items and explore what made the patient provide the respective answers. It can be seen as advisable combining the Picture-Geloph<9> with, subsequently, the GELOPH<15> (this order of use may help not to prime individuals when answering the questionnaire), in order to safeguard diagnostic decisions against a selective method bias (e.g., when individuals have an acquiescent tendency answering the GELOPH<15>). The Picture-Geloph<9> and the coding manual can be retrieved from the first author of this paper.

Conclusion

In the present paper, an additional diagnostic tool for the assessment of gelotophobia was evaluated within a large sample of gelotophobes. The proposed 9-item standard scale allows for an economic and valid assessment of the fear of being laughed at. Furthermore, the phenomenon of the fear of being laughed at was demonstrated by means other than subjective self-reports in its full spectrum.

AUTHOR CONTRIBUTIONS

WR initiated the project and designed the concepts, TP collected data, RB, TP, and WR designed the coding scheme, RB and RĎ coded the answers. All authors contributed to the writing of the manuscript, read it critically, and gave consent to its publication.

FUNDING

The research leading to these results has received funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement n°270780 (ILHAIRE project).

ACKNOWLEDGMENT

The authors thank Alexander Stahlmann for commenting on an earlier version of this manuscript.

REFERENCES

Chen, H. C., Chan, Y. C., Ruch, W., and Proyer, R. T. (2011). Evaluating the reliability and validity of a traditional Chinese version of the PhoPhiKat-45. *Psychol. Test.* 58, 119–145.

Chen, H. C., Chan, Y. C., Ruch, W., and Proyer, R. T. (2013). "Cross-cultural perspectives on agents and targets of ridicule: a Taiwanese and Swiss perspective," in *Modern and Contemporary Approaches to Humour in China*, Vol. 2, eds J. R. M. Davis and J. V. Chey (Hong Kong: Hong Kong University Press), 215–229.

Durka, R., and Ruch, W. (2015). The location of three dispositions towards ridicule in the five-factor personality model in the population of Slovak adults. *Pers. Individ. Dif.* 72, 177–181. doi: 10.1016/j.paid.2014.08.045

Fink, A., Schwab, D., and Papousek, I. (2011). Sensitivity of EEG upper alpha activity to cognitive and affective creativity interventions. *Int. J. Psychophysiol.* 82, 233–239. doi: 10.1016/j.ijpsycho.2011.09.003

Forabosco, G., Ruch, W., and Nucera, P. (2009). The fear of being laughed at among psychiatric patients. *Humor* 22, 233–251. doi: 10.1515/HUMR.2009.011

Greenstein, F., and Tarrow, S. (1970). Political orientations of children: the use of a semi-projective technique in three nations. *Sage Prof. Papers Comp. Polit.* 1, 479–558.

Gregory, R. J. (2004). *Psychological Testing: History, Principles, and Applications*. Needham Heights, MA: Allyn & Bacon.

Hofmann, J., Platt, T., Ruch, W., and Proyer, R. (2015). Individual differences in gelotophobia predict responses to joy and contempt. *Sage Open* 5, 1–12. doi: 10.1177/2158244015581191

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185. doi: 10.1007/BF02289447

Ivanova, A., Stefanenko, E., Enikopolov, S., Proyer, R. T., and Ruch, W. (2012). The fear of being laughed at in healthy people and psychiatry patients. Assessing gelotophobia in Russia. *Brid. East. West. Psychiatry* 8, 10–17.

McGraw, K. O., and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1, 30–46. doi: 10.1037/1082-989X.1.1.30

Papousek, I., Aydin, N., Lackner, H., Weiss, E., Bühner, M., Schulter, G., et al. (2014). Laughter as a social rejection cue: gelotophobia and transient cardiac responses to other persons' laughter and insult. *Psychophysiology* 51, 1112–1121. doi: 10.1111/psyp.12259

Platt, T. (2008). Emotional responses to ridicule and teasing: should gelotophobes react differently? *Humor* 21, 105–128. doi: 10.1515/HUMOR.2008.005

Platt, T., and Forabosco, G. (2012). "Gelotophobia: the fear of being laughed at," in *Humor and Health Promotion*, ed. P. Gremigni (Hauppauge, NY: Nova Science Publishers), 229–252.

Platt, T., Hofmann, J., Ruch, W., and Proyer, R. T. (2013). Duchenne display responses towards sixteen enjoyable emotions: individual differences between no and high fear of being laughed at. *Mot. Emot.* 37, 776–786. doi: 10.1007/s11031-013-9342-9

Platt, T., Proyer, R. T., Hofmann, J., and Ventis, W. L. (2016). Gelotophobia in practice and the implications of ignoring it. *Eur. J. Humour Res.* 2, 46–46. doi: 10.7592/EJHR2016.4.2.platt

Platt, T., and Ruch, W. (2009). The emotions of gelotophobes: shameful, fearful, and joyless? *Humor* 22, 91–110. doi: 10.1515/HUMR.2009.005

Platt, T., Ruch, W., Hofmann, J., and Proyer, R. T. (2012). Extreme fear of being laughed at: components of gelotophobia. *Isr. J. Humor Res.* 1, 86–106.

Proyer, R. T. (2007). Convergence of conventional and behavior-based measures: towards a multimethod approach in the assessment of vocational interests. *Psychol. Sci.* 49, 168–183.

Proyer, R. T., Ruch, W., Ali, N. S., Al-Olimat, H. S., Amemiya, T., Andualem, T., et al. (2009). Breaking ground in cross-cultural research on the fear of being laughed at (gelotophobia): a multi-national study involving 73 countries. *Humor* 22, 253–279. doi: 10.1515/HUMR.2009.012

Rosenzweig, S. (1978). *The Rosenzweig Picture Frustration (P-F) Study*. St. Louis: Rana House.

Ruch, W., Altfreder, O., and Proyer, R. T. (2009). How do gelotophobes interpret laughter in ambiguous situations? An experimental validation of the concept. *Humor* 22, 63–89. doi: 10.1515/HUMR.2009.004

Ruch, W., Hofmann, J., and Platt, T. (2015). Individual differences in gelotophobia and responses to laughter-eliciting emotions. *Pers. Individ. Dif.* 72, 117–121. doi: 10.1016/j.paid.2014.08.034

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. T. (2014a). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., Platt, T., Hofmann, J., Niewiadomski, R., Urbain, J., Mancini, M., et al. (2014b). Gelotophobia and the challenges of implementing laughter into virtual agents interactions. *Front. Hum. Neurosci.* 8:928. doi: 10.3389/fnhum.2014.00928

Ruch, W., and Proyer, R. T. (2008). Who is gelotophobic? Assessment criteria for the fear of being laughed at. *Swiss J. Psychol.* 67, 19–27. doi: 10.1024/1421-0185.67.1.19

Samson, A. C., Huber, O., and Ruch, W. (2011). Teasing, ridiculing and the relation to the fear of being laughed at in individuals with Asperger's syndrome. *J. Autism Dev. Disord.* 41, 475–483. doi: 10.1007/s10803-010-1071-2

Shott, S. (1979). Emotion and social life: a symbolic interactionist analysis. *Am. J. Soc.* 84, 1317–1334.

Sokolowski, K., Schmalt, H.-D., Langens, T. A., and Puca, R. M. (2000). Assessing achievement, affiliation, and power motives all at once: the multi-motive grid (MMG). *J. Pers. Assess.* 74, 126–145. doi: 10.1207/S15327752JPA740109

Stefanenko, E. A., Ivanova, E. M., Enikopolov, S. N., Proyer, R. T., and Ruch, W. (2011). Diagnosing the fear of being laughed at: Russian adaptation of the gelotophobia questionnaire [Диагностика страха выглядеть смешным: русскоязычная адаптация опросника гелотофобии]. *Psychol. J. [Психологический журнал]* 32, 94–108.

Tangney, J., Miller, R. S., Flicker, L., and Barlow, D. (1996). Are shame, guilt, and embarrassment distinct emotions? *J. Pers. Soc. Psychol.* 70, 1256–1269. doi: 10.1037/0022-3514.70.6.1256

Titze, M. (2009). Gelotophobia: the fear of being laughed at. *Humor* 22, 27–48. doi: 10.1515/HUMR.2009.002

Ziegler, M., Schmidt-Atzert, L., Bühner, M., and Krumm, S. (2007). Fakability of different measurement methods for achievement motivation: questionnaire, semi-projective, and objective. *Psychol. Sci.* 49, 291–307.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Ruch, Platt, Bruntsch and Ďurka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 February 2018
doi: 10.3389/fpsyg.2018.00078

Relations of Dispositions toward Ridicule and Histrionic Self-Presentation with Quantitative and Qualitative Humor Creation Abilities

Karl-Heinz Renner[1] and Leonie Manthey[2†]*

[1] Department of Psychology, Bundeswehr University Munich, Munich, Germany, [2] Department of Psychology, FernUniversität in Hagen, Hagen, Germany

OPEN ACCESS

Edited by:
Willibald Ruch,
University of Zurich, Switzerland

Reviewed by:
Xiaodong Yue,
City University of Hong Kong, Hong Kong
Claudia Harzer,
Technische Universität Darmstadt, Germany

****Correspondence:***
Karl-Heinz Renner
karl-heinz.renner@unibw.de

†Present address:
Leonie Manthey,
Youke Sterke Jeugd, Utrecht, Netherlands

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *22 October 2017*
Accepted: *18 January 2018*
Published: *13 February 2018*

Citation:
Renner K-H and Manthey L (2018) Relations of Dispositions toward Ridicule and Histrionic Self-Presentation with Quantitative and Qualitative Humor Creation Abilities. Front. Psychol. 9:78. doi: 10.3389/fpsyg.2018.00078

Previous research has shown that humor and self-presentation are linked in several ways. With regard to individual differences, it turned out that gelotophilia (the joy of being laughed at) and katagelasticism (the joy of laughing at others) are substantially associated with the histrionic self-presentation style that is characterized by performing explicit As-If-behaviors (e.g., irony, parodying others) in everyday interactions. By contrast, gelotophobia (the fear of being laughed at) shows a negative correlation with histrionic self-presentation. In order to further contribute to the nomological network, we have explored whether the three dispositions toward ridicule and laughter as well as histrionic self-presentation are related to humor creation abilities. In doing so, we have assessed the four constructs in a study with 337 participants that also completed the Cartoon Punch line Production Test (CPPT, Köhler and Ruch, 1993, unpublished). In the CPPT, subjects were asked to generate as many funny punch lines as possible for six caption-removed cartoons. The created punch lines were then analyzed with regard to quantitative (e.g., number of punch lines) and qualitative (e.g., wittiness of the punch lines and overall wittiness of the person as evaluated by three independent raters) humor creation abilities. Results show that both gelotophilia and histrionic self-presentation were positively correlated with quantitative and qualitative humor creation abilities. By contrast, gelotophobia showed slightly negative and katagelasticism no associations with the assessed humor creation abilities. These findings especially apply to the subgroup of participants that created punch lines for each of the six cartoons and partly replicate and extend the results of a previous study by Ruch et al. (2009). Altogether, the results of our study show that individual differences in humor-related traits are associated with the quantity and quality of humorous punch lines. It is argued that behavior-related or performative humor creation tasks should be considered in addition to the CPPT in order to open up new avenues that can cross-fertilize research on individual differences in humor and self-presentation.

Keywords: humor, self-presentation, gelotophobia, gelotophilia, katagelasticism, self-presentation styles, histrionic self-presentation

INTRODUCTION

Since the introduction of gelotophobia, gelotophilia, and katagelasticism as individual differences variables (Ruch and Proyer, 2009a,b), these dispositions toward ridicule and laughter have attracted considerable research activities. Most of the extant studies refer to gelotophobia, the fear of being laughed at (see Ruch et al., 2014 for an overview), that has also stimulated cross-cultural research (e.g., Proyer et al., 2012b; Kamble et al., 2014) and turned out to be a decisive predictor of bullying (Platt et al., 2009). But also gelotophilia, the joy of being laughed at, and katagelasticism, the joy of laughing at others, were investigated in many studies and showed relations, e.g., with character strengths (Proyer et al., 2014), the Big Five (Ruch et al., 2013), parenting styles (Proyer et al., 2012a) in addition to gelotophobia. Furthermore, Renner and Heydasch (2010) have applied a self-presentational view with regard to the three dispositions toward laughter and ridicule. In the remainder of this introduction, we will explicate why self-presentation and especially the histrionic self-presentation style is related to humor and laughter and adds to the nomological network of gelotophobia, gelotophilia, and katagelasticism. Furthermore, we derive hypotheses regarding the relations of the three dispositions toward ridicule and the histrionic self-presentation style with quantitative and qualitative humor creation abilities.

Gelotophobia, Gelotophilia, Katagelasticism, and Self-Presentation Styles

Self-presentation and humor are linked in several ways (Renner and Heydasch, 2010): First, people need certain presentational skills to create or increase humorous effects. Obviously, some people are more capable in telling jokes and making puns than others. These presentational or performative aspects of joking (e.g., gestures, voice shifts, timing, pantomime) are also referred to as non-verbal humor (Norrick, 2004). Second, self-presentation aims at influencing and managing the impressions of others and this can be achieved by humor and laughter (Rosenfeld et al., 1983), e.g., people can laugh about a joke in order to be perceived as friendly and agreeable, or a young man may amuse his beloved by making jokes in order to enhance his attractiveness (Cooper, 2005). Third, individual differences in self-presentation have turned out to predict wittiness, e.g., high self-monitors according to Snyder (1974), i.e., persons who are skilled and motivated to engage in self-presentation, were assessed as wittier than their low-self-monitoring counterparts in a study by Turner (1980).

Renner and Heydasch (2010) have shown theoretical and empirical links between the three dispositions toward ridicule and laughter with the acquisitive and the protective self-presentation style (Arkin, 1981; Wolfe et al., 1986; Laux and Renner, 2002), and especially the histrionic self-presentation style (Renner et al., 2008).

The first two styles refer to individual differences in self-presentation that are either success- or failure-oriented: Acquisitive Self-Presenters are motivated to adapt their behaviors according to the requirements of a given social situation in order to win social approval. By contrast, protective self-presenters change their behaviors in social situations in order to avoid social disapproval. The theoretical and empirical analysis of the three dispositions toward ridicule in terms of (individual differences in) self-presentation was suggested because Ruch and Proyer (2009a, p. 184) also used a role-theoretical framework in order to explicate gelotophobia, gelotophilia and katagelasticism. Thus, gelotophiles and katagelasticists play rather active roles because they create humor that is directed toward themselves (gelotophilia) or at others' expense (katagelasticism), whereas gelotophobes play the passive role of being the target of laughter.

It turned out that gelotophobia is markedly associated with the protective self-presentation style that aims at avoiding social disapproval (Renner and Heydasch, 2010). As expected, protective self-presenters tend to interpret being laughed at by others as an indicator of social rejection and disapproval. By contrast, the acquisitive self-presentation style that aims at winning social approval, is negatively correlated with gelotophobia but showed a small positive correlation with gelotophilia. Winning social approval may sometimes be accomplished by making other people laugh at one's own expense.

The most pronounced positive associations emerged between gelotophilia, katagelasticism and the histrionic self-presentation style, a personality variable that comprises individual differences in using As-If-behaviors in everyday interactions. Histrionic self-presentation is defined "...as a way of shaping everyday interactions by *explicit* As-If-behaviors. Histrionic self-presenters regard daily situations as opportunities for role playing and for transforming such situations into 'dramatic scenes'" (Renner et al., 2008, p. 1303). Histrionic As-If-behaviors are not meant seriously and often appear in the form of jokes and teasing. Ironic remarks are subtle forms of As-If-behavior whereas imitating another person by changing one's voice, mimic, gestures or posture and trying to involve other people in such role plays would be an example of a small dramatic scene. Renner and Heydasch (2010) have pointed out that histrionic As-If-behaviors pervade our everyday life and are often used by entertainers in the media. But even certain politicians sometimes use As-If-behaviors, e.g., the former German minister of defense, Peter Struck, imitated a "Blues Brother" during an election campaign and thus a character from the cult movie of the same name by John Landis.

Based on their theoretical and empirical analyses, Renner and Heydasch (2010) have argued that both dispositions toward ridicule, especially gelotophilia and katagelasticim, and the histrionic self-presentation style may contribute to their respective nomological networks. They suggested that gelotophilia and katagelasticism may be interpreted as specific types of humorous As-If-behaviors that are at the same time associated with different preferences regarding humor appreciation (laughing at oneself with the audience vs. laughing at others). On the other hand, histrionic self-presentation highlights a possible mechanism (doing as-if) that is especially important concerning the performative (or non-verbal) aspects of making jokes about oneself or about others.

Aims of the Current Research and Hypotheses

In order to further contribute to their nomological networks, we have explored the differential impact of the three dispositions toward laughter and ridicule and histrionic self-presentation with regard to humor creation abilities as assessed in a performance test. As Cronbach and Meehl have already pointed out in 1955, in order to validate a construct that is said to have certain meanings or in more concrete terms in order to "...'make clear what something is' ..." (Cronbach and Meehl, 1955, p. 290) we need to specify the "laws" in which a construct occurs. Cronbach and Meehl (1955, p. 290) "...refer to the interlocking system of laws which constitute a theory as *a nomological network*" (italics in the original text). Although the term "associative network" seems to be more appropriate in psychology because we seldom can specify real laws but only probabilistic associations, the decisive message is that testable hypotheses may be derived from the proposed meaning or interpretation of a construct. The respective results, then, may or may not contribute to the nomological or associative network and thus to the question what a construct actually "is".

Ruch et al. (2009) have already shown that gelotophilia and katagelasticism tend to be positively associated with the ability to create humor as assessed with the Cartoon Punch line Production Test (CPPT), in which subjects are asked to write as many witty punch lines to caption-removed cartoons as they can think of. Gelotophobia was unrelated but not negatively associated with humor creation ability as assessed in the CPPT. These results refer to measures of *qualitative* humor creation ability, i.e., the ability to produce punch lines that are evaluated as witty by independent raters and the evaluation of the entire person as witty based on all generated punch lines. Interestingly, neither gelotophilia nor katagelasticism were significantly associated with quantitative humor creation, that is, the number of punch lines created in the CPPT. Gelotophobia even showed a low but not significant negative relation with the quantity of punch lines.

Up to now, the histrionic self-presentation style was shown to be associated with humor in behavior-related tasks, e.g., the rated humorousness of a presentation task and a simulated talk show in which participants had to play different guests by quickly changing between the respective roles. Histrionic self-presentation was also related to the use of humor as a coping reaction (Renner et al., 2008). Based on these previous findings, one may argue that histrionic self-presenters do need ideas that are at least witty at a medium level as a starting point for their As-If-performances. Indeed, in a single-item originality test, histrionic participants were asked to list as many different and original descriptions as possible for a trivial figure (Renner, 2006, see **Figure 1**).

Here are some of the most original solutions that were determined by a rating procedure:

- Instructions for putting together an IKEA cupboard
- Installation by Josef Beuys: felt rolls made of low fat quark
- Man peeing with a wide range of sprinkling in super slow-motion
- Gymnastics for seniors: Let's practice lying on our backs

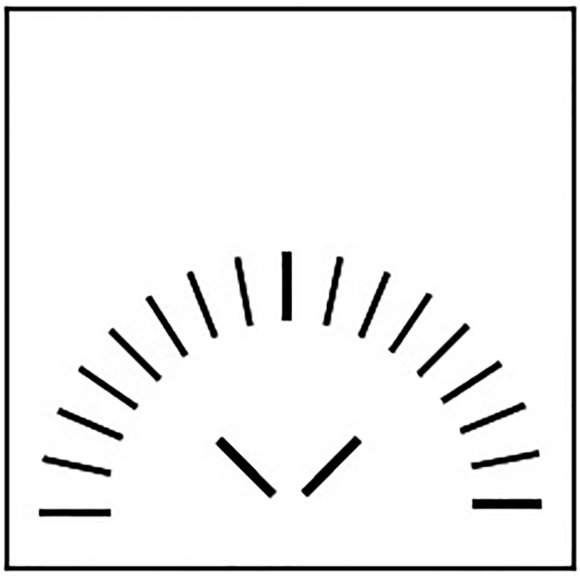


FIGURE 1 | Single-item-originality test.

Based on these preliminary findings, we hypothesize that the histrionic self-presentation style should be positively associated with the ability to create witty punch lines in the CPPT. In addition, we also expect a positive relation between histrionic self-presentation and the quantity of punch lines produced. During the course of a spontaneous and improvised histrionic performance the actors need to quickly adapt to the reactions of their interaction partners and generate new witty ideas to be successful. Finally, we wanted to check whether the findings by Ruch et al. (2009) with regard to the humor creation abilities of gelotophobes, gelotophiles, and katagelasticists are replicable.

MATERIALS AND METHODS

Participants and Procedure

This study is based on a total of $N = 337$ participants (254 women) with a mean age of $M = 33.17$ years ($SD = 10.70$, Mdn = 30) ranging from 14 to 63. Participants were first-year undergraduate students of a distance learning program in psychology at a German University (B.Sc.; $N = 256$) and persons from the circle of acquaintances ($N = 81$) of the research group who conducted the data collection of the study. The students of the distance learning university differ regarding age and occupation from the common population of young, mostly female psychology freshmen: Apart from the higher mean for age, the majority of the participants (65,3%) were employed and an additional 21,7% were currently not employed, but had been employed in the past for at least 6 months. Only 13,1% of the sample had never been employed for a period longer than 6 months. Of the 81 participants from the circle of acquaintances of the research group 59 were studying very diverse subjects (9 economics, 6 different teacher training programs,

5 psychology, 4 informatics, 3 educational science, 3 German philology, 2 biology, 2 mathematics and the other 25 participants other specific subjects) at universities and universities for applied sciences throughout Germany.

Procedure

Participants were invited by email to take part in the study that was conducted completely online using the UNIPARK program of questback. The study was accessible via a link in the emails; this link was also available on the website of the psychology department. Participants had to generate an individual six-digit code according to fixed specifications at the beginning of the study that was used to match their data with other studies. After entering this code, the purpose of the survey and details on data protection were provided. Then, questions on demographic characteristics and the questionnaires on the three dispositions toward laughter and ridicule and the histrionic self-presentation style followed. The quantitative and the qualitative humor creation ability were assessed next using the Cartoon Punch line Production Test (see next section). At the end of each survey, students received a certification of participation.

Instruments

Cartoon Punch Line Production Test (CPPT)

Qualitative and quantitative humor creation abilities were assessed with the German version of the CPPT-K (Köhler and Ruch, 1993, unpublished; Köhler and Ruch, 1996) that consists of six caption-removed cartoons related to the three humor categories incongruity resolution, nonsense, and sexual humor (2 each). These three humor categories were derived from factor-analytic studies and pertain to jokes and cartoons that (1) contain an irritating incongruity that can be completely resolved, (2) contain an incongruity that cannot be or cannot be completely resolved or produces new absurdities (nonsense) and (3) are characterized by more or less explicit sexual content (Ruch, 1992). Although the differentiation between these three categories is important and underlines the background or even "rootedness" of the CPPT in extant humor research, "... it still remains to be shown that the type of humor depicted in the cartoons as well as the contents of the produced responses do indeed matter in the process of humor production (Ruch and Heintz, forthcoming, p. 34). Thus, in the analysis of the punch lines that are produced in the CPPT, the three humor categories are not specifically considered. Ruch and Heintz (forthcoming) report on the validity of the CPPT: It turned out that the CPPT scores were positively correlated with openness to experience and several other self-report measures of humor production, but showed negative associations with seriousness.

Subjects were instructed to create as many punch lines for each of the six cartoons as they were able to. Contrary to the original paper-pencil mode of the CPPT, but similar to the administration of this test in the study by Ruch et al. (2009), we presented the six cartoons online. In accordance with the initial instruction, we administered the CPPT with a time restriction: each cartoon was shown for 2.5 min and then the next cartoon was shown. Thus, each participant produced punch lines in a total period of 15 min.

The total number of punch lines produced by each participant across the six cartoons indicates the *quantitative humor creation ability* (CPPT NP score). Overall, 2771 punch lines were created by the total sample. The number of cartoons for which a punch line was created by each participant (CPPT NC score) may be used as a fluency score. The *qualitative humor creation ability* was assessed by three independent female psychology students who rated the best punch line for each cartoon on a 10-point Likert-scale ranging from 1 = "not witty at all" to 10 = "extremely witty". Each rater was free to select the punch line she perceived as most witty after reading the created lines for each of the six cartoons. If only one punch line was produced for a cartoon, it was this punch line that was assessed regarding wittiness on the 10-point Likert scale. First, the total score (sum) of the wittiness of the best punch line (averaged across the three raters) for all cartoons (CPPT WP) was calculated. The wittiness ratings were then averaged across the cartoons *for which a punch line was provided* (CPPT WPM score). Thus, this last score is not simply the CPPT WP score divided by six (cartoons) but only divided by the individual number of cartoons for which a punch line was created at all. As a consequence, the CPPT WP score and the CPPT WPM score represent different aspects of the qualitative humor creation ability. Whereas the CPPT WPM score represents the average *maximum* wittiness, the CPPT WP score indicates a combination of fluency and wittiness. The two scores can differ dramatically, e.g., a person who has only created a single punch line that is rated with a high score of, say, 8, will receive this score for both the CPPT WP and the CPPT WPM, whereas another person who has created punch lines for each of the six cartoons that were rated with a score of, say, 6 each, will receive a CPPT WP of 36 and a CPPT WPM of 6.

After the ratings for each punch line were provided, the raters judged the overall wit and fantasy of the person. In general, it is expected that witty persons produce witty punch lines. It may be, however, that also a non-witty person produces a witty punch line once in a while but this exception may not lead to the overall assessment of the person as very witty. The raters were asked how pronounced the ability of a given subject is to produce a witty effect and answered this question on a 10-point Likert scale ranging from 1 = "not at all" to 10 = "extremely strong". The fantasy of the person was assessed on a bipolar scale ranging from – 4 = "unimaginative" to + 4 = "imaginative" and thus on a 9-point Likert scale (0 is included).

The interrater reliabilities (treating the ratings of the three raters as items) were calculated for each of the six cartoons and the overall wit of the person. The reliabilities for the six cartoons were 0.55, 0.82, 0.83, 0.85, 0.84 and 0.87. Thus, only the reliability for the first cartoon was low and the mean reliability across the six cartoons is still .81. The reliabilities for the overall wit and the fantasy of the person were 0.64 and 0.71, respectively, which may be evaluated as acceptable.

PhoPhiKat

The German version of the PhoPhiKat-45 (Ruch and Proyer, 2009a) was administered to assess gelotophobia, gelotophilia, and katagelasticism. The PhoPhiKat-45 measures these three humor-related traits with 15 items per dimension.

Ruch and Proyer (2009a) reported internal consistencies of $\alpha = 0.88$ for gelotophobia, $\alpha = 0.87$ for gelotophilia, and $\alpha = 0.84$ for katagelasticism. The respective reliability coefficients in the study at hand were quite similar with $\alpha = 0.87$ for gelotophobia, $\alpha = 0.83$ for gelotophilia, and $\alpha = 0.83$ for katagelasticism. Sample items include "When I have made a fool of myself in front of others I grow completely stiff and lose my ability to behave adequately" (gelotophobia), "For raising laughs I pleasurably make the most out of embarrassments or misfortunes that happen to me which other people would be ashamed of" (gelotophilia) and "Since it is only fun, I do not see any problems in compromising others in a funny way" (katagelasticism). Items are administered with a four-point-scale (1 = strongly disagree; 2 = moderately disagree; 3 = moderately agree; 4 = strongly agree). The validity of the PhoPhiKat-45 was shown in several studies, e.g., in the initial Ruch and Proyer (2009a) study, it is shown that the scores of the three dispositions toward laughter and ridicule were differently related to remembered experiences of being laughed at during childhood. Furthermore, the three dispositions toward laughter and ridicule showed the expected associations with the three dimensions of Eysenck's PEN model of personality (Eysenck and Eysenck, 1991): Gelotophobia was positively correlated with neuroticism and negatively with extraversion, gelotophilia was primariliy related to extraversion and katagelasticism was positively associated with extraversion and psychoticism (Proyer and Ruch, 2010).

As-If-Scale (AIS)

The histrionic self-presentation style was measured by the German version of the As-If-Scale (AIS; Renner et al., 2008). The AIS is an 8-Item scale that covers subtle histrionic forms ("I formulate my statements in such a way that they could have more than one meaning to others"), dramatic performances ("I enjoy putting on a real show for others"), and As-If behaviors that are especially related to changes in body language or nonverbal communication ("When I tell stories I act out the roles of the different participants by imitating their body language and the way they talk."). The internal consistency was $\alpha = 0.82$ in this study. The validity of the AIS is shown in Renner et al. (2008): The AIS-score predicted several concrete As-If-behaviors in role-playing tasks and was also associated with the rated wittiness across several role plays. Furthermore, subjects with high scores on the AIS were able two quickly change between different roles in a simulated talk show.

RESULTS

Table 1 shows the descriptive statistics and reliabilities for as well as gender differences with regard to the dispositions toward laughter and ridicule and histrionic self-presentation. Means and standard deviations for gelotophobia, gelotophilia, and katagelasticism in our sample were quite similar to those reported by Ruch and Proyer (2009a), and men showed higher katagelasticism than women as well. By contrast, the histrionic self-presentation style was more pronounced in the sample at hand than in most of the previous studies (Renner et al., 2008, study 1, sample 1: $t(477) = 5.65$, $p < 0.05$, $d = 0.57$; Renner and Heydasch, 2010: $t(978) = 13.09$, $p < 0.05$, $d = 0.86$). No differences, however, emerged between the histrionic self-presentation style in the study at hand and study 2 in Renner et al. (2008. $t(451) = -1.42$ n.s.). This result may be due to the fact that both the extant study and study 2 were announced with an explicit hint on humor and As-If behaviors. Thus, self-selection of participants with humorous and histrionic tendencies obviously was the case. As in these previous studies, men scored higher on histrionic self-presentation than women. The skewness and kurtosis statistics show that the distributions of the four traits were reasonably normal.

Also, as in previous studies (Ruch and Proyer, 2009a; Renner and Heydasch, 2010), gelotophobia was negatively associated with gelotophilia ($r = -0.24$, $p < 0.01$), whereas gelotophilia was markedly and positively correlated with katagelasticism ($r = 0.39$, $p < 0.01$). As in the study by Renner and Heydasch (2010), but contrary to Ruch and Proyer (2009a), gelotophobia was slightly correlated with katagelasticism in the extant sample ($r = 0.15$, $p < 0.01$). Again (see Renner and Heydasch, 2010), the histrionic self-presentation style showed marked positive associations with gelotophilia ($r = 0.47$, $p < 0.01$) and katagelasticism ($r = 0.37$, $p < 0.01$), but no relation to gelotophobia ($r = -0.03$).

Table 2 shows the descriptive statistics for the CPPT-scores that indicate quantitative and qualitative humor creation abilities. With regard to the quantitative scores, participants generated an average of approximately 8 punch lines across the six cartoons with a considerable range between 1 and 29 (CPPT NP). Furthermore, punch lines were created for 4 to 5 cartoons on average (CPPT NC), in detail: 40.4% (136 subjects) of the participants created punch lines for the entire six cartoons, 21.7% for five, 17.8% for four, 9.5% for 3, 5.6% for two cartoons and 5.0% for only one cartoon. These quantitative scores were slightly higher than in the study by Ruch et al. (2009) in which the mean for CPPT NP was 7.54 ($SD = 4.55$, range 1–23) and the mean for CPPT NC was 4.36 ($SD = 1.68$, range 1–6).

Concerning the scores that indicate qualitative humor creation abilities, the total score (sum) of the wittiness of the best punch line for all cartoons (CPPT WP) was also higher than in the study by Ruch et al. (2009). The same is true for the mean wittiness of the best punch line across cartoons and raters (CPPT WPM) that is located, however, still below the midpoint of the scale that ranges from 1 to 10. The latter also applies to the wit and the fantasy of the person (CPPT WI). The wit of the person cannot be compared with the respective score in Ruch et al. (2009), because a different scaling (1–7 instead of 1–10) was used in this study and the fantasy scale was not applied.

The skewness and kurtosis statistics show that the distributions of the qualitative scores are reasonably normal. By contrast, the distribution of the total number of punch lines (CPPT NP) is positively skewed and leptokurtic (positive excess kurtosis), i.e., there are few participants that generated punch lines above the average and the distribution shows fatter tails. In addition, several outliers were identified in the distribution of the total number of punch lines. The distribution of the number of cartoons for which a punch line was created (CPPT NC), is negatively skewed and shows a near zero excess kurtosis, i.e., few participants generated punch lines for less than

TABLE 1 | Descriptive statistics and gender differences for the dispositions toward laughter and ridicule and histrionic self-presentation.

	Total ($N = 337$)				Women ($N = 254$)	Men ($N = 83$)	Gender differences	
Personality constructs	***M* (*SD*)**	***Sk***	***K***	**α**	***M* (*SD*)**	***M* (*SD*)**	***t* (*df* = 335)**	***d***
Gelotophobia	1.83 (0.51)	0.72	0.12	0.87	1.86 (0.52)	1.76 (0.49)	1.62	0.20
Gelotophilia	2.38 (0.49)	−0.05	−0.01	0.83	2.35 (0.48)	2.47 (0.49)	−1.85^{+}	−0.25
Katagelasticism	1.95 (0.46)	0.32	0.13	0.83	1.88 (0.44)	2.17 (0.46)	−5.25**	−0.64
Histrionic SPS	2.40 (0.59)	0.59	−0.50	0.82	2.35 (0.58)	2.54 (0.60)	−2.60*	−0.32

SPS = Self-Presentation Style, Sk = skewness, K = kurtosis, $^{+}p < 0.10$, $*p < 0.05$, $**p < 0.01$.

TABLE 2 | Descriptive statistics for CPPT scores that indicate quantitative and qualitative humor creation abilities.

CPPT-scores	***M***	***SD***	***Min***	***Max***	***Sk***	***K***	***Sex***	***Age***
Quantitative scores								
CPPT NP	8.22	5.17	1	29	1.12	1.20	0.15**	0.15**
CPPT NC	4.66	1.47	1	6	−0.97	0.01	0.14**	0.21**
Qualitative scores								
CPPT WP	17.80	7.15	2	39	0.02	−0.36	0.13*	0.09
CPPT WPM	3.83	0.95	1	7	−0.12	0.15	0.04	−0.07
CPPT WI	3.92	1.36	1	8	0.28	−0.13	0.07	−0.03
CPPT FA	−0.24	1.65	−4	4	−0.09	−0.40	0.08	0.00

Sk = skewness, K = kurtosis, CPPT NP = total number of punch lines, CPPT NC = number of cartoons for which a punch line was created, CPPT WP = sum of the wittiness of the best punch line for all cartoons, CPPT WPM = mean of best punch line across cartoons and raters, CPPT WI = mean wit of the person across raters, CPPT FA = mean fantasy of the person across raters. Correlations with sex and age are non-parametric Spearman correlations. $*p < 0.05$, $**p < 0.01$.

the average number of cartoons. Due to these slight deviations from the normal distribution and especially because of the outliers, we calculated non-parametric Spearman correlations. As these correlations show, age and sex were positively associated with the quantitative indicators of humor creation ability, meaning that male gender and higher age were associated with the creation of more punch lines for more cartoons. In addition, the total score (sum) of the wittiness of the best punch line tended to be higher for men, whereas age showed no significant relation with this indicator of qualitative humor creation ability.

Table 3 shows the associations between the three dispositions toward laughter and ridicule and histrionic self-presentation with the CPPT-scores. Due to deviations from the normal distribution and outliers with regard to the CPPT, but also the gelotophobia and histrionic self-presentation scores, Spearman rank correlations were calculated as in the study by Ruch et al. (2009). In addition to the analyses in the total sample, we also calculated the correlations separately for the subgroup of participants that were able to create punch lines for each cartoon (i.e., group 6), and for the subgroup of participants that could not provide captions to all cartoons (i.e., group 1–5). This procedure is in accordance with the approach of Ruch et al. (2009) who argued that the group succeeding in generating at least one punch line for each of the six cartoons is of special interest, because one may assume that these participants are characterized by the highest humor production abilities. In addition, one may also argue that only those participants who provided at least one punch line for each of the six cartoons really completed the CPPT.

Concerning the scores that indicate quantitative humor creation abilities (see **Table 3**), the expected positive association emerged between histrionic self-presentation and the total number of punch lines (CPPT NP). The respective correlation, however, was only small and marginally significant, whereas the correlation between gelotophilia and the total number of punch lines was significant at the 0.05-level but also only small. Gelotophobia and Katagelasticism were unrelated to the total number of punch lines and none of the four traits showed associations with the number of cartoons for which a punch line was created (CPPT NC).

Both gelotophilia and histrionic self-presentation showed small correlations with the CPPT scores that indicate qualitative humor creation abilities in the total sample. Thus, high scores for gelotophilia and histrionic self-presentation were associated with more wittiness of the best punch line (CPPT WP and WPM) and a higher degree of the estimated wit and fantasy of the person. These associations, however, were clearly more pronounced and sometimes twice as high as in the total sample in the subgroup that created at least one punch line for each of the six cartoons (group 6 in **Table 3**), and non-existent within the group that only managed to generate punch lines for 1–5 cartoons (group 1–5 in **Table 3**). The respective correlations in group 1–5 and group 6 differ significantly at $p < 0.05$ for gelotophilia with CPPT WP, CPPT WPM, and CPPT WI and for histrionic self-presentation with CPPT WP and CPPT WI. In addition, the correlations with the wittiness of the best punch lines (CPPT WP and WPM) were a little bit higher for gelotophilia than for histrionic self-presentation.

TABLE 3 | Spearman rank correlations between dispositions toward laughter and ridicule, histrionic self-presentation, and the CPPT scores.

CPPT scores	*Gelotophobia*	*Gelotophilia*	*Katagelasticism*	*Histrionic SPS*
Quantitative scores				
CPPT NP	0.08	0.11*	0.01	0.10^{+}
CPPT NC	−0.06	0.04	0.02	0.03
Qualitative scores				
CPPT WP	−0.10	0.15**	0.04	0.10^{+}
Group 1–5	−0.03	0.07	0.03	−0.01
Group 6	−0.21*	0.33**	0.01	0.22*
CPPT WPM	−0.11*	0.21**	0.04	0.17**
Group 1–5	−0.05	0.12^{+}	0.05	0.14*
Group 6	−0.21**	0.33**	0.01	0.22*
CPPT WI	−0.08	0.17**	0.06	0.14*
Group 1–5	−0.03	0.07	0.06	0.05
Group 6	−0.12	0.30**	0.04	0.27**
CPPT FA	−0.09	0.18**	0.08	0.17**
Group 1–5	−0.07	0.13^{+}	0.11	0.10
Group 6	−0.08	0.25**	–0.02	0.26**

SPS = Self-Presentation Style, CPPT NP = total number of punch lines, CPPT NC = number of cartoons for which a punch line was created, CPPT WP = sum of the wittiness of the best punch line for all cartoons, CPPT WPM = mean of best punch line across cartoons and raters, CPPT WI = mean wit of the person across raters, CPPT FA = mean fantasy of the person across raters, group 1–5 = 201 participants that created punch lines for 1–5 cartoons, group 6 = 136 participants that created punch lines for each of the 6 cartoons. $^{+}p < 0.10$, $*p < 0.05$, $**p < 0.01$.

Small negative correlations were found between gelotophobia and the scores indicating qualitative humor creation abilities. However, they were only significant for the mean of the best punch line (CPPT WPM) and for the total score of the best punch line in group 6. No relations were found between katagelasticism and qualitative humor creation abilities.

DISCUSSION

This study aimed at (1) extending the nomological network of the histrionic self-presentation style by determining associations with humor creation abilities and (2) replicating the results of a former study by Ruch et al. (2009) regarding three dispositions toward ridicule and humor creation abilities. In doing so, relations between gelotophobia, gelotophilia, katagelasticism, and histrionic self-presentation with quantitative and qualitative humor creation abilities as measured in the CPPT were examined. Although the correlation coefficient was only marginally significant, the histrionic self-presentation style was associated with the total number of punch lines created across the six cartoons as hypothesized. Furthermore, histrionic self-presentation was also associated with the wittiness of the best punch lines, as well as the wit and the fantasy of the person. The associations between histrionic-self-presentation and these quantitative and qualitative humor creation abilities were, however, only small to medium. Since it is always more difficult to show associations across different data sources regarding the same or similar constructs, these small to medium correlations may be evaluated as meaningful anyway. Thus, histrionic self-presenters seem to be able to quickly generate many witty ideas that can serve as a starting point for their non-verbal As-If-performances; the witty ideas may be interpreted as a kind of raw material or potential that needs to be elaborated in a concrete interpersonal situation. This potential may be a basic idea for a histrionic role playing game. In order to perform this role playing game, however, the histrionic self-presenter needs to embody it by exhibiting non-verbal As-If-behaviors. In addition, it is also necessary to be responsive to possible interaction partners and consider situational circumstances in order to perform a successful histrionic role playing game.

The most pronounced associations between the three dispositions toward ridicule and laughter with the quantitative and qualitative humor creation abilities as assessed in the CPPT were found for gelotophilia, especially with regard to the subgroup of subjects that managed to create punch lines for each of the six cartoons. The correlations between gelotophilia and the CPPT scores indicating quantitative humor creation ability replicate the findings of the Ruch et al. (2009) study. Since the sample in the study at hand encompasses nearly three times as many subjects as in the Ruch et al. (2009) study, the small correlation between the total number of punch lines with gelotophilia was significant. The correlations between gelotophilia and qualitative humor creation abilities were comparable or higher than in the Ruch et al. (2009) study, especially regarding the six cartoons subgroup. As in the Ruch et al. (2009) study, no associations between gelotophobia and katagelasticism with quantitative humor creation ability were found. Contrary to the findings in Ruch et al. (2009), katagelasticism was unrelated to qualitative humor creation ability even in the group that created punch lines for each of the six cartoons. Thus, in our sample there seem to be only some katagelasticists who are evaluated as witty, whereas others are not. In addition, the relations between gelotophobia and qualitative humor creation ability

were slightly negative, especially in the subgroup that created punch lines for each of the six cartoons. Thus, and also in slight contrast to the study by Ruch et al. (2009), the gelotophobes in our sample were evaluated as less witty. In sum, we could only partially replicate the findings of this previous study. In our view, possible reasons for these partially different results in the study at hand are not that much our sample that was in fact bigger but showed similar socio-demographic characteristics. Also, the fact that we only used 3 raters and not 10 as Ruch et al. (2009), did not seem to have a decisive effect in terms of reliability. The most important difference between the two studies was the fact that we used a time restriction – 2.5 min for each cartoon – and Ruch et al. (2009) did not. Research has shown that creative and original solutions usually do need time (e.g., Hennessey and Amabile, 2010). Since there is a strong correlation between originality and wittiness, the time restriction may have impaired the potential for witty solutions at least in gelotophobes and katagelasticists. What argues for this interpretation is the finding that even in the groups that created punch lines for each cartoon, negative or no relations between gelotophobia and katagelasticism, respectively, with qualitative humor creation ability were found. However, this argument does not apply for gelotophiles and histrionic self-presenters, who seem to be superior regarding quantitative and qualitative humor creation abilities even under time pressure. As already argued before, histrionic self-presenters need to quickly generate witty ideas in an ongoing interaction that may be used as a kind of raw material for their histrionic role playing games.

The low to medium correlations between age and sex with the CPPT-scores show that these basic socio-demographic variables have to be taken into account when it comes to predictors of quantitative humor creation ability in particular. Men and subjects with higher age tended to create more punch lines for more cartoons in our study. This result is partly surprising because although men usually show higher humor creation ability (e.g., Mickes et al., 2012) than women, a decline in humor creation ability is assumed with age (e.g., Greengross, 2013). The probably unexpected correlations between age with the total number of punch lines as well as the number of cartoons for which a punch line was created, needs to be qualified with regard to the near-zero correlations of age with the wittiness ratings of the best punch lines. Thus, although older participants produced more punch lines, there seemed to be no significant relation of age with qualitative humor creation ability, i.e., the wittiness of the punch lines.

This study has strengths as well as weaknesses. The results are based on a comparably big and diverse sample and did not only rely on self-report measures but established, partially replicated and extended associations between a cognitive performance task regarding quantitative and qualitative humor creation ability with dispositions toward ridicule and laughter, as well as histrionic self-presentation. Although web-based studies using self-report measures are comparable with paper–pencil studies (Gosling et al., 2004), the same need not be the case with performance tests (Noyes and Garland, 2008): Our participants completed the CPPT online and thus under quite different situational conditions regarding time and place, that might have influenced their performance. Thus, it would have been preferable to administer the CPPT under the same conditions for each participant, e.g., in a big lecture hall with the entire sample. Against the background of research on creativity and originality (Hennessey and Amabile, 2010), the time limit for the generation of witty punch lines could have been a disadvantage regarding the assessment of the maximum humor performance as well. Thus, future studies should administer the CPPT under controlled conditions without time restrictions for each cartoon. In our view, more than six cartoons are not necessarily needed when there are no time limits. Another interesting research question regarding the CPPT would be to determine whether the preferences of the raters for incongruity resolution, nonsense and sexual humor do influence the evaluations of the wittiness of the punch lines. If this obvious hypothesis should be supported, it would be necessary to control for the respective humor preferences of the raters. The same could be considered regarding the raters scores on gelotophobia, gelotophilia, and katagelasticism and possible gelotophobic, gelotophilic, and katagelasticistic contents of the punch lines. Ruch et al. (2009) have already scored the punch lines created in the CPPT regarding these contents and did not find associations with the gelotophobia, gelotophilia, and katagelasticism scores of their subjects. It could, however, well be that the respective scores *of the raters* on these three dispositions toward ridicule influence the ratings of the punch lines that mirror the respective contents.

Further limitations pertain to the sampling procedure in and the gender distribution of our study. As in most psychological studies, we recruited a non-probability self-selection sample, i.e., our sample is neither probabilistic nor representative and thus generalization to whatever population is not possible. In addition, and as already pointed out at the beginning of the results section, participants self-selected to our study that was announced with an explicit hint on humor and as-if behaviors. This self-selection bias and also the fact that much more women than men participated in the study further impede generalization. Thus, further studies with more representative samples that replicate and consolidate the previous findings are necessary.

Future studies should also explore the humor performance of gelotophobes, katagelasticists, and gelotophiles in a behavior-related social interaction task. In doing so, it would be informative to combine a performance test like the CPPT with a task that requires creating humor at the behavioral and interactive level. Thus, in the first part of such a study, participants could be asked to generate as many ideas for shaping an upcoming social interaction as humorous as possible; in the second part, participants could be asked to select the best idea and to perform it together with possible interaction partners. From a self-presentational view, the ability to perform As-If-behaviors especially with gelotophiles and katagelasticists could be determined. From the point of view of dispositions toward ridicule and laughter, it could be explored whether

there are gelotophilic and katagelasticistic histrionic self-presenters that exhibit As-If-behaviors aiming at laughing at him or herself together with the audience or at laughing at the expense of others. In doing so, the interactive effects of histrionic self-presentation and gelotophilia and katagelasticism respectively should be determined; e.g., subjects that score high on both histrionic self-presentation and gelotophilia should show the most witty ideas and As-If-behaviors. In addition, the question whether gelotophobes really lack humor could be explored at the behavioral level.

CONCLUSION

This study shows that both gelotophilia and histrionic self-presentation are associated with quantitative and qualitative humor creation abilities as measured in the Cartoon Punch line production test, whereas gelotophobia and katagelasticism show slightly negative or no relations regarding humor creation abilities. These findings partly replicate and extend the results of a previous study by Ruch et al. (2009) and open up new avenues that can cross-fertilize research on individual differences in humor and self-presentation.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations and ethical guidelines of the German Psychological Society. All subjects participated anonymously and voluntarily and could quit their participation whenever they wanted without any disadvantages.

AUTHOR CONTRIBUTIONS

K-HR and LM planned the study and derived the hypotheses. LM has carried out the entire data collection. K-HR organized the rating procedure with regard to the CPPT, carried out the data analyses and did the writing of the manuscript.

ACKNOWLEDGMENTS

We would like to thank Nora-Corina Jacob and Stephanie Klee for useful comments on an earlier version of this manuscript.

REFERENCES

Arkin, R. M. (1981). "Self-presentation styles," in *Impression Management Theory and Social Psychology*, ed. J. T. Tedeschi (New York, NY: Academic Press), 311–333. doi: 10.1016/B978-0-12-685180-9.50020-8

Cooper, C. D. (2005). Just joking around? Employee humor expression as an ingratiatory behaviour. *Acad. Manag. Rev.* 30, 765–776. doi: 10.5465/AMR.2005.18378877

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.* 52, 281–302. doi: 10.1037/h0040957

Eysenck, H. J., and Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scales (EPQ Adults)*. London: Hodder and Stoughton.

Gosling, S. D., Vazire, S., Srivastava, S., and John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *Am. Psychol.* 59, 93–104. doi: 10.1037/0003-066X.59.2.93

Greengross, G. (2013). Humor and aging – a mini-review. *Gerontology* 59, 448–453. doi: 10.1159/000351005

Hennessey, B. A., and Amabile, T. M. (2010). Creativity. *Annu. Rev. Psychol.* 61, 569–598. doi: 10.1146/annurev.psych

Kamble, S. V., Proyer, R. T., and Ruch, W. (2014). Gelotophobia in India: the assessment of the fear of being laughed at with the Kanada version of the GELOPH <15> *Psychol. Stud.* 59, 337–344. doi: 10.1007/s12646-014-0254-x

Köhler, G., and Ruch, W. (1996). Sources of variance in current sense of humor inventories: how much substance, how much method variance? *Humor: International Journal of Humor Research* 9, 363–397. doi: 10.1515/humr.1996.9.3-4.363

Laux, L., and Renner, K.-H. (2002). Self-Monitoring und authentizität: die verkannten selbstdarsteller. *Z. Diff. Diagnost. Psychol.* 23, 129–148. doi: 10.1024//0170-1789.23.2.129

Mickes, L., Walker, D. E., Parris, J. L., Mankoff, R., and Christenfeld, N. J. S. (2012). Who's funny: gender stereotypes, humor production, and memory bias. *Psychon. Bull. Rev.* 19, 108–112. doi: 10.3758/s13423-011-0161-2

Norrick, N. R. (2004). Non-verbal humor and joke performance. *Humor* 17, 401–409. doi: 10.1515/humr.2004.17.4.401

Noyes, J. M., and Garland, K. J. (2008). Computer- vs. paper-based tasks: are they equivalent? *Ergonomics* 51, 1352–1375. doi: 10.1080/00140130802170387

Platt, T., Proyer, R. T., and Ruch, W. (2009). Gelotophobia and bullying: the assessment of the fear of being laughed at and its application among bullying victims. *Psychol. Sci.* 51, 135–147.

Proyer, R. T., Estoppey, S., and Ruch, W. (2012a). An initial study on how families deal with ridicule and being laughed at: parenting styles and parent-child-relations with respect to gelotophobia, gelotophilia, and katagelasticism. *J. Adult Dev.* 19, 228–237. doi: 10.1007/s10804-012-9150-6

Proyer, R. T., and Ruch, W. (2010). Enjoying and fearing laughter: personality characteristics of gelotophobes, gelotophiles, and katagelasticists. *Psychol. Test Assess. Model.* 52, 148–160.

Proyer, R. T., Ruch, W., and Chen, G.-H. (2012b). Positive psychology and the fear of being laughed at: gelotophobia and its relations to orientations to happiness and life satisfaction in Austria, China, and Switzerland. *Humor* 25, 23–40.

Proyer, R. T., Wellenzohn, S., and Ruch, W. (2014). Character and dealing with laughter: the relation of self and peer-reported strengths of character with gelotophobia, gelotophilia, and katagelasticism. *J. Psychol. Interdiscipl. Appl.* 148, 113–132. doi: 10.1080/00223980.2012.752336

Renner, K.-H. (2006). *Der histrionische Selbstdarstellungsstil. Theoretische Grundlegung und empirische Studien zu einem neuen Selbstdarstellungskonstrukt*. Bamberg: Habilitationsschrift.

Renner, K.-H., Enz, S., Friedel, H., Merzbacher, G., and Laux, L. (2008). Doing as if: the histrionic self-presentation style. *J. Res. Pers.* 42, 1303–1322. doi: 10.1016/j.jrp.2008.04.005

Renner, K. H., and Heydasch, T. (2010). Performing humor: on the relations between self-presentation styles, gelotophobia, gelotophilia, and katagelasticism. *Psychol. Test Assessm. Model.* 52, 171–190.

Rosenfeld, P., Gialcone, R. A., and Tedeschi, J. T. (1983). Humor and impression management. *J. Soc. Psychol.* 121, 59–63. doi: 10.1080/00224545.1983.9924467

Ruch, W. (1992). Assessment of appreciation of humor: studies with the 3 WD humor test. *Adv. Personal. Assess.* 9, 27–75. doi: 10.5167/uzh-77946

Ruch, W., Beermann, U., and Proyer, R. T. (2009). Investigating the humor of gelotophobes: does feeling ridiculous equal being humorless? *Humor* 22, 111–143. doi: 10.1515/HUMR.2009.006

Ruch, W., Harzer, C., and Proyer, R. T. (2013). Beyond being timid, witty, and cynical: big five personality characteristics of gelotophobes, gelotophiles, and katagelasticists. *Int. Stud. Humor* 2, 24–42.

Ruch, W., and Heintz, S. (forthcoming). "Humor production and creativity: overview and recommendations," in *Creativity and Humor*, eds S. R. Luria and J. C. Kaufman (San Diego, CA: Elsevier Academic Press).

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. T. (2014). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., and Proyer, R. T. (2009a). Extending the study of gelotophobia: on gelotophiles and katagelasticists. *Humor* 22, 183–212. doi: 10.1515/HUMR.2009.009

Ruch, W., and Proyer, R. T. (2009b). The fear of being laughed at: Individual and group differences in gelotophobia. *Humor* 21, 47–67.

Snyder, M. (1974). The self-monitoring of expressive behavior. *J. Pers. Soc. Psychol.* 30, 526–537. doi: 10.1037/h0037039

Turner, R. G. (1980). Self-monitoring and humor production. *J. Pers.* 48, 163–172. doi: 10.1111/j.1467-6494.1980.tb00825.x

Wolfe, R., Lennox, R., and Cutler, B. (1986). Getting along and getting ahead: empirical support for a theory of protective and acquisitive self-presentation. *J. Pers. Soc. Psychol.* 50, 256–361. doi: 10.1037/0022-3514.50.2.356

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Renner and Manthey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 14 August 2018
doi: 10.3389/fpsyg.2018.01447

Fear of Being Laughed at in Children and Adolescents: Exploring the Importance of Overweight, Underweight, and Teasing

Carl-Walter Kohlmann[1], Heike Eschenbeck[1], Uwe Heim-Dreger[1], Michael Hock[2], Tracey Platt[3] and Willibald Ruch[4]*

[1] Psychology, University of Education Schwäbisch Gmünd, Schwäbisch Gmünd, Germany, [2] Psychology, University of Bamberg, Bamberg, Germany, [3] Institute of Sport and Human Sciences, University of Wolverhampton, Wolverhampton, United Kingdom, [4] Department of Psychology, University of Zurich, Zurich, Switzerland

Weight bias toward obese youths is often accompanied by the experience of psychological stress in those affected. Therefore, the fear of being laughed at (i.e., gelotophobia) in overweight children and adolescents can be rather serious. In four explorative studies, the importance of relative weight, self-awareness of weight (incl. satisfaction with weight), experiences of teasing and ridicule, as well as the role of social-evaluative situations in school were analyzed with regard to gelotophobia. In two online interviews of adults with pronounced gelotophobia (Study I: 102 English-speaking participants, Study II: 22 German-speaking participants) relating to reasons they assumed for their development of gelotophobia, there was evidence of injurious appearance-related experiences during childhood and adolescence. In Study III (75 Swiss adolescents) associations between the experience of weight-related teasing and mockery with overweight, self-perceptions of weight, and gelotophobia were analyzed. Especially in girls, overweight was associated with the experience of weight-related teasing and ridicule, which in turn was accompanied by gelotophobia. Study IV included 178 German adolescents who were asked to report their body image ("Do you think you are. . . too thin, just the right weight, or too fat?"). In addition, gelotophobia, teasing, BMI based on self-reports, and joy at school were measured. In particular, girls who felt too fat and boys who felt too thin reported teasing. Teasing was related to diminished joy at school and to gelotophobia. Among boys, underweight mediated by weight-related teasing contributed to gelotophobia. The results suggest that more research should be devoted to gelotophobia and the experience of weight-related teasing and mocking to better understand factors contributing to the well-being of children and adolescents with weight problems.

Keywords: gelotophobia, teasing, victimization, overweight, underweight, body image, well-being

OPEN ACCESS

Edited by:
Marina A. Pavlova, Universität Tübingen, Germany

Reviewed by:
Michele Guerreschi, Independent Researcher, Brescia, Italy
Hugo Carretero-Dios, Universidad de Granada, Spain

****Correspondence:***
Carl-Walter Kohlmann
carl-walter.kohlmann@ph-gmuend.de

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *23 January 2018*
Accepted: *23 July 2018*
Published: *14 August 2018*

Citation:
Kohlmann C-W, Eschenbeck H, Heim-Dreger U, Hock M, Platt T and Ruch W (2018) Fear of Being Laughed at in Children and Adolescents: Exploring the Importance of Overweight, Underweight, and Teasing. Front. Psychol. 9:1447. doi: 10.3389/fpsyg.2018.01447

INTRODUCTION

The aim of the present study is to explore the role of the fear of being laughed at (i.e., gelotophobia) in the context of overweight, teasing and well-being in children and adolescents.

Gelotophobia describes the fear of being an object of laughter (for an overview, see Ruch et al., 2014). Gelotophobia is conceptualized as an individual difference variable considerably varying

in non-clinical samples (Ruch and Proyer, 2008). Individuals high in gelotophobia fear exposing themselves to others because they expect that others screen them for evidence of ridiculousness. Experiences of teasing and victimization (including being a target of destructive humor) during childhood and adolescence seem to play a role in the development of gelotophobia (Ruch et al., 2010).

Being teased about one's weight is common among overweight and obese adolescents (Puhl and Latner, 2007; Haines et al., 2013). Obesity is associated with impaired well-being and mental health in children and adolescents (Eschenbeck et al., 2009; Griffiths et al., 2010). Weight bias toward obese youths seems to play a central role in the association between overweight and emotional well-being (Eisenberg et al., 2003). Weight-based teasing from peers and parents in adolescence can even result in weight gain, unhealthy weight control and eating to cope 15 years later (Puhl et al., 2017).

Studies have demonstrated that weight bias begins early in childhood. It becomes worse as children get older (Puhl and Latner, 2007). Puhl et al. (2008) applied an online interview with a sample of overweight and obese adults to identify and describe their subjective experiences of weight bias. When asked to describe their worst experiences of weight stigmatization participants reported experiencing weight stigma across a wide range of contexts and involving a variety of interpersonal sources. Most of the participants (77%) reported verbal bias, which included intentional negative comments, insults, derogatory names, teasing, ridicule or being made fun of because their weight. Emotional consequences of weight stigmatization were not assessed directly. However, when asked about what others should know about what it is like to be overweight besides mostly weight-based responses (e.g., difficulty of weight loss, physical challenges of excess weight) also emotional aspects have been reported (e.g., depressive feelings, feelings hurt, humiliation, embarrassment, sadness, and sorrow) in 36% of the interviews.

For individuals with pronounced gelotophobia, teasing experiences are documented as well. For example, adults high in gelotophobia reported that they have been teased quiet often during school time, that teachers made fun of them during lessons, and that they felt punished by their parents by means of ironic and sarcastic comments (Ruch et al., 2010). Edwards et al. (2010) found that among undergraduate students gelotophobia was positively correlated with a history of being teased. Teasing domains included social behavior, academic excellence, performance, and appearance. The appearance scale was comprised of several aspects of appearance (e.g., color or style of hair, wore glasses, weight, fatter than other kids). Therefore, the fear of being laughed at (i.e., gelotophobia) particularly in overweight children and adolescents can be rather serious. However, research combining the perspectives on teasing and victimization in overweight and obese children and adolescents on the one hand and teasing and victimization as related to gelotophobia on the other hand, has not yet received systematic attention.

To study the role of gelotophobia in relation to overweight, teasing, and emotional well-being in children and adolescents, two different perspectives were combined: do adults with gelotophobia remember weight-related teasing in younger age? Is weight-related teasing a mediator between weight status and gelotophobia in children and adolescents?

The following research questions will be addressed: (1) Are physical appearance and weight-related teasing and ridicule in childhood and adolescence among the assumed reasons for being laughed at in the memories of gelotophobic adults? (2) Is the experience of weight-based teasing a mediator between overweight and gelotophobia in adolescents? (3) Are gelotophobia and experiences of teasing associated with reduced joy at school, especially in social contexts (i.e., in the schoolyard)?

STUDY I

In order to explore the association between gelotophobia, overweight, weight-related teasing and victimization. It is of interest to study whether gelotophobes report physical appearance (including weight-related aspects of appearance) as assumed reasons for being laughed at. In addition, the source of victimization and onset of stigma are of interest. It is hypothesized that striking features in appearance are the predominantly assumed reasons for being laughed at. For gelotophobes, laughter-related experiences in childhood and youth, but not in adulthood, are supposed to be highly relevant. Peers should therefore play the central role as interpersonal source of victimization experiences.

Method

Procedure and Participants

Participants were recruited by Internet contact. Media coverage of feature stories on gelotophobia were utilized to recruit participants by providing a URL that directed interested people to a website. Participants were invited to complete questionnaires on the site. No personal identifying information was taken (for details, see Platt et al., 2012).

Gelotophobia was assessed by the GELOPH<15> (Ruch and Proyer, 2008). The questionnaire for the assessment of gelotophobia is comprised of 15 items (e.g., "When they laugh in my presence I get suspicious."). Items are positively keyed and responses are given on a four-point Likert scale (1 = strongly disagree; 2 = moderately disagree; 3 = moderately agree; 4 = strongly agree). The GELOPH<15> was used in previous studies and proved to be a reliable and valid instrument for the assessment of gelotophobia (Ruch and Proyer, 2008; Proyer et al., 2009). In the present sample, the GELOPH<15> proved to be reliable, yielding a high internal consistency (Cronbach's $\alpha = 0.91$; see Platt et al., 2012).

Of the original English-speaking sample of 622 adults (Platt et al., 2012, Sample I), a sub-sample of 113 participants reported extreme fear of being laughed at ($M \geq 3.5$; Ruch, 2009; Ruch et al., 2014). Interview data were available for 102 of these participants (63 men, 39 women; mainly from United States, United Kingdom, Australia, India, and Canada; age varied between 18 and 63 years, $M = 26.70$, $SD = 12.01$). Civil status of the participants was distributed as follows: single ($n = 72$), married ($n = 16$), cohabiting ($n = 11$), separated ($n = 2$), and widowed ($n = 1$).

The *Structured Gelotophobia Interview—Written Experimental Form* (Platt and Ruch, 2007, Unpublished; see also Platt et al., 2012) contains a list of 20 questions relating to a variety of issues regarding the onset of the fear of being laughed at, typical ways of dealing with it, thoughts, emotions and actions while being laughed at, bullying experience as well as socio-demographic variables. The interview was administered in a written format. For the present study, only certain aspects of the interview have been analyzed.

Data Coding and Analysis

The written responses to the questions submitted online were coded using a stage model of qualitative content analysis (Berg, 2004). After reading the responses for content, primary analytic categories were identified. A coding template was developed based on these responses to establish variables and categories and determine criteria for selection and sorting of content into variables and categories. For the present analysis, only three variables of the interview were analyzed: assumed reason for being laughed at (*assumed reason*), the remembered period of life or age (*onset*), and the interpersonal source of threat (*source*) when the first victimization experience associated with the onset of the fear of being laughed at occurred.

Categories for *assumed reason* were social behavior (e.g., social interaction, having eye contact, having seemed weird to classmates, awkward moments, being over-reactive, grieving for a loved one), physical appearance, tribal [(i.e., ethnic background, religion), residual, and not evident from response; for stigma categories see Goffman, 1963]. The physical appearance category included three subcategories: physical stable and overweight-related (e.g., overweight, fat, obesity, and chubby), physical stable and not overweight-related (e.g., bad skin, big lips, red hair, and being ugly), and physical flexible (e.g., clothes, got glasses). Categories for *onset* (for time period of stigma, see Puhl et al., 2008) were childhood (including early childhood), elementary school, middle school, high school, adulthood, and not evident from response. *Source* categories were peers, family (i.e., brother, siblings, mother, and father), others (teacher, other person), peers and family, peers and others, and not evident from response.

All responses were double coded by the first author and a student research assistant[1], and inter-rater agreement (Cohen, 1960) was calculated. With all coefficients $\kappa > 0.80$, a high agreement between the two coders was achieved.

Results

Descriptive statistics are presented in **Table 1**. Among the *assumed reasons* for being laughed at the categories social behavior (n = 22, 22%), physical appearance (n = 18, 18%), and not evident from response (n = 58, 57%) were coded most often. The latter category plus tribal (n = 1) and residual (n = 3) were summarized as the new aggregated category residual (n = 62). Among the physical appearance category the sub-category physical stable and overweight-related was reported by six participants (i.e., 33% of the physical appearance category and 6% of all assumed reason categories).

Onset of the fear of being laughed at had its maximum in elementary school (n = 38, 37%) and its minimum in adulthood (n = 1, 1%). For the frequencies for the other onset categories see **Table 1**. For further analyses aggregated categories of onset were computed: childhood and elementary school (n = 50), middle school and high school (n = 29), and residual (n = 23) which was comprised of adulthood (n = 1) and not evident from response (n = 22).

For almost half of the participants (n = 49, 48%) interpersonal *source* of victimization behavior was not evident from response. From the remaining responses, the majority of participants reported that their victimization experiences were enacted by peers (n = 39, 38% from total sample), followed by family members (n = 6, 6%), or a combination of peers and family members (n = 6, 6%). All the responses involving peers combined (i.e., peers, peers and family, peers, and others) resulted in total of reported 46 victimization behaviors (45%) related to peers. An aggregated category for source was not computed.

Separate cross tabulations of gender with *assumed reason* and *onset* were performed. χ^2 tests (ps > 0.23) did not show any evidence of an association of gender of participants with the two dimensions.

A cross tabulation of *assumed reason* and *onset* indicated that the distribution of the two interview variables were not independent from each other (see **Table 2**); $\chi^2(4, N = 102) = 13.95$, $p < 0.01$. Whereas social behavior as the assumed reason for being victimized was not related to the onset categories, physical appearance was especially an assumed reason for being victimized during middle school and high school (n_o = 10, n_e = 5.1, z = 2.8) but not in childhood and elementary school (n_o = 7, n_e = 8.8, z = 0.9). Results for the residual categories (e.g., overrepresentation of the onset category residual within the assumed reason category residual) will not be interpreted given the heterogeneous nature of both categories.

To summarize, findings showed that among extreme gelotophobic individuals social behavior and physical appearance were reported as assumed reasons for being laughed at. For physical appearance this was especially prominent during the age when attending middle or high school.

Discussion

For individuals with extreme gelotophobia, social behavior and physical appearance were the alleged main reasons for laughter. Weight-related aspects of appearance accounted for one-third of this category. It has also been shown that severe experiences related to the development of gelotophobia for childhood and schooling are reported. Peers seem to be the primary interpersonal source of remembered threat. Therefore, weight-related teasing and mockery by peers in childhood and youth are at least a remarkable aspect in the broader understanding of gelotophobia development.

Findings of the present study are compatible with findings from gelotophobia research. In a sample of 6 to 9-year-olds gelotophobia was positively related with victim status (Proyer et al., 2012). Führ (2010) found that children and adolescents who

[1]We would like to thank Andreas Vuori for his valuable support in coding the *Structured Gelotophobia Interview—Written Experimental Form* in Study I.

TABLE 1 | Variables and categories derived from the Structured Gelotophobia Interview (Study I).

Variable	Description of variable	Categories	Aggregated categories (if applicable)
Assumed reason	Assumed reason for being laughed at		
		Social behavior (*n* = 22)	Social behavior (*n* = 22)
		Physical appearance (*n* = 18)*	Physical appearance (*n* = 18)
		Tribal (*n* = 1)	Residual (*n* = 62)
		Residual (*n* = 3)	
		Not evident from response (*n* = 58)	
Onset	Onset of fear of being laughed at		
		Childhood (*n* = 12) Elementary school (*n* = 38)	Childhood and elementary school (*n* = 50)
		Middle school (*n* = 20) High school (*n* = 9)	Middle school and high school (*n* = 29)
		Adulthood (*n* = 1)	Residual (*n* = 23)
		Not evident from response (*n* = 22)	
Source	Interpersonal source of threat		
		Peers (*n* = 39)	
		Family (*n* = 6)	
		Others (*n* = 1)	
		Peers and family (*n* = 6)	
		Peers and others (*n* = 1)	
		Not evident from response (*n* = 49)	

*N = 102. *The physical appearance category was comprised of the sub-categories physical stable and overweight-related (n = 6, 33%), physical stable and not overweight-related (n = 9, 50%), and physical flexible (n = 3, 17%).*

TABLE 2 | Cross tabulation of assumed reason for being laughed at by onset of fear of being laughed at (Study I): observed frequencies and expected frequencies (in parentheses).

	Onset			
Assumed reason	**Childhood and elementary school**	**Middle school and high school**	**Residual**	**Total**
Social behavior	14 (10.8)	6 (6.3)	2 (5.0)	22
Physical appearance	7 (8.8)	10* (5.1)	1 (4.1)	18
Residual	29 (30.4)	13* (17.6)	20* (14.0)	62
Total	50	29	23	102

$p < 0.05$ (corrected standardized residuals > |1.96|).

reported having been a victim of bullying expressed higher levels of the fear of being laughed at. However, these studies did not have a particular focus on weight-related victimization. A study among undergraduate students (Edwards et al., 2010) revealed that higher gelotophobia was associated with memories of being teased about social behavior, performance, academic excellence, and appearance (including the item "I was teased about my weight"). However, the strongest association of gelotophobia was found for teasing related to social behavior.

A limitation of the present study is that in about half of the cases, the interpersonal source of threat and the suspected reason for being laughed at were not apparent from the transcribed interviews. Unfortunately, weight status of participants was also not assessed in the original Structured Gelotophobia Interview. Therefore, a second interview study was conducted to overcome these limitations.

STUDY II

Aim of the second study was to test whether main findings of Study I can be replicated in a more structured assessment. It was hypothesized that even in a more structured online interview, individuals high in gelotophobia cite aspects of their physical appearance, including weight aspects, in addition to their conspicuous social behavior as assumed reasons for being laughed at. Furthermore, peers are supposed to be the main source of teasing and mockery remembered, with most of the victimization experiences starting in childhood and schooling.

Method

Procedure and Participants

A German-speaking sample of 35 adults (14 men, 21 women) was recruited by Internet contact based on a newspaper article on gelotophobia ("Hoffentlich lacht keiner", [Hoping nobody is laughing], Frankfurter Allgemeine Sonntagszeitung, published February 19, 2012). Gelotophobia was assessed by the GELOPH<15> (Ruch and Proyer, 2008; Cronbach's $\alpha = 0.83$ in the present sample).

Besides demographics (gender, age, and civil status), self-reports of body weight and height were assessed. Of this sample, 63% ($n = 22$) reported at least substantial fear of being laughed at ($M > 2.5$; slight fear [M between 2.5 and 3.0], $n = 13$; marked fear

[M between 3.0 and 3.5, $n = 8$; extreme fear [$M \geq 3.5$], $n = 1$; for criteria see Ruch, 2009). Data of these 22 participants (10 men, 12 women; 21 Germans, 1 Italian) were used for further analyses. Age varied between 17 and 66 years ($M = 44.91$, $SD = 14.06$). Civil status of the participants was distributed as follows: single ($n = 7$), married ($n = 11$), cohabiting ($n = 4$). Self-reports of weight and height were given by 20 participants (10 men, 10 women). BMI varied between 17.21 and 42.98 ($M = 25.04$, $SD = 6.02$). Distribution of weight categories were as follows: underweight ($n = 1$ woman), normal weight ($n = 11$; 5 men, 6 women), overweight ($n = 6$; 5 men, 1 woman), and obese ($n = 2$ women).

A slightly modified version of the *Structured Gelotophobia Interview—Written Experimental Form* (Platt and Ruch, 2007, Unpublished; see Study I) was administered in a written format with fixed response categories. Questions included the three variables from Study I: assumed reason for being laughed at (*assumed reason*), the remembered period of life or age (*onset*), and the interpersonal source of threat (*source*) when the first victimization experience associated with the onset of the fear of being laughed at occurred. Categories for *assumed reason* (multiple responses possible) were social behavior, physical appearance, mistake, other, and without reason. Categories for *onset* (only one response possible) were early childhood and preschool (age: 0–5 years), elementary school (6–10 years), middle school (11–15 years), high school (16–20 years), and adulthood (sub-categories were 20–30 years, 30–40 years, . . ., 60–70 years, >70 years). Categories were based on the German school system. *Source* categories (multiple responses possible) were schoolmates, other peers, family members, friends, teachers, and other adults.

Results

Descriptive statistics are presented in **Table 3**. Among the *assumed reasons* for being laughed at (with multiple responses allowed) the category of social behavior was reported most frequently ($n = 9$, 41%). Physical appearance was mentioned by five individuals (23%, see **Table 3** for the other categories). *Onset* of victimization experiences related to the fear of being laughed at had its maximum in early childhood and preschool ($n = 10$, 45%), followed by elementary school ($n = 7$, 32%) and middle school ($n = 5$, 23%). High school and adulthood were not mentioned. To compare findings for *onset* with those from Study 1, we additionally aggregated categories of *onset*: this yielded *n*s of 17 (77%) for childhood and elementary school, 5 (23%) for middle school and high school, and 0 for adulthood. As interpersonal *sources* of victimization behavior (with multiple responses allowed) participants most often reported schoolmates ($n = 15$, 68%) and other peers ($n = 9$, 41%), family members ($n = 12$, 55%), teachers ($n = 9$, 41%), and other adults ($n = 7$, 32%; see **Table 3** for the other categories).

In contrast to Study I, the gelotophobia interview variables assumed reason and source were based on a multiple response format. Therefore, associations of these variables with gender and weight status were analyzed separately for each category. Associations of *assumed reason, onset,* and *source* with gender ($N = 22$) and weight status ($N = 20$ due to missing values) were generally low (χ^2 tests and Fisher's exact tests, ps > 0.10). There was only a significant association between gender and *onset*; Fisher's exact test, $p < 0.05$. More women than men reported an early onset for the first victimization experience associated with the fear of being laughed at: early childhood and preschool (men: $n = 2$, women: $n = 8$), elementary school or middle school (men: $n = 8$, women: $n = 4$).

Findings showed that for early childhood and the period attending elementary school, first victimization experiences as assumed reasons for the onset of gelotophobia were reported. Schoolmates and other peers as well as family members and teachers were the main *sources* of victimization. Social behavior was more predominant than physical appearance among the most often reported reasons assumed for being laughed at.

Discussion

In a structured interview, people with gelotophobia also report that they were teased and mocked mainly because of their social behavior but also because of their appearance. These unpleasant experiences began mainly in childhood and during elementary school. In these two points, Study II confirms the results of Study I, although the criterion for the diagnosis of gelotophobia was not as strict as in Study I. However, a striking difference to Study I resulted for the reported source of teasing and harassment. Although in both studies peers and classmates are the main attackers, in Study II family members and teachers were also mentioned remarkably often.

It cannot be ruled out that the greater importance of family members and teachers as sources of teasing in Study I compared to Study II is related to the fact that the samples come from different cultural backgrounds or that the average age of the interviewees in Study II was significantly higher (around 45 years) than in Study I (27 years). Davies (2009) assumes that pressure to conform and maintain harmony and the existence and maintenance of hierarchies are important social variables related to victimization and the development of gelotophobia. The differences in the role of family members and teachers between the two studies may be due to the fact that there may have been a shift from one generation to another in the social factors mentioned by Davies.

Weight status was independent of the variables derived from the interviews. Although almost a quarter of all respondents cited physical appearance as a reason for laughing, the small sample size (22 participants, eight of them were overweight or obese) was far from being ideal for investigating associations between weight status and specific forms of teasing.

To summarize, both interview studies showed that physical appearance was reported by at least some gelotophobic adults as an assumed reason for being laughed at in childhood and adolescence. These preliminary findings, however, cannot support the assumption that the weight status of children and adolescents (i.e., overweight and obesity) may be associated with gelotophobia mediated by weight-related teasing. To test this assumption more directly, Studies III and IV were conducted with students who were not pre-selected according to their degree of gelotophobia.

TABLE 3 | Variables and categories derived from the modified Structured Gelotophobia Interview (Study II).

Variable	Description of variable	Categories	Aggregated categories (if applicable)
Assumed reason*	Assumed reason for being laughed at		
		Social behavior (*n* = 9)	
		Physical appearance (*n* = 5)	
		Mistake (*n* = 4)	
		Other (*n* = 6)	
		Without reason (*n* = 8)	
Onset	Onset of fear of being laughed at		
		Early childhood and preschool (*n* = 10) Elementary school (*n* = 7)	Childhood and elementary school (*n* = 17)
		Middle school (*n* = 5) High school (*n* = 0)	Middle school and high school (*n* = 5)
		Adulthood (*n* = 0)	Adulthood (*n* = 0)
Source*	Interpersonal source of threat		
		Schoolmates (*n* = 15)	
		Other peers (*n* = 9)	
		Family (*n* = 12)	
		Friends (*n* = 4)	
		Teachers (*n* = 9)	
		Other adults (*n* = 7)	
		Unknown (*n* = 0)	

*N = 22. *Multiple responses possible. Therefore, no aggregated categories were computed for this variable.*

STUDY III

Overweight and obese children and adolescents experience more weight-based teasing and victimization than children and adolescents with normal weight (Puhl and Latner, 2007). Research suggests that weight status is predictive of vulnerability to bullying in peer relationships (Neumark-Sztainer et al., 2002). For example, the prevalence of weight-based teasing by peers was significantly higher among overweight and obese youth (45%) than among normal weight youth (22%; Goldfield et al., 2010). Overweight and obese girls experienced even more weight-related teasing by peers than overweight and obese boys (52% vs. 30%). In addition, teasing about body weight was associated with impaired well-being (anxiety, depression, and negative self-esteem). Among overweight children and adolescents, appearance-related teasing seems to be prevalent, frequent, upsetting, and focusing more on weight than on less stigmatized aspects of appearance (Hayden-Wade et al., 2005).

The present study aimed at analyzing the relationship between weight status, weight-based victimization and gelotophobia among adolescent boys and girls. The main research question is whether weight-related teasing mediates the relationship between weight status (e.g., being overweight or obese) and gelotophobia (see **Figure 1**).

In addition, the association of weight status, teasing and gelotophobia with the adolescents' body image will be explored. In a representative German sample of 3,254 girls and 3,415 boys aged 11–17 years, a high proportion of normal-weighted girls and boys had a negative overweight-associated body image (i.e., "I think I'm a bit too fat" or ". . . far too fat"; 49% of the girls compared to 26% of the boys; Kurth and Ellert, 2008), which was associated with reduced well-being. On the other hand, overweight adolescents who perceive their weight about right report a better psychological and physical health than those with realistic self-perceptions (Fuchs et al., 2012). Therefore, not only the association of weight status but also the association of body image with weight-related teasing and gelotophobia is of interest.

Method

Participants and Procedure

Participants were 75 adolescents (boys: n = 23, girls: n = 52; age: M = 13.97 years, SD = 1.08 years, Range = 12–16 years), recruited from Swiss middle schools in the Cantons of Zurich and Aargau. Parental and participant consent was required to participate in the study. All data (i.e., demographic information, reports of weight and height, questionnaires) were assessed online (Unipark).

Variables and Questionnaires

Gelotophobia was assessed by the GELOPH<15>(Ruch and Proyer, 2008; see Study I; Cronbach's α = 0.87 in the present sample).

Teasing

The Teasing Questionnaire-Revised (Strawser et al., 2005; see also Storch et al., 2004) is a 29-item self-report scale designed to measure teasing experiences. Responses are given on five-point Likert-type scale (0 = "I was never teased about this," 1 = "I was rarely teased about this," 2 = "I was sometimes teased about this," 3 = "I was often teased about this," and 4 = "I was always teased about this"). Domains considered are performance (e.g., "not good at sports"), academic excellence (e.g., "not 'nerdy"'), social behavior (e.g., "often looked nervous"), family background (e.g., "I had a 'funny' name"), and appearance (e.g., "fatter than other

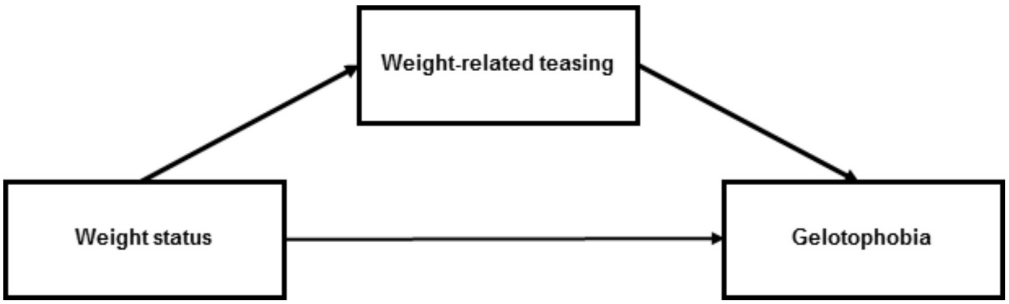


FIGURE 1 | Proposed model: Weight-related teasing as mediator between weight status and gelotophobia.

kids," "color or style of hair"). The *teasing (total)* scale showed a good reliability with Cronbach's α = 0.92 in the present sample.

Based on the literature on the main topics of weight-related teasing (Puhl and Latner, 2007), an additional 6-item subscale on *weight-related teasing* was composed for the present study. One item of the performance factor ("not good at sports") and five items of the appearance factor ("aspects of my appearance," "being ugly/unattractive," "weight," "way that I dressed," "fatter than other kids") were used (Cronbach's α = 0.88; for similar scales for the assessment of weight-related teasing, see Thompson et al., 1995; Gros et al., 2012).

Weight categories

Self-reports of weight and height were assessed to compute BMI scores. Based on age- and gender-specific norms, participants were allocated to weight categories (Kromeyer-Hauschild et al., 2001; see also Cole et al., 2000; Stamm et al., 2010). The original five categories were reduced to three categories by combining the two underweight categories (i.e., extremely underweight: $n = 4$; underweight: $n = 9$) as well as the overweight ($n = 5$) and the obesity category ($n = 1$) in one category each, resulting in the weight categories underweight ($n = 13$; 4 boys, 9 girls), normal weight ($n = 56$; 18 boys, 38 girls) and overweight ($n = 6$; 1 boy, 5 girls).

Body image

Participants were asked to evaluate their own weight using a single item: "Do you think you are . . . (1) far too thin, (2) a bit too thin, (3) just the right weight, (4) a bit too fat, (5) far too fat?" (Kurth and Ellert, 2008). The original five categories were reduced to three categories: 1 ($n = 2$) and 2 ($n = 14$) were combined into "too thin" and 4 ($n = 30$) and 5 ($n = 0$) into "too fat," resulting in the following distribution of body image: "too thin" ($n = 16$; 5 boys, 11 girls), "just the right weight" ($n = 29$, 9 boys, 20 girls), and "too fat" ($n = 30$, 9 boys, 21 girls).

In addition, dummy-coded variables for weight categories (underweight: yes = 1, no = 0; overweight: yes = 1, no = 0) and body image ("too thin": yes = 1, no = 0; "too fat": yes = 1, no = 0) were computed.

Data Analysis

The correspondence between weight status and body image was evaluated by Cohen's kappa (Cohen, 1960). For an initial examination, descriptive statistics and correlations among gelotophobia and teasing and with weight status and body image were computed separately for boys and girls. As a second step, to examine whether there will be an indirect association between overweight and gelotophobia mediated by weight-related teasing, path analyses were computed (Preacher and Hayes, 2004; Hayes, 2013). Mediation analyses were conducted with a SPSS macro using bootstrapping with $z = 5{,}000$ resamples to compute 95% confidence intervals for the indirect effect.

Results

Weight categories and body image show a low but significant association (Cohen's $\kappa = 0.25$, $p < 0.001$; 95% confidence interval: 0.08–0.43; see **Table 4**). Separate analyses for male and female adolescents did not result in different associations (κs = 0.25 for boys and girls, resp.). However, among the normal weight group a substantial number of adolescents (23 out of 56, i.e., 41%) perceive themselves as being "too fat."

Correlations between gelotophobia, teasing (total) and weight-related teasing as well as the correlations of these variables with weight categories and body image are presented in **Table 5**. Gelotophobia was positively correlated with teasing (total) in girls ($r = 0.42$, $p < 0.01$) and in boys ($r = 0.41$, $p = 0.06$). Weight-related teasing, however, was associated with gelotophobia for girls only ($r = 0.40$, $p < 0.01$). Gelotophobia did not show significant associations with weight categories or body image (e.g., correlation between gelotophobia and underweight in boys: $r = 0.25$, $p = 0.26$; correlation between gelotophobia and overweight in girls: $r = 0.21$, $p = 0.14$). In girls, overweight was

TABLE 4 | Cross tabulation of weight categories and body image (Study III): observed frequencies.

Weight categories	Body image			
	"Too thin"	"Just right"	"Too fat"	Total
Underweight	**8**	4	1	13
Normal weight	8	**25**	23	56
Overweight	0	0	**6**	6
Total	16	29	30	75

"Just right" = "just the right weight". Cells with an expected agreement between weight categories and body image are printed in bold. κ = 0.25, p < 0.001.

TABLE 5 | Correlations of gelotophobia and teasing with weight categories and body image and joy at school (Study III).

				Variable		
Variable		***M***	***SD***	**Gelotophobia**	**Teasing (total)**	**Weight-related teasing**
Gelotophobia	Boys	2.06	0.48	1.00		
	Girls	2.19	0.58	1.00		
Teasing (total)	Boys	0.41	0.50	0.41^{+}	1.00	
	Girls	0.40	0.40	0.42**	1.00	
Weight-related teasing	Boys	0.49	0.61	0.18	0.78**	1.00
	Girls	0.55	0.79	0.40**	0.82**	1.00
Underweight	Boys	0.17	0.39	0.25	−0.06	−0.22
	Girls	0.17	0.38	−0.11	0.10	−0.02
Overweight	Boys	0.04	0.21	−0.03	−0.12	0.00
	Girls	0.09	0.30	0.21	0.32*	0.41**
Body image "too thin"	Boys	0.22	0.42	−0.10	−0.12	−0.17
	Girls	0.21	0.41	−0.07	−0.09	−0.15
Body image "too fat"	Boys	0.39	0.50	−0.03	−0.33	−0.07
	Girls	0.40	0.50	0.17	0.04	0.24

Boys: n = 23, girls: n = 52. Gelotophobia and teasing scores are reported as item means (sum scores divided by number of items; Gelotophobia, range: 1–4; Teasing, range: 0–4, Joy, range: 1–4. Weight categories and body image are dummy-coded (1 = yes, 0 = no). $^{+}p < 0.06$, $*p < 0.05$, $**p < 0.01$ (two-tailed).

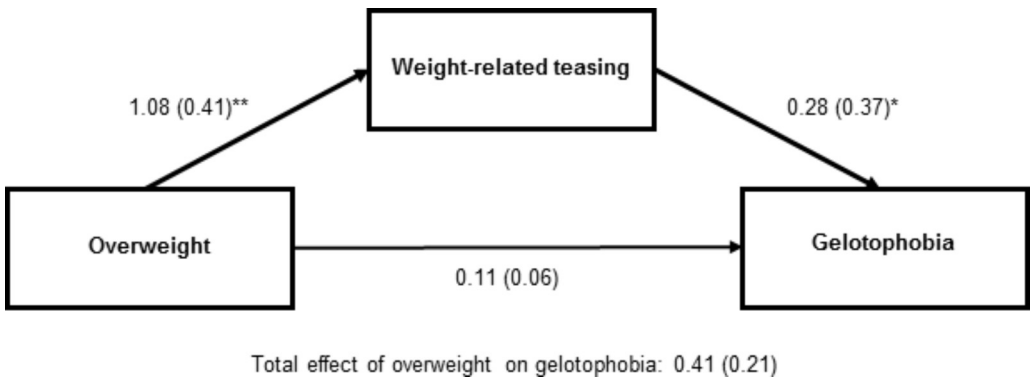


FIGURE 2 | Weight-related teasing as mediator between overweight and gelotophobia in girls ($N = 52$, Study III). Unstandardized regressions coefficients and standardized regression coefficients (in parentheses) are based on item mean scores for weight-related teasing and gelotophobia, respectively. Overweight categories: 0 = no, 1 = yes. $*p < 0.05$, $**p < 0.01$.

significantly associated with teasing (total, $r = 0.32$, $p < 0.05$) as well as weight-related teasing ($r = 0.41$, $p < 0.01$).

In the tested mediation model the independent variable was overweight (dummy-coded), mediator was weight-related teasing, and the dependent variable was gelotophobia. Only among girls the indirect effect was significant. The 95% CI obtained for the indirect effect of overweight status on gelotophobia by bootstrapping was 0.23 (CFI: 0.03 to 0.63) and did not include 0 (z = 5,000 bootstrap resamples), indicating an indirect-only mediation (Zhao et al., 2010) since the direct path was not significant[2]. The results for the interplay between overweight status, weight-related teasing, and gelotophobia in girls are shown in **Figure 2**. Weight-related teasing was higher in overweight than in non-overweight adolescent girls. Furthermore, gelotophobia was higher with enhanced weight-related teasing. Overweight status was not directly related to gelotophobia. However, overweight status and gelotophobia were indirectly associated by weight-related teasing.

Discussion

Weight status and body image showed a significant but small association. Especially the high proportion of normal weight adolescents with an unrealistic overweight-related body image was in accordance with a previous study by Kurth and Ellert (2008). In some contrast to their study, however, is the finding that in the present study body image was unrelated to gelotophobia. According to Kurth and Ellert (2008) body image seems to be crucial for psychological well-being. Given the conception of gelotophobia as a shame-bound anxiety, this could have been expected. However, gelotophobia cannot be fully explained by negative affect and is also sufficiently different

[2]Similar findings resulted when the total sample ($N = 75$) with both boys and girls or teasing (total) as the mediator were analyzed.

from social anxiety (Ruch et al., 2014). To further explore the association between body image and gelotophobia, a larger sample size seems to be needed.

Findings of the present study give some support for the assumption that weight-related teasing may play a role in gelotophobia, at least in girls. Overweight adolescent girls reported increased weight-related teasing which itself was associated with gelotophobia. A similar association could not be observed in boys. The strong association between weight status and weight-related teasing in girls is in accordance with the findings by Goldfield et al. (2010). However, they found an (albeit smaller) association for boys as well. In general, the significant path between weight-related teasing and gelotophobia in girls parallels findings on memories of appearance-related teasing in gelotophobic adults (Studies I and II) and adolescents (Edwards et al., 2010). In these studies, however, findings were independent from gender or gender differences were not explicitly tested. The fact that no hypothesis-compliant correlations could be demonstrated for boys in the present study must also be viewed critically with regard to the sample examined. The proportion of boys was not only low overall, but only one boy was overweight. To overcome the limitations of the study a replication of the performed analyses based on a larger sample size is needed.

STUDY IV

Main objective of the study was to further examine the relationship between weight status, weight-based victimization and gelotophobia among adolescent boys and girls. Based on a larger sample, the hypothesis to be tested is whether weight-related teasing mediates the relationship between weight status (e.g., being overweight or obese) and gelotophobia. Again, correlations of weight status, teasing, and gelotophobia with the adolescents' body image will be analyzed.

Previous research has repeatedly demonstrated that weight-related teasing and victimization in the school setting is associated with negative affect in adolescent boys and girls (for example, see Puhl and Luedicke, 2012). Given the importance of positive emotions at school (for an overview, see Pekrun et al., 2018), however, the present study will broaden the perspective by adding measures of joy for two prototypical school situations (i.e., writing a class test, being together with others in the school yard; Eschenbeck et al., unpublished). It is hypothesized that especially joy in the social situation (i.e., at the schoolyard) shows a strong negative association with teasing and gelotophobia.

Method

Participants and Procedure

Participants were recruited from secondary schools in southern Germany. Complete data sets were available for 178 adolescents (boys: n = 93, girls: n = 85; age: M = 13.80 years, SD = 1.13, Range = 12–16 years). Parental consent was required to participate in the study. Participants and their parents provided their informed consent prior to the start of the study. Adolescents completed a self-report questionnaire in their classes. The measures were administered by trained students.

Variables

Gelotophobia was assessed by the GELOPH<15>(Ruch and Proyer, 2008; see Study I), with Cronbach's α = 0.87 in the present study.

Teasing

The Teasing Questionnaire-Revised (Strawser et al., 2005; see Study III) was applied, yielding in a score for teasing (total) with α = 0.93 and a score for weight-related teasing with α = 0.86 (for scale description see Study III).

Joy at school was assessed by two subscales of the Multidimensional Anxiety Inventory for Children and Adolescents (MAICA; Eschenbeck et al., unpublished). Two school-related scenarios were presented: *class test* ("Imagine you are taking a test at school") and *schoolyard* ("Imagine you are together with your class mates [e.g., in the schoolyard, in the classroom during your break, in the locker room, on the way to school]"). Participants rated their emotions on a four-point Likert scale (almost never/never = 1, sometimes = 2, often = 3, almost always/always = 4) on five items (e.g., "I feel good," "I'm cheerful") for each of the two school scenarios presented. Reliabilities (Cronbach's α) were α = 0.89 for joy during school test and α = 0.87 for joy at the schoolyard.

Weight status

Participants reported height and weight. BMI was calculated from both self-reported height and weight (BMI = weight in kilograms/height in meter^2). Overweight and obesity were defined using the BMI reference values of Kromeyer-Hauschild et al. (2001) according to the recommendations of the German Working Committee on Obesity in Children and Adolescents. As in Study III, the original five categories were reduced to three categories by combining the two underweight categories (i.e., extremely underweight: n = 10; underweight: n = 12) as well as the overweight (n = 7) and the obesity category (n = 5) in one category each, resulting in the weight categories underweight (n = 22; 8 boys, 14 girls), normal weight (n = 144; 76 boys, 68 girls) and overweight (n = 12; 9 boys, 3 girls).

Body image was assessed by single item: "Do you think you are . . . (1) far too thin, (2) a bit too thin, (3) just the right weight, (4) a bit too fat, (5) far too fat?" (Kurth and Ellert, 2008; see Study III). As in Study III, the original five categories were reduced to three categories: "too thin" (n = 21; "far too thin": n = 4; "a bit too thin": n = 17), "just the right weight" (n = 96), and "too fat" (n = 61; "a bit too fat": n = 54; "far too fat": n = 7).

In addition, dummy-coded variables for weight categories (underweight: yes = 1, no = 0; overweight: yes = 1, no = 0) and body image ("too thin": yes = 1, no = 0; "too fat": yes = 1, no = 0) were computed.

Data Analysis

Data analysis was similar to Study III. Agreement between weight status and body image was evaluated by Cohen's kappa. Descriptive statistics and correlations among gelotophobia and teasing and with weight status, body image and joy at school were computed. Again, the analyses were performed separately

for boys and girls. In addition, 2 × 3 ANOVAs with the between-subject factors gender and weight status (underweight, normal weight, overweight) or body image ("too thin," "just right," "too fat") and the dependent variables gelotophobia and teasing were calculated. To examine whether there will be an indirect association between weight status and gelotophobia mediated by weight-related teasing, path analyses were computed as described for Study III.

Results

Weight categories and body image show a low but significant association (Cohen's $\kappa = 0.22$, $p < 0.001$; 95% confidence interval: 0.08–0.36; see **Table 6**). Separate analyses for male and female adolescents yielded a substantial correspondence between weight category and body image for boys ($\kappa = 0.44$, $p < 0.001$; 95% confidence interval: 0.22–0.65) and a much smaller one for girls ($\kappa = 0.08$, $p = 0.07$; 95% confidence interval: 0.00–0.24). The difference between boys and girls was mainly a result of a high proportion of girls with normal weight reporting a body image of "too fat" (54%), whereas in boys this proportion was lower (14%). Correspondingly, among the underweight girls the percentage of those with a body image of "just right" or "too fat" was 64%, whereas this proportion was 38% in boys.

Correlations between gelotophobia and teasing as well as the correlations of these variables with weight categories, body image and joy are presented in **Table 7**. For both boys and girls, gelotophobia was correlated positively with teasing (total) and weight-related teasing and negatively with joy at the schoolyard. The negative association between gelotophobia and joy at school was more pronounced among girls ($r = -0.50$) than among boys ($r = -0.20$; $z = 2.27$, $p < 0.05$). With regard to associations of gelotophobia with weight status and body image, only a significant correlation between gelotophobia and underweight status in boys emerged. Underweight boys reported higher fear auf being laughed at as non-underweight boys.

In general, results for teasing are similar to those reported for gelotophobia. Both teasing (total) and weight-related teasing were elevated in underweight boys. Also, joy at school was correlated with both teasing measures, whereas again the correlations were stronger for girls than for boys (zs > 3.00, ps < 0.01). A significant negative correlation for joy at class test with teasing (total) emerged but only for boys.

TABLE 6 | Cross tabulation of weight categories and body image (Study IV): observed frequencies.

Weight categories	Body image			
	"Too thin"	"Just right"	"Too fat"	Total
Underweight	**10**	9	3	22
Normal weight	11	**85**	48	144
Overweight	0	2	**10**	12
Total	21	96	61	75

"Just right" = "just the right weight." Cells with an expected agreement between weight categories and body image are printed in bold. $\kappa = 0.22$, $p < 0.001$.

Associations of body image with weight-related teasing varied as a function of gender. For boys, a body image of seeing oneself as "too thin" was associated with weight-related teasing as well as with teasing (total). In contrast, for girls, a body image of seeing oneself as "too fat" that was associated with weight-related teasing.

For a simultaneous analysis of either gender and weight status or gender and body image 2 × 3 ANOVAs with the dependent variables gelotophobia and teasing were calculated. The three 2 × 3 ANOVAs with gender and weight status (underweight, normal weight, overweight) all resulted in significant interactions of gender by weight status for gelotophobia; $F(2,172) = 3.55$, $p < 0.05$; and for teasing (total); $F(2,172) = 7.02$, $p < 0.01$; and for weight-related teasing; $F(2,172) = 4.94$, $p < 0.01$. **Figure 3** illustrates the findings for gelotophobia. Underweight boys reported the highest gelotophobia among boys whereas both normal weight and overweight girls reported the highest gelotophobia among girls. Probably due to low number of underweight and overweight participants, however, only for the normal weight groups gender differences were statistically significant.

Both analyses for the teasing variables indicated that underweight boys (teasing [total]: $M = 0.73$; weight-related teasing: $M = 0.63$) and overweight girls (teasing [total]: $M = 0.36$; weight-related teasing: $M = 0.89$) report the highest amount of teasing compared to almost all other groups (teasing [total]: Ms between 0.16 and 0.36; weight-related teasing: Ms between 0.15 and 0.50). Only for girls, teasing (total) in both overweight and normal weight girls was the same (Ms = 0.36). **Figure 4** illustrates the findings for weight-related teasing as a function of gender and weight category.

The 2 × 3 ANOVAs with gender and body image ("too thin," "just the right weight," "too fat" resulted in significant gender main effect for gelotophobia with higher scores for girls than for boys; $F(1,172) = 5.86$, $p < 0.05$ (see **Table 7**)[3].

For the two teasing variables interactions of gender by body image were significant. Interactions for teasing (total), $F(2,172) = 5.61$, $p < 0.005$, as well as weight-related teasing, $F(2,172) = 3.08$, $p < 0.05$, are shown in **Figures 5**, **6**, resp. For both teasing (total) and weight-related teasing, boys who perceived themselves as "too thin" and girls who perceived themselves as "too fat" showed the highest scores.

Finally, we examined whether there was an indirect association between weight status and gelotophobia mediated by weight-related teasing. Path analyses were computed separately for boy boys and girls, each with either overweight or underweight (dummy-coded) as the predictor. Only for boys a significant mediation effect was obtained. The independent variable was underweight (dummy-coded), mediator was teasing (total), and the dependent variable was gelotophobia. The 95% CI obtained for the indirect effect of underweight status on gelotophobia by bootstrapping was 0.24 (CFI: 0.04–0.64) and did not include 0 (z = 5,000 bootstrap resamples). The results for the interplay between underweight status, weight-related

[3]In the ANOVAs of gender by weight status the main effect of gender on gelotophobia did not reach significance.

TABLE 7 | Correlations of gelotophobia and teasing with weight categories, body image and joy at school (Study IV).

				Variable		
Variable		***M***	***SD***	**Gelotophobia**	**Teasing (total)**	**Weight-related teasing**
Gelotophobia	Boys	1.57	0.47	1.00		
	Girls	1.89	0.55	1.00		
Teasing (total)	Boys	0.25	0.31	0.52**	1.00	
	Girls	0.33	0.40	0.62**	1.00	
Weight-related teasing	Boys	0.28	0.44	0.49**	0.75**	1.00
	Girls	0.46	0.75	0.53**	0.90**	1.00
Underweight	Boys	0.09	0.28	0.30**	0.38**	0.32**
	Girls	0.16	0.37	−0.07	−0.18	−0.18
Overweight	Boys	0.10	0.30	0.01	−0.07	0.14
	Girls	0.04	0.19	−0.07	0.02	0.11
Body image "too thin"	Boys	0.11	0.31	0.17	0.36**	0.28*
	Girls	0.13	0.34	0.00	0.13	−0.14
Body image "too fat"	Boys	0.19	0.40	0.05	−0.12	0.08
	Girls	0.51	0.50	0.14	0.21	0.27*
Joy (class test)	Boys	2.14	0.79	0.01	−0.21*	−0.12
	Girls	1.79	0.67	−0.04	−0.08	−0.13
Joy (schoolyard)	Boys	3.25	0.72	−0.20*	−0.22*	−0.26*
	Girls	3.22	0.79	−0.50**	−0.62**	−0.61**

*Boys: n = 93, girls: n = 85. Gelotophobia, teasing, and joy scores are reported as item means (sum scores divided by number of items; Gelotophobia, range: 1–4; Teasing, range: 0–4, Joy, range: 1–4. Weight categories and body image are dummy-coded (1 = yes, 0 = no). *p < 0.05, **p < 0.01 (two-tailed).*

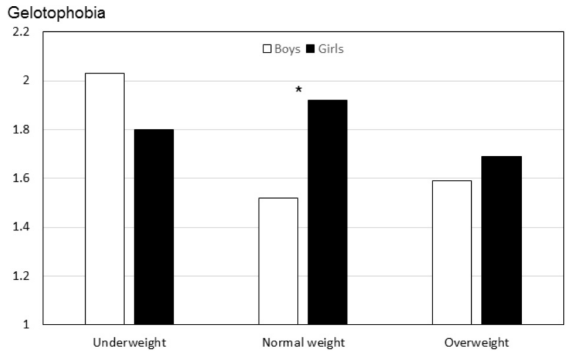


FIGURE 3 | Gelotophobia as a function of gender and weight categories (Study IV). Underweight (boys: *n* = 8, girls: *n* = 14), Normal weight (boys: *n* = 76, girls: *n* = 68), Overweight (boys: *n* = 9, girls: *n* = 3). Significant gender differences within weight groups are marked. *$p < 0.05$.

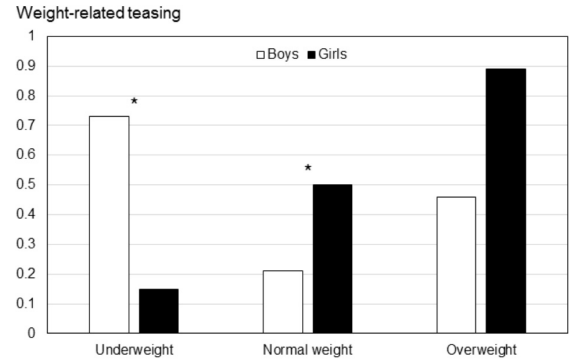


FIGURE 4 | Weight-related teasing as a function of gender and weight categories (Study IV). Underweight (boys: *n* = 8, girls: *n* = 14), Normal weight (boys: *n* = 76, girls: *n* = 68), Overweight (boys: *n* = 9, girls: *n* = 3). Significant gender differences within weight groups are marked. *$p < 0.05$.

teasing and gelotophobia are shown in **Figure 7**. The direct effect of being underweight on gelotophobia was fully mediated by weight-related teasing[4]. The findings indicate that gelotophobia in underweight adolescent boys is mainly a function of their experience of being teased.

Discussion

Findings of this study in part replicate the findings of Study III. First of all, the positive association between weight-related teasing and gelotophobia could be replicated for girls and extended to boys. Therefore, findings are in accordance with previous research on appearance-related teasing and gelotophobia (Edwards et al., 2010) and on weight-related teasing and psychological well-being (Goldfield et al., 2010; Puhl and Luedicke, 2012; Zuba and Warschburger, 2017).

Once again, weight status and body image were significantly but weakly associated. Above all, the high proportion of normal-weight participants with an unrealistically overweight body image, which was particularly pronounced in the present study among girls, corresponded to the previous findings (Kurth and Ellert, 2008).

[4]Similar findings resulted when teasing (total) as the mediator was analyzed. The 95% CI obtained for the indirect effect of underweight status on gelotophobia by bootstrapping was 0.30 (CFI: 0.02 to 0.87).

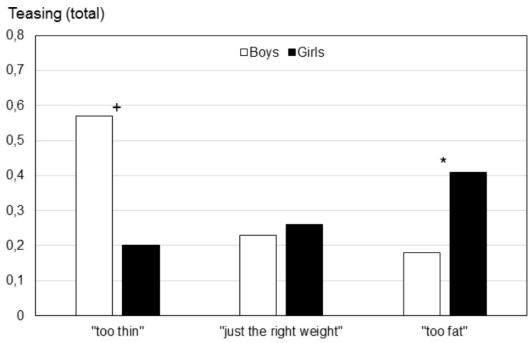


FIGURE 5 | Teasing (total) as a function of gender and body image (Study IV). "too thin" (boys: $n = 10$, girls: $n = 11$), "just the right weight" (boys: $n = 65$, girls: $n = 31$), "too fat" (boys: $n = 18$, girls: $n = 43$). Significant gender differences within body image groups are marked. $^{*}p < 0.05$, $^{+}p < 0.07$.

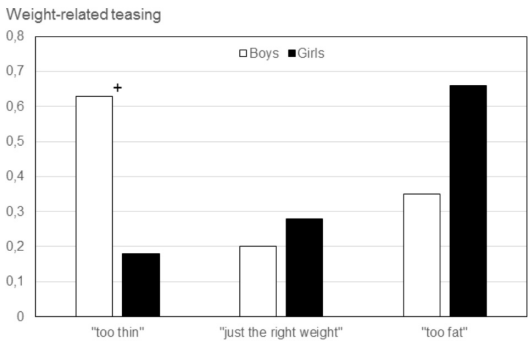


FIGURE 6 | Weight-related teasing as a function of gender and body image (Study IV). "too thin" (boys: $n = 10$, girls: $n = 11$), "just the right weight" (boys: $n = 65$, girls: $n = 31$), "too fat" (boys: $n = 18$, girls: $n = 43$). Significant gender differences within body image groups are marked. $^{+}p < 0.06$.

Concerning the association of weight status with gelotophobia and the mediation effect for weight-related teasing, a comparison of the present study with Study III becomes more difficult. A mediation effect emerged again. In the present study, underweight contributed to gelotophobia via weight-related teasing in boys only. In Study III, however, overweight predicted gelotophobia via weight-related teasing in girls only. The rather low number of boys in Study III and the low number of overweight girls in Study IV make it difficult to directly compare the findings of the two studies. Accepting these limitations, however, the findings of these first studies on the joint analysis of weight status, weight-related teasing and gelotophobia point to an association between weight status and gelotophobia mediated by weight-related teasing, albeit gender differences seem to play a crucial role. This was obvious also for the findings on body image.

Girls who perceived themselves as "too fat" as well as boys who perceived themselves as "too thin" reported the highest teasing scores. According to a review by Cohane and Pope (2001), although boys generally display less overall body concern than girls, many boys of all ages report dissatisfaction with their bodies, often associated with reduced self-esteem. Whereas girls typically wanted to be thinner, boys frequently wanted to be bigger. However, most studies failed to distinguish between "bigness" due to increased muscle and that due to fat. The male muscular body type seems to represent the dominant cultural ideal (Wienke, 1998). Assessment of body image in boys and men, therefore, should not only rely on fat but also on muscularity, resulting in two-dimensional assessment approaches (Cafri and Thompson, 2004). A recent study with male and female adolescents aged 12–16 years (Hoffmann and Warschburger, 2017) revealed more pronounced weight and shape concern in females than males and more pronounced muscularity concern in males than females.

For both teasing (total) and weight-related teasing, negative correlations with joy at school emerged. This finding is in accordance with findings from a study involving more than 90,000 14–16 year old Finish adolescents (Konu et al., 2002). Not being victimized at school was an important social relationship variable associated with psychological health and well-being.

Gelotophobia was related to reduced joy at the schoolyard but not during class test. Although the significant correlations were stronger in girls than in boys, the pattern of findings

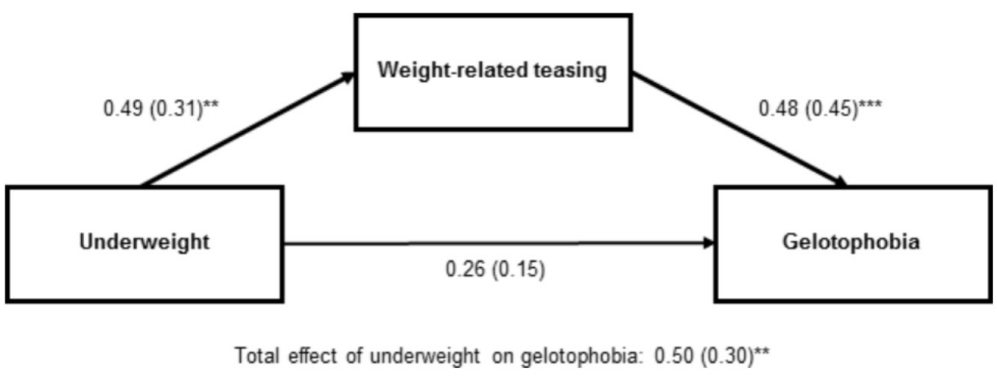


FIGURE 7 | Weight-related teasing as mediator between underweight and gelotophobia in adolescent boys ($n = 93$, Study IV). Unstandardized regressions coefficients and standardized regression coefficients (in parentheses) are based on item mean scores for weight-related teasing and gelotophobia, respectively. Underweight categories: 0 = no, 1 = yes. $^{**}p < 0.01$, $^{***}p < 0.001$.

for the public situation (i.e., at the schoolyard) and the more private performance situation (class test) support the construct validity of the gelotophobia concept. According to Titze (2009) gelotophobic individuals lack liveliness, spontaneity, and joy (see also Platt and Forabosco, 2012). According to Ruch et al. (2014) fear of being laughed cannot simply be reduced to a facet of anxiety, negative affectivity, or neuroticism. The social context is crucial as well. Emotions (either joy or anxiety) when writing a class test seem to be unrelated to gelotophobia because in this situation there is no risk of becoming a victim of teasing or ridicule. Being together with others at the schoolyard, however, may be extremely stressful for gelotophobic students, especially if one takes into account that the behavior on the schoolyard is less standardized than during a class test and that a possible intervening teacher is not nearby.

GENERAL DISCUSSION

The present series of four studies is the first attempt to apply the findings on the connections between obesity, victimization, and well-being (e.g., Puhl and Latner, 2007; Lampard et al., 2014) to the research on teasing and gelotophobia (Edwards et al., 2010; Ruch et al., 2010). It was investigated whether overweight, mediated by weight-related teasing, was also related to gelotophobia. The general discussion will focus on this assumed association pattern.

The two interview studies with adults with pronounced gelotophobia (Studies I and II) indicated that in childhood and adolescence the external appearance (including weight-related aspects) was seen as a possible cause (in addition to social behavior) for teasing and ridicule related to gelotophobia development.

The two correlational studies with adolescents (Studies III and IV) principally confirmed the connection between teasing and gelotophobia reported in the literature (e.g., Edwards et al., 2010). In addition, however, there were also first indications that weight status, mediated by weight-related teasing, may be associated with gelotophobia. Overweight status in girls (in Study III) and underweight status in boys (in Study IV) was related to weight-related teasing, which in turn was accompanied by increased gelotophobia. The reported mediation effect for girls from Study III could not be replicated in Study IV. Similarly, the mediation effect found for boys found in Study IV could not be observed in Study III. This may be due to the limitations within both samples already discussed within these studies.

Especially the sample sizes in these studies can be viewed as critical. Therefore, for the pronounced finding in Study IV of weight-related teasing as mediator between underweight and gelotophobia in adolescent boys, *post hoc* simulation studies were performed. Two Monte Carlo simulations run for the mediation model in two artificial data sets and 10,000 iterations revealed that power was only 0.62 for the direct effect of underweight on gelotophobia (power of the other regressions was sufficient, i.e., 0.80). Monte Carlo simulations with 10,000 iterations replicated in two artificial data sets showed that $N = 151$ boys yielded sufficient power for the regressions involved in the proposed mediation model.[5] Further research with more statistical power is warranted to investigate the role of teasing on the relationship between weight status and gelotophobia in more detail.

The findings of this set of preliminarily studies on appearance, body weight, weight-related teasing and gelotophobia make it worthwhile to further investigate the role of weight-related teasing for gelotophobia in male and female children, adolescents, and adults. However, several major improvements and extensions could then be made: (1) As already stated above, larger samples of underweight and overweight boys and girls would have to be examined. This would not only result in more robust findings but also allow to include measures of coping (e.g., social support) as potential buffers in the mediation models (Reiter-Purtill et al., 2017). (2) It would be advisable to objectively record the weight status. (3) Multi-dimensional assessment procedures could be used to capture the body image, taking into account not only weight perceptions but also figure and muscularity. (4) If all these points were implemented in a longitudinal design with children and adolescents, important further insights into the role of body weight and body image as well as teasing and victimization for mental well-being and especially gelotophobia could be expected. Recent research by Zuba and Warschburger (2017) suggests that the experience of weight teasing and internalization of weight bias is more important than weight status in explaining psychological functioning among children and indicate a need for appropriate prevention and intervention approaches. Further knowledge on the fear of being laughed at may contribute to deal with this challenge.

ETHICS STATEMENT

Studies I, II, and III were conducted following the ethical guidance of the University of Zurich ethics checklist. Full disclosure and informed consent was provided prior to participation in the study by clicking on an accept and continue link on the website. No participant had access to the study without agreeing. Study IV was conducted according to the ethical guidelines of the German Psychological Society. Study IV was approved by the state education authority. In Studies III and IV, children and their parents gave their informed consent prior to the start of the study.

AUTHOR CONTRIBUTIONS

All listed authors contributed meaningfully to the paper. TP and WR developed the concept of Study I. C-WK, TP, and WR developed the concept of Studies II and III. C-WK, HE, MH, and UH-D developed the concept and design of

[5]To estimate *post hoc* the necessary N to reach a power of 0.80 for each regression, we produced 500 simulations with $N = 20$, 500 simulations with $N = 30$, 500 simulations with $N = 40$ . . . and 500 simulations with $N = 300$ (Beaujean, 2014). We did this simulation with two different seeds (565 and 123). Both simulations showed approximately the same result for each factor of interest. The power analysis was done with R. For fitting the mediation for the simulation we used the lavaan package (Rosseel, 2012) and for running the Monte Carlo simulation we used the simsem package (Pornprasertmanit et al., 2016).

Study IV. C-WK, TP, and WR contributed to the design of Studies I, II, and III. C-WK, TP, and WR analyzed and interpreted the data of Studies I, II, and III. C-WK, HE, UH-D, and MH analyzed and interpreted the data of Study IV. C-WK prepared the draft manuscript, and HE, UH-D, MH, TP, and WR provided critical revisions. All authors approved the final version to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This research was supported in part by a Swiss National Science Foundation International Short Visit awarded to C-WK.

REFERENCES

Beaujean, A. A. (2014). Sample size determination for regression models using Monte Carlo methods in R. *Pract. Assess. Res. Eval.* 19, 1–16.

Berg, B. L. (2004). *Qualitative Research Methods for the Social Sciences*, 5th edn. Boston, MA: Pearson.

Cafri, G., and Thompson, J. K. (2004). Measuring male body image: a review of the current methodology. *Psychol. Men Mascul.* 5, 18–29. doi: 10.1037/1524-9220.5.1.18

Cohane, G. H., and Pope, H. G. (2001). Body image in boys: a review of the literature. *Int. J. Eat. Disord.* 29, 373–379. doi: 10.1002/eat.1033

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Measur.* 20, 213–220. doi: 10.1177/001316446002000104

Cole, T. J., Bellizzi, M. C., Flegal, K. M., and Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. *Br. Med. J.* 320:1240. doi: 10.1136/bmj.320.7244.1240

Davies, C. (2009). Humor theory and the fear of being laughed at. *Humor* 22, 49–62. doi: 10.1515/HUMR.2009.003

Edwards, K. R., Martin, R. A., and Dozois, D. J. A. (2010). The fear of being laughed at, social anxiety, and memories of being teased during childhood. *Psychol. Test Assess. Model.* 52, 94–107.

Eisenberg, M. E., Neumark-Sztainer, D., and Story, M. (2003). Associations of weight-based teasing and emotional well-being among adolescents. *Arch. Pediatr. Adolesc. Med.* 157, 733–738. doi: 10.1001/archpedi.157.8.733

Eschenbeck, H., Kohlmann, C.-W., Dudey, S., and Schürholz, T. (2009). Physician-diagnosed obesity in German 6- to 14-year-olds: prevalence and comorbidity of internalizing disorders, externalizing disorders, and sleep disorders. *Obesity Facts* 2, 67–73. doi: 10.1159/000209987

Fuchs, T., Eschenbeck, H., Krug, S., Schlaud, M., and Kohlmann, C.-W. (2012). Perception makes the difference: the association of actual weight and perceived weight status with self-reported and parent-reported personal resources and well-being in adolescents. *Appl. Psychol.* 4, 321–340. doi: 10.1111/j.1758-0854.2012.01077.x

Führ, M. (2010). The applicability of the GELOPH in children and adolescents: first evaluation in a large sample of Danish pupils. *Psychol. Test Assess. Model.* 52, 60–76.

Goffman, E. (1963). *Stigma: Notes on the Management of Spoiled Identity*. New York, NY: Simon & Schuster.

Goldfield, G., Moore, C., Henderson, K., Buchholz, A., Obeid, N., and Flament, M. (2010). The relation between weight-based teasing and psychological adjustment in adolescents. *Paediatr. Child Health* 15, 283–288. doi: 10.1093/pch/15.5.283

Griffiths, L. J., Parsons, T., and Hill, A. J. (2010). Self-esteem and quality of life in obese children and adolescents: a systematic review. *Int. J. Pediatr. Obes.* 5, 282–304. doi: 10.3109/17477160903473697

Gros, D. F., Stauffacher Gros, K., Randi, E., McCabe, R. E., and Antony, M. M. (2012). Psychometric evaluation of the factor structure of the Teasing Questionnaire – Revised (TQ-R). *J. Psychopathol. Behav. Assess.* 34, 542–551. doi: 10.3109/17477160903473697

Haines, J., Hannan, P. J., van den Berg, P., Eisenberg, M. E., and Neumark-Sztainer, D. (2013). Weight-related teasing from adolescence to young adulthood: longitudinal and secular trends between 1999 and 2010. *Obesity* 21, E428–E434. doi: 10.1002/oby.20092

Hayden-Wade, H. A., Stein, R. I., Ghaderi, A., Saelens, B. E., Zabinski, M. F., and Wilfley, D. E. (2005). Prevalence, characteristics, and correlates of teasing experiences among overweight children vs. non-overweight peers. *Obes. Res.* 13, 1381–1392. doi: 10.1038/oby.2005.167

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford.

Hoffmann, S., and Warschburger, P. (2017). Weight, shape, and muscularity concerns in male and female adolescents: predictors of change and influences on eating concern. *Int. J. Eat. Disord.* 50, 139–147. doi: 10.1002/eat.22635

Konu, A. I., Lintonen, T. P., and Rimpelä, M. K. (2002). Factors associated with schoolchildren's general subjective well-being. *Health Educ. Res.* 17, 155–165. doi: 10.1093/her/17.2.155

Kromeyer-Hauschild, K., Wabitsch, M., Kunze, D., Geller, F., Geiß, H. C., Hesse, V., et al. (2001). Perzentile für den Body-Mass-Index für das Kindes- und Jugendalter unter Heranziehung verschiedener deutscher Stichproben. *Monatsschrift Kinderheilkunde* 149, 807–818. doi: 10.1007/s001120170107

Kurth, B.-M., and Ellert, U. (2008). Perceived or true obesity: which causes more suffering in adolescents? *Deutsches Ärzteblatt Int.* 105, 406–412. doi: 10.3238/arztebl.2008.0406

Lampard, A. M., MacLehose, R. F. E., Eisenberg, M. E., Neumark-Sztainer, D., and Davison, K. K. (2014). Weight-related teasing in the school environment: associations with psychosocial health and weight control practices among adolescent boys and girls. *J. Youth Adolesc.* 43, 1770–1780. doi: 10.1007/s10964-013-0086-3

Neumark-Sztainer, D., Falkner, N., Story, M., Perry, C., Hannan, P. J., and Mulert, S. (2002). Weight-teasing among adolescents: correlations with weight status and disordered eating behaviors. *Int. J. Obes. Relat. Metab. Disord.* 26, 123–131. doi: 10.1038/sj.ijo.0801853

Pekrun, R., Muis, K. R., Frenzel, A. C., and Goetz, T. (2018). *Emotions at School*. New York, NY: Routledge.

Platt, T., and Forabosco, G. (2012). "Gelotophobia: the fear of being laughed at," in *Humor and Health Promotion*, ed. P. Gremigni (Hauppauge, NY: Nova Science Publishers), 229–252.

Platt, T., Ruch, W., Hofmann, J., and Proyer, R. T. (2012). Extreme fear of being laughed at: components of gelotophobia. *Israeli Journal of Humor Research* 1, 86–106.

Pornprasertmanit, S., Miller, P., and Schoemann, A. (2016). *Simsem. R Package for Simulated Structural Equation Modeling*. Available at: http://simsem.org

Preacher, K. J., and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav. Res. Methods Instr. Comput.* 36, 717–731. doi: 10.3758/BF03206553

Proyer, R. T., Neukom, M., Platt, T., and Ruch, W. (2012). Assessing gelotophobia, gelotophilia, and katagelasticism in children: an initial study on how six to nine-year-olds deal with laughter and ridicule and how this relates to bullying and victimization. *Child Indicat. Res.* 5, 297–316. doi: 10.1007/s12187-011-9127-1

Proyer, R. T., Ruch, W., Ali, N. S., Al-Olimat, H. S., Andualem Adal, T., Aziz Ansari, S., et al. (2009). Breaking ground in cross-cultural research on the fear of being laughed at (gelotophobia): a multi-national study involving 73 countries. *Humor* 22, 253–279. doi: 10.1515/HUMR.2009.012

Puhl, R. M., and Latner, J. D. (2007). Stigma, obesity, and the health of the Nation's children. *Psychol. Bull.* 133, 557–580. doi: 10.1037/0033-2909.133.4.557

Puhl, R. M., and Luedicke, J. (2012). Weight-based victimization among adolescents in the school setting: emotional reactions and coping behaviors. *J. Youth Adolesc.* 41, 27–40. doi: 10.1007/s10964-011-9713-z

Puhl, R. M., Moss-Racusin, C. A., Schwarz, M. B., and Brownell, K. D. (2008). Weight stigmatization and bias reduction: perspectives of overweight and obese adults. *Health Educ. Res.* 23, 347–358. doi: 10.1093/her/cym052

Puhl, R. M., Wall, M. M., Chen, C., Austin, B., Eisenberg, M., and Neumark-Sztainer, D. (2017). Experiences of weight teasing in adolescence and weight-related outcomes in adulthood: a 15-year longitudinal study. *Prevent. Med.* 100, 173–179. doi: 10.1016/j.ypmed.2017.04.023

Reiter-Purtill, J., Gowey, M. A., Austin, H., Smith, K. C., Rofey, D. L., Jenkins, T. M., et al. (2017). Peer victimization in adolescents with severe obesity: the roles of self-worth and social support in associations with psychosocial adjustment. *J. Pediatr. Psychol.* 42, 272–282. doi: 10.1093/jpepsy/jsw078

Rosseel, Y. (2012). Lavaan: an R package for structural equation modeling. *J. Statist. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Ruch, W. (2009). Fearing humor? Gelotophobia: the fear of being laughed at. Introduction and overview. *Humor* 22, 1–25. doi: 10.1515/HUMR.2009.001

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. T. (2014). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., and Proyer, R. T. (2008). Who is gelotophobic? Assessment criteria for the fear of being laughed at. *Swiss J. Psychol.* 67, 19–27. doi: 10.1024/1421-0185.67.1.19

Ruch, W., Proyer, R. T., and Ventis, L. (2010). The relationship of teasing in childhood to the expression of gelotophobia in adults. *Psychol. Test Assess. Model.* 52, 77–93.

Stamm, H., Wiegand, D., and Lamprecht, M. (2010). *Monitoring der Gewichtsdaten von Kindern und Jugendlichen in den Kantonen Graubünden, Wallis, Jura, Genf und Basel-Stadt sowie den Städten Freiburg, Bern und Zürich*. Bern: Gesundheitsförderung Schweiz.

Storch, E. A., Roth, D. A., Coles, M. E., Heimberg, R. G., Bravata, E. A., and Moser, J. (2004). The measurement and impact of childhood teasing in a sample of young adults. *J. Anxiety Disord.* 18, 681–694. doi: 10.1016/j.janxdis.2003.09.003

Strawser, M. S., Storch, E. A., and Roberti, J. W. (2005). The Teasing Questionnaire—Revised: measurement of childhood teasing in adults. *J. Anxiety Disord.* 19, 780–792. doi: 10.1016/j.janxdis.2004.09.005

Thompson, J. K., Cattarin, J., Fowler, B., and Fisher, E. (1995). The perception of teasing scale (POTS): a revision and extension of the physical appearance related teasing Scale (PARTS). *J. Pers. Assess.* 65, 146–157. doi: 10.1207/s15327752jpa6501_11

Titze, M. (2009). Gelotophobia: the fear of being laughed at. *Humor* 22, 27–48. doi: 10.1515/HUMR.2009.002

Wienke, C. (1998). Negotiating the male body: men, masculinity, and cultural ideals. *J. Men's Stud.* 6, 255–282. doi: 10.1177/106082659800600301

Zhao, X., Lynch, J. G., and Chen, Q. (2010). Reconsidering Baron and Kenny: myths and truths about mediation analysis. *J. Consum. Res.* 37, 197–206. doi: 10.1086/651257

Zuba, A., and Warschburger, P. (2017). The role of weight teasing and weight bias internalization in psychological functioning: a prospective study among school-aged children. *Eur. Child Adolesc. Psychiatry* 26, 1245–1255. doi: 10.1007/s00787-017-0982-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Kohlmann, Eschenbeck, Heim-Dreger, Hock, Platt and Ruch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 23 January 2018
doi: 10.3389/fpsyg.2018.00004

Fear of Being Laughed at in Borderline Personality Disorder

Carolin Brück, Stephanie Derstroff and Dirk Wildgruber**

Department of Psychiatry and Psychotherapy, University Medical Center, Eberhard Karls University of Tübingen, Tübingen, Germany

Building on the assumption of a possible link between biases in social information processing frequently associated with borderline personality disorder (BPD) and the occurrence of gelotophobia (i.e., a fear of being laughed at), the present study aimed at evaluating the prevalence rate of gelotophobia among BPD patients. Using the Geloph<15> , a questionnaire that allows a standardized assessment of the presence and severity of gelotophobia symptoms, rates of gelotophobia were assessed in a group of 30 female BPD patients and compared to data gathered in clinical and non-clinical reference groups. Results indicate a high prevalence of gelotophobia among BPD patients with 87% of BPD patients meeting the Geloph<15> criterion for being classified as gelotophobic. Compared to other clinical and non-clinical reference groups, the rate of gelotophobia among BPD patients appears to be remarkably high, far exceeding the numbers reported for other groups in the literature to date, with 30% of BPD patients reaching extreme levels, 37% pronounced levels, and 20% slight levels of gelotophobia.

Keywords: borderline personality disorder, gelotophobia, social cognition, laughter, fear of being laughed at

OPEN ACCESS

Edited by:
Hsueh-Chih Chen, National Taiwan Normal University, Taiwan

Reviewed by:
Jennifer Hofmann, University of Zurich, Switzerland
Liudmila Liutsko, Global Health Institute Barcelona (ISGlobal), Spain

****Correspondence:***
Carolin Brück
carolin.brueck@gmx.net
Dirk Wildgruber
dirk.wildgruber@med.uni-tuebingen.de

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *04 October 2017*
Accepted: *03 January 2018*
Published: *23 January 2018*

Citation:
Brück C, Derstroff S and Wildgruber D (2018) Fear of Being Laughed at in Borderline Personality Disorder. Front. Psychol. 9:4. doi: 10.3389/fpsyg.2018.00004

INTRODUCTION

Aside a pervasive pattern of instability in affect regulation, self-image, and impulse control, the list of traits that conceptualize borderline personality disorder (BPD) includes disturbances in interpersonal functioning as core clinical feature of the diagnosis (American Psychiatric Association, 2000; Lieb et al., 2004). In an effort to understand the underlying causes of BPD-related interpersonal dysfunction, research in recent years has devoted increasing attention to impairments in social cognition, i.e., impairments in the "mental processes involved in perceiving, attending to, remembering, thinking about, and making senses of the people in our social world" (Moskowitz, 2005, p. 3), as one likely contributor to BPD-related difficulties in social interaction. In this context, particularly failures to understand the emotions and intentions of others have frequently been the focus of research (Domes et al., 2008, 2009; Daros et al., 2013). Findings derived from this line of research lead to the assumption of a negativity bias in the evaluation of others (e.g., Arntz and Veen, 2001; Barnow et al., 2009; Domes et al., 2009) that drives patients to misperceive or misinterpret the nature of a social exchange possibly linked to a heightened "rejection sensitivity" (Miano et al., 2013), a "disposition to anxiously expect, readily perceive and intensely react to rejection" (Downey et al., 2004, p. 668) in BPD patients (Staebler et al., 2011).

Much like a lens through which BPD patients perceive the world, expectations of rejection may cloud patients' interpretations of social interactions in a way that leads them to almost automatically perceive signs of rejections in others and to interpret perhaps even innocent or friendly interactions as rejecting (Downey et al., 2004).

At the level of observable behavior this processing disposition may reflect itself in a variety of forms – one possibly being the experience of *gelotophobia,* a fear of being laughed at (Ruch, 2009), based on the misinterpretation of laughter signals generally as signs of hostility and rejection. Studies of groups of people taken form the general population have contributed to the concept of gelotophobia as a continuum (Ruch et al., 2014), whereas particularly the high end of the range described as fear of being shamed by the ridicule of others with a paranoid sensitivity to anticipate ridicule, a disproportional negative response to laughter (Ruch et al., 2014), difficulties in regulating emotional states and an anger proneness (Weiss et al., 2012) may provide compelling suggestions of an overlap with BPD and BPD-related concepts such as rejection sensitivity. Though at first glance similar to personality dimensions such as shame-proneness or social anxiety, gelotophobia has been shown to transcend global personality traits and has been established as a unique concept distinct from more general traits used to describe an adult personality (Ruch et al., 2014). As far as links between social impairments in BPD and the concept of gelotophobia are concerned, behavioral studies conducted in groups of gelotophobic individuals once more draw attention to commonalities in responding. Similar to reports of a negative perception bias in BPD (e.g., Brück et al., 2017) studies show that gelotophobic individuals tend to perceive benevolent laughter as more unpleasant, misinterpret the affective state of a laughing individual as negative in valence and judge cartoons depicting social scenes involving laughter as displays of mockery and ridicule (Ruch, 2009).

Given the aforementioned phenomenological similarities between social cognitive impairments in BPD and markers of gelotophobia, one might assume a high prevalence of gelotophobics among BPD patients. While research conducted in groups of patients suffering from other mental disorders such as anxiety disorders, eating disorders, mood disorders, schizophrenic disorder or autism, confirm relatively high rates of gelotophobia among the studied patient groups (Forabosco et al., 2009; Samson et al., 2011), rates of gelotophobia among BPD patients remain unknown.

Bridging the current gap in research, this study aimed at evaluating the prevalence rate of gelotophobia among BPD patients. To this end, a group of BPD patients was asked to complete the Geloph<15> (Ruch and Proyer, 2008), a questionnaire that allows a standardized assessment of the presence and severity of gelotophobia symptoms.

MATERIALS AND METHODS

Participants

A total number of 30 female patients (M_{age} = 23.47 years, age range: 19–34 years, $M_{education}$ = 11.23 years ± 1.72 *SD*) diagnosed with BPD volunteered to participate. Participants were chosen from a pool of patients seeking treatment. Patients had to be 18 years or older and diagnosed with BPD in order to participate. Patients were diagnosed by trained psychiatrists or clinical psychologists based on the criteria provided in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; Saß et al., 1996) using the International Personality Disorder Examination (IPDE; Loranger et al., 1998).

In addition to the diagnosis BPD, 15 of the 30 recruited BPD patients (=50%) also met the diagnostic criteria for mood disorders, nine (=30%) for eating disorders, nine (=30%) for anxiety disorders, five for disturbances of activity and attention (=17%), and 17 (= 57%) for substance abuse disorders. Seventeen of the 30 patients received some form of psychiatric medication: Six patients were treated with antidepressants, one with antipsychotics, ten with a combination of antidepressants and either antipsychotics, anxiolytics or mood stabilizers.

Materials

To quantify symptoms of gelotophobia, each participant was provided with a German version of the Geloph<15>, a standard instrument to determine the presence and intensity of the fear of being laughed at (Ruch and Proyer, 2008). Measures of gelotophobia are obtained using a set of 15 statements describing typical behaviors and attitudes of gelotophobes. Participants are asked to indicate the extent of their agreement with each statement choosing one of four answer alternatives: strongly disagree, moderately disagree, moderately agree, and strongly agree.

To determine individual gelotophobia scores for each participant, ratings are assigned numeric values ranging from 1 to 4 with higher values indicating a higher degree of agreement with each statement (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree). The respective rating values then are averaged among all 15 items of the questionnaire resulting in a total score ranging from 0 to 4. Interpretation guidelines and cutoff values provided by the authors allow further detailing the findings: Averaged scores of 3.5 or higher are generally interpreted to represent expressions of extreme gelotophobia, while scores falling between the limits of ≥3.0 and <3.5 are interpreted to indicate a marked/pronounced gelotophobia, and scores between the limits of ≥2.5 and <3.0 a slight form of gelotophobia. Averaged scores lower than 2.5 are interpreted to indicate that the respective individual experiences no fear of being laughed at. Based on these cutoff values and their individual scores patients were categorized into one of four groups – patients without gelotophobia, with slight gelotophobia, with pronounced gelotophobia, and with extreme gelotophobia – and the percentage of BPD patients falling within each group was determined.

Procedure

After admission to an inpatient-treatment program targeting individuals suffering from BPD, patients were approached by the first or second author and informed about ongoing studies concerning BPD and asked for their participation. Participation was voluntarily and did not interfere with the treatment program. Patients were given time to consider their participation and were revisited a few days after the first meeting. If a patient agreed to participate, a date at the patient's earliest convenience was scheduled to collect the data. Data were collected in a quiet room separate from the ward

patients sought treatment at. Participants were seated at a desk, handed a paper version of the Geloph<15> and asked to answer each questions without any constraints on the time needed to fill in the questionnaire. All participants had the opportunity to withdraw consent at any time during the study. Participants did not receive immediate feedback on their results. However, if a person asked for feedback, the results were explained.

Data Analysis

For a more in-depth interpretation of the gathered data, findings on the relative frequency with which the different severity levels of gelotophobia occur in BPD patients were compared to prevalence rates obtained in other clinical and non-clinical reference groups. As far as the non-clinical reference group is concerned, a sample of 30 female volunteers was recruited from the general population. The selection of individuals was based on the criteria of a similar age (M_{age} = 23.93 years, age range: 18–31 years) and level of primary education ($M_{education}$ = 11.47 years ± 1.57 *SD*) as BPD patients included in this study as well as on the criterion of not having been diagnosed with a psychiatric disorder. Prevalence rates for clinical groups were derived from research reports published in the current literature (Forabosco et al., 2009; Samson et al., 2011).

Chi-square tests were used to statistically compare proportions of gelotophobes and non-gelotophobes within the different groups, and odds ratios were calculated to estimate effect sizes (=odds of gelotophobia in BPD group (number of BPD patients with gelotophobia divided by number of BPD patients without gelotophobia) divided by odds of gelotophobia in reference group (number of individuals in reference group with gelotophobia divided by number of individuals in reference group without gelotophobia).

RESULTS

BPD patients' Geloph<15> scores ranged from 2.00 to 3.93. Across patients, the scores averaged to a mean of M_{Geloph} = 3.13 (*SD* = 0.53). Categorizations based on cutoff values provided by Ruch and Proyer (2008) revealed that 26 out of the 30 BPD patients included in this study (=86.67%) could be classified as gelotophobic (i.e., Geloph<15> score > 2.5). 30.00% of all patients (=9/30) indicated extreme levels of gelotophobia, 36.67% (=11/30) pronounced levels, and 20.00% (=6/30) slight levels.

Comparisons conducted with prevalence rates in other clinical or non-clinical reference group (**Table 1**) indicated significantly higher rates of gelotophobia among BPD patients as relative to females without mental disorders [$\chi^2(1, N = 60) = 38.57$, $p < 0.01$ two-tailed] or patient with autism [$\chi^2(1, N = 70) = 12.75$, $p < 0.01$ two-tailed], schizophrenia [$\chi^2(1, N = 56) = 8.86$, $p < 0.01$ two-tailed], or mood disorders [$\chi^2(1, N = 62) = 28.60$, $p < 0.01$ two-tailed]. Based on odd ratios calculated on the data, the odds of BPD patients exhibiting gelotophobia were 91.0 times higher than in females without mental disorders, 7.9 times higher than in patients diagnosed with autism spectrum disorder, 6.5 times higher than in patients diagnosed with a schizophrenia spectrum disorder, and 28.2 times higher than in patients diagnosed with mood disorders.

DISCUSSION

In sum, the data gathered in this study evidence a high prevalence of gelotophobia among BPD patients: Roughly 9 out of 10 patients meet the Geloph<15> criterion for being classified as gelotophobic with more than 60% of all BPD patients exhibiting pronounced to extreme levels of symptom manifestation. Relative to occurrences in other clinical and non-clinical samples, the rates of gelotophobia among BPD patients appears to be rather high, far exceeding the numbers reported for other mental disorders (see **Table 1**) in the literature to date. Keeping in mind limitations of a medium-sized all female sample of patients currently seeking intensive psychotherapy, interpretations concerning gelotophobic traits in BPD patients (male and female) within the general population must remain cautious at this time. Particularly when considering data suggesting responses of female BPD patients on self-report measures to be clouded by a more negative world view, a greater dissatisfaction and critical views of themselves as compared to male BPD patients (McCormick et al., 2007), further studies with a

TABLE 1 | Percentages of individuals with no, slight, pronounced, and extreme gelotophobia summarized for patients with borderline personality disorders as well as non-clinical and clinical reference groups.

	N	*NG*	G_{total}	G_s	G_p	G_e
Borderline personality disorder	30	13.33%	86.67%	20.00%	36.67%	30.00%
Non-clinical reference group	30	93.33%	6.67%	6.67%	0.00%	0.00%
Clinical reference group						
Autism spectrum disorders[1]	40	55.00%	45.00%	27.50%	10.00%	7.50%
Schizophrenia[2]	26	50.00%	50.00%	15.38%	30.77%	3.85%
Mood disorders[2]	32	81.25%	18.75%	12.50%	3.12%	3.12%

NG, no gelotophobia; Gtotal, gelotophobia (regardless of severity); Gs, slight gelotophobia; Gp, pronounced geloto-phobia; Ge, extreme gelotophobia. [1]*Data provided in Samson et al. (2011),* [2]*Data provided in Forabosco et al. (2009).*

mixed-sex sample become necessary to substantiate the current observation.

Besides a phenomenological description, at this point of time further studies should aim to advance our understanding as to what contributes to such extraordinarily high prevalence rate of gelotophobia and to detail how co-occurrences of gelotophobia may affect or even further complicate social lives of BPD patients.

As suggested earlier, specific cognitive-affective dispositions in the processing of social information – particularly the overreaching expectation of rejection – may provide an initial stepping stone to explain the occurrence of gelotophobia in BPD patients. Latter assumption builds on the idea that instead of correctly inferring different communicative intentions, the expectation of rejection leads patients to perceive laughter generally as sign of rejection and that coupled with the emotional turmoil of feeling rejected laughter becomes a signal to be feared – a misjudgment with perhaps dire consequences on a patients well-being and life in the community (e.g., social withdrawal, low self-esteem and social competence, or a lack of liveliness, spontaneity, and joy; Ruch, 2009). In this context, attention needs to be devoted to studying whether or not BPD patients indeed show impairments in the decoding of laughter signals in order to further substantiate the initial claims. While the hypothesis may be in line with reports of a reduced ability to correctly derive social information from other communication signals such as facial expressions (Domes et al., 2009; Daros et al., 2013), for example, to our knowledge no study to date has sought to investigate BPD-related alterations in the perception of laughter. Behavioral profiles derived from such studies combined with measures of gelotophobia and rejection sensitivity in the same samples of patients ultimately may allow to test the suggested model in which gelotophobia is mediated by specific dispositions of information processing and thus serves as another marker of a biased social perception in BPD.

With respect to the field of BPD treatment, knowledge about BPD-related phenomena such as gelotophobia may aid treatment planning, in a sense, that it may suggest to raise awareness to gelotophobic tendencies and their effects on social perception and to further discrimination learning with respect to social signals which ultimately may facilitate the cognitive restructuring of negative schemas regarding social interaction associated with BPD.

ETHICS STATEMENT

The study was performed according to the principles of the Code of Ethics of the World Medical Association (Declaration of Helsinki) and with the approval of the ethics review board of the University of Tübingen. Before inclusion in the study, all participants gave written informed consent.

AUTHOR CONTRIBUTIONS

CB contributed to the conception and design of the study, to acquisition of participants, collection, analysis, and interpretation of data, as well as manuscript drafting and revision. SD contributed to the acquisition of participants, as well as the collection, analysis and interpretation of the data. DW contributed to the conception and design of the study, to analysis, and interpretation of data, as well as manuscript drafting and revision.

FUNDING

We acknowledge financial support by the German Research Foundation (Deutsche Forschungsgemeinschaft) and Open Access Publishing Fund of the University of Tübingen to cover publication costs of this research project.

ACKNOWLEDGMENTS

We would like to acknowledge the help of Prof. Willibald Ruch in the conception of this work, and would like to thank him for providing us with the Geloph<15> questionnaire used in the study.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edn. Washington, DC: American Psychiatric Association.

Arntz, A., and Veen, G. (2001). Evaluations of others by borderline patients. *J. Nerv. Ment. Dis.* 189, 513–521. doi: 10.1097/00005053-200108000-00004

Barnow, S., Stopsack, M., Grabe, H. J., Meinke, C., Spitzer, C., Kronmuller, K., et al. (2009). Interpersonal evaluation bias in borderline personality disorder. *Behav. Res. Ther.* 47, 359–365. doi: 10.1016/j.brat.2009.02.003

Brück, C., Derstroff, S., Wolf-Arehult, M., Wekenmann, S., and Wildgruber, D. (2017). Perception of verbal and nonverbal emotional signals in women with borderline personality disorder: evidence of a negative bias and an increased reliance on nonverbal cues. *J. Pers. Disord.* 31, 221–231. doi: 10.1521/pedi_2016_30_245

Daros, A. R., Zakzanis, K. K., and Ruocco, A. C. (2013). Facial emotion recognition in borderline personality disorder. *Psychol. Med.* 9, 1953–1963. doi: 10.1017/S0033291712002607

Domes, G., Czieschnek, D., Weidler, F., Berger, C., Fast, K., and Herpertz, S. C. (2008). Recognition of facial affect in borderline personality disorder. *J. Pers. Disord.* 22, 135–147. doi: 10.1521/pedi.2008.22.2.135

Domes, G., Schulze, L., and Herpertz, S. C. (2009). Emotion recognition in borderline personality disorder-a review of the literature. *J. Pers. Disord.* 23, 6–19. doi: 10.1521/pedi.2009.23.1.6

Downey, G., Mougios, V., Ayduk, O., London, B. E., and Shoda, Y. (2004). Rejection sensitivity and the defensive motivational system: insights from the startle response to rejection cues. *Psychol. Sci.* 15, 668–673. doi: 10.1111/j.0956-7976.2004.00738.x

Forabosco, G., Ruch, W., and Nucera, P. (2009). The fear of being laughed at among psychiatric patients. *Humor Int. J. Humor Res.* 22, 233–251. doi: 10.1515/HUMR.2009.011

Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M. M., and Bohus, M. (2004). Borderline personality disorder. *Lancet* 364, 453–461. doi: 10.1016/S0140-6736(04)16770-6

Loranger, A. W., Sartorius, N., Andreoli, A., Berger, P., Buchheim, P., Channabasavanna, S. M., et al. (1998). *Deutschsprachige Fassung der*

International Personality Disorder Examination: IPDE. Geneva: World Health Organization.

McCormick, B., Blum, N., Hansel, R., Franklin, J. A., St John, D., Pfohl, B., et al. (2007). Relationship of sex to symptom severity, psychiatric comorbidity, and health care utilization in 163 subjects with borderline personality disorder. *Compr. Psychiatry* 48, 406–412. doi: 10.1016/j.comppsych.2007.05.005

Miano, A., Fertuck, E. A., Arntz, A., and Stanley, B. (2013). Rejection sensitivity is a mediator between borderline personality disorder features and facial trust appraisal. *J. Pers. Disord.* 27, 442–456. doi: 10.1521/pedi_2013_27_096

Moskowitz, G. B. (2005). *Social Cognition: Understanding Self and Others*. New York, NY: The Guilford Press.

Ruch, W. (2009). Fearing humor? Gelotophobia: the fear of being laughed at introduction and overview. *Humor Int. J. Humor Res.* 22, 1–25. doi: 10.1515/HUMR.2009.001

Ruch, W., Hofmann, J., Platt, T., and Proyer, R. T. (2014). The state-of-the art in gelotophobia research: a review and some theoretical extensions. *Humor Int. J. Humor Res.* 27, 23–45. doi: 10.1515/humor-2013-0046

Ruch, W., and Proyer, R. T. (2008). Who is gelotophobic? Assessment criteria for the fear of being laughed at. *Swiss J. Psychol.* 67, 19–27. doi: 10.1024/1421-0185.67.1.19

Samson, A., Huber, O., and Ruch, W. (2011). Teasing, ridiculing and the relation to the fear of being laughed at in individuals with Asperger's syndrome. *J. Autism. Dev. Disord.* 41, 475–483. doi: 10.1007/s10803-010-1071-2

Saß, H., Wittchen, H., and Zaudig, M. (1996). *Diagnostisches und Statistisches Manual Psychischer Störungen DSM-IV*. Göttingen: Hogrefe.

Staebler, K., Helbing, E., Rosenbach, C., and Renneberg, B. (2011). Rejection sensitivity and borderline personality disorder. *Clin. Psychol. Psychother.* 18, 275–283. doi: 10.1002/cpp.705

Weiss, E. M., Schulter, G., Freudenthaler, H. H., Hofer, E., Pichler, N., and Papousek, I. (2012). Potential markers of aggressive behavior: the fear of other persons' laughter and its overlap with mental disorders. *PLOS ONE* 7:e38088. doi: 10.1371/journal.pone.0038088

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Brück, Derstroff and Wildgruber. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 01 March 2018
doi: 10.3389/fpsyg.2018.00232

The Education of Playful Boys: Class Clowns in the Classroom

*Lynn A. Barnett**

Recreation, Sport and Tourism, University of Illinois at Urbana–Champaign, Champaign, IL, United States

This longitudinal study identified degrees of playfulness in 278 kindergarten-aged children, and followed them through their next three school years to determine how playfulness was viewed by the children themselves, their classmates, and teachers. Perceptions of the social competence, disruptiveness, and labeling as the class clown, were assessed from all perspectives in each of first through third grades. Hierarchical linear modeling was conducted to account for the nesting of the data (children within classrooms within schools) and for the lack of independence between the measures. A central finding confirmed extant literature in that gender differences were dominant, with playful boys regarded as distinct from their less playful counterparts, while no such discrepancies appeared for girls. Playful boys were increasingly negatively regarded as rebellious and intrusive and were labeled as the "class clown" by their teachers. These findings were in direct contrast with children's self-perceptions and those of their peers, who initially regarded more playful boys as appealing and engaging playmates. The data further revealed that the playful boys were stigmatized by their teachers, and this was communicated through verbal and non-verbal reprimands, and classmates assimilated this message and became increasingly denigrating of the playful quality in the boys. In stark contrast, girls' playfulness levels were not a consideration in ratings by teachers or peers at any grade, nor did their classroom behaviors show significant variation. These negative perceptions were likely transferred by teachers to peers and to the children themselves, whereupon they changed their positive perceptions to be increasingly negative by third grade. The results contribute to the literature by demonstrating that playfulness in boys (but not girls) is often associated with the "class clown" designation, and is viewed as an increasingly lethal characteristic in school classrooms, where compelling efforts are undertaken to discourage its expression and persistence.

Keywords: children's playfulness, class clown, disruptiveness, classroom behavior, teacher-student relationship

OPEN ACCESS

Edited by:
René T. Proyer, Martin Luther University of Halle-Wittenberg, Germany

Reviewed by:
Ben Mardell, Harvard University, United States
Ciara Laverty, University of Cambridge, United Kingdom

***Correspondence:**
Lynn A. Barnett
lynnbm@illinois.edu

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *19 December 2017*
Accepted: *12 February 2018*
Published: *01 March 2018*

Citation:
Barnett LA (2018) The Education of Playful Boys: Class Clowns in the Classroom. Front. Psychol. 9:232. doi: 10.3389/fpsyg.2018.00232

INTRODUCTION

A stream of research has systematically investigated young children's playfulness by endeavoring to determine its underlying structure, dynamics, correlates, and nomological network (Lieberman, 1977; Barnett, 1990, 1991a,b). The most consistent findings have determined that there are five constituent determinants of the playfulness quality, and that their combined effect is highly predictive (Lieberman, 1966; Barnett, 1990, 1991a). The physical spontaneity dimension reflects the child's activity level and physical coordination; social spontaneity captures his or her ability to move in and out of social play situations fluidly, to share, and to show leadership during peer play;

cognitive spontaneity reflects the degree to which imagination and creativity are shown in play by the child inventing games, roles, and characters; manifest joy is demonstrated by the degree of exuberance, joy, enthusiasm, and heightened positive emotions the child exhibits in play; and sense of humor encompasses the teasing, rhyming, humor appreciation, and joke-telling aspects shown during play.

Empirical studies have also sought to identify descriptive lexicons that appear to distinguish between children who have more and less of the playful quality (Singer et al., 1980). Barnett (1991b) found that the characteristics that differentiated high and low playful children were "bright," "active," "aggressive," "curious," "imaginative," "impulsive," "mischievous," "cheerful," "confident," "dependent," and "responsible." These results were informed by children's teachers and recreation leaders, as well as by adult observers naive to the children they rated—so they appear to be communally understood. Rogers et al. (1998) obtained significant correlations with the personality descriptors of approachable, adaptable, persistent, aggressive, impatient, competitive, and dependent with similar aged children. Several of these characteristics appear to be at variance with the consistently positive ones that the public seems to identify.

The Advent of a Structured Setting: The Formal School Years

All of these empirical studies on children's playfulness have been conducted with very young children, spanning the ages of 2 to 6 years, and there has been a dearth of explorations into playfulness with school-aged children. The preschool and kindergarten settings contrast sharply with those comprising primary school grades in that they are much more relaxed and informal, and the degree of structure, the number of rules, and the degree of adult supervision and scrutiny are all less. It is therefore quite reasonable that virtually all of the extant research on children's playfulness has been conducted in permissive settings where manifest playful behaviors are plenteous.

The transition from kindergarten represents a process of significant and extensive adaptation in which children must quickly learn to follow directions, pay persistent attention, and internalize new expectations and rules (McClelland et al., 2007; Suchodoletz et al., 2009). Hence, the move to formal schooling requires that children transition to a more structured environment that demands self- discipline and -control. Children's ability to regulate their behavior is critically important, in that it has been shown to be predictive of how well they adapt to school (Blair, 2002), and to their academic achievement through the primary grades and middle school (McClelland et al., 2000, 2006, 2007; Vitaro et al., 2005). Research has demonstrated that how well children are able to navigate this transition also forecasts their long-term educational trajectory (McClelland et al., 2006). Conversely, children who experience substantial difficulty are at greater risk for poor academic achievement, problems with social peer relationships, emotional and conduct problems, and dropping out of school before or during adolescence (Eisenberg et al., 2000; McClelland et al., 2000; Vitaro et al., 2005).

Children's adjustment to the formal classroom setting is typically assessed by teachers using three metrics (Perry and Weinstein, 1998): academic functioning, social functioning, and classroom behavioral functioning, with the latter regarded as the most essential to school readiness (Petriwskyj et al., 2005). Successful adjustment is defined by teachers as accommodation to the classroom culture, rules, and behavioral expectations (Petriwskyj et al., 2005). To transition and function effectively requires that children be able to exercise self-control over their behaviors (McClelland et al., 2007) and restraint in expressing emotions (Diener and Kim, 2004). Teachers view children who they perceive to be distracting or disruptive as a detriment to the classroom learning environment, and they endeavor to control, shape or extinguish these behaviors in multiple ways (Jones and Dindia, 2004).

The characteristics that depict young playful children and expound on their exuberant, physically active, spontaneous, and impulsive qualities appear to be incompatible with the more restrictive school setting where rules and structure prevail and where the requirements for children to constrain their behaviors are intensified. The representations of playful kindergarteners would posit they might have a problem in negotiating a less familiar and more stringently controlled environment and one which demands well-developed behavioral self-regulation. This would portend that playful children might encounter problems successfully adapting to the classroom setting and maintaining obedience to classroom rules and teachers' demands. Playful children might well find themselves in conflict with their teacher, who—because of the emphasis on "appropriate" classroom behaviors—might view playful characteristics as disruptive and troublesome. It was thus a major focus of the present study to investigate how playful children transition to the primary school setting by investigating how they are perceived by their teachers, particularly the extent to which they are viewed as disruptive to classroom decorum. We also wondered whether—as the degree of structure present in the school classroom increased—the difficulties children would incur in trying to manage their behavior (stifle their playful expression) would also increase. We speculated that more playful children might be perceived as increasingly disruptive by their teachers as they progress through the first three primary school years.

Playful Children and Their Classmates

The school classroom can also be regarded as a principal setting in which children interact with their peers (Rubin et al., 2006). Their ability to manage their behavior appropriately relates to their social competence, interpersonal skills, social status, and success in peer interactions (Vitaro et al., 2005; Trentacosta and Izard, 2007). Conversely, children's difficulties normalizing their behavior have been shown to cause problems with forming friendships, and more generally in developing social competence (McClelland et al., 2000).

Research has also shown that teachers exert influence on children's peer relationships (Hughes et al., 2001) by providing social cues about how likeable a peer is (Hughes et al., 2001; Farmer et al., 2011). The teacher's prominent visible role in the classroom provides extensive opportunities for students to

observe exchanges with classmates, and to develop ideas about those who are liked and disliked, which then influences their own affective evaluations (Hughes et al., 2001, 2014). In accordance with social referencing theory, Hughes et al. (2001, 2014) found that students reported liking classmates who they viewed as having a positive relationship with the teacher, and disliking those they observed to be conflictual. Their findings emphasize the significant role the teacher plays in serving as a socializing agent who can significantly influence children's social relationships with peers.

It is thus crucial to children's social development and status that they are perceived by their teacher as agreeable and affable. Classmates will recognize children whose relations and exchanges with their teacher are disapproving, and they may adjust their impressions to dislike or shun these peers. The children who are most likely to have negative interactions with their teachers are those with a higher tendency to show disruptive classroom behavior and less able to constrain or redirect their frequent offtask activities (Kean, 1995; Rothbart and Bates, 2006). Their teachers come to regard them as less competent academically, and more challenging to manage and teach (Rothbart and Bates, 2006). The characteristics that have been identified as distinguishing playful children, including the propensity to be more physically active (Lieberman, 1977; Singer et al., 1980; Barnett, 1991b), verbal (Singer et al., 1980), impulsive (Barnett, 1991b; Rogers et al., 1998), aggressive (Barnett, 1991b; Rogers et al., 1998), and mischievous (Barnett, 1991b), would portend a poor relationship with teachers, which could then be transmitted to their classmates. Consistent with this literature, we wondered if more playful children would, at least initially, be viewed by their peers different in social status than children who are less playful, and whether their views would change across time (grades).

Alternatively, in studies with middle school children, research has demonstrated that there is often a difference in perspective between children and their teachers. One such area in which divergences have been detected is in the extent to which various classroom behaviors are viewed as disruptive, and how serious misbehaviors are judged to be. While some studies have found teachers to regard disobediences as more unforgiving (Corsaro and Eder, 1990), other research has shown students attach harsher views of classroom transgressions (Dursley and Betts, 2015). What is significant for the present study is the divergence between the perceptions of students and their teachers, particularly in evaluations of disruptive behaviors in the classroom—one central focus of this study. We thus sought separate assessments regarding the extent to which incidences of classroom behavior might be considered disruptive, to consider the perspectives of teachers and students individually, and to examine the role of children's playfulness as predictive of any differences.

How Playful Children View Themselves

After little more than a month in their classroom, children as young as first graders have been shown to make inferences about their own abilities from cues provided within the classroom (Stipek, 1981; Stipek and Tannatt, 1984). They attend to the differential ways in which teachers respond to other students, how and where praise and criticism are overtly rendered, responses to questions that are asked, how assessments and grades are assigned, and groupings of students based on ability (Jussim, 1986; Weinstein et al., 1987). They are keenly aware of the academic and social expectations of their teacher and the differential treatment that ensues from varying degrees of compliance, and they ultimately adopt these expectations as their own (Weinstein et al., 1987).

Numerous studies have demonstrated that teachers may formulate expectations for their students based on their classroom behaviors (Dusek and Joseph, 1983), and that these expectations can influence students' performance and motivation (for a review see Jussim and Harber, 2005). Known as the "Pygmalion Effect" (Rosenthal and Jacobson, 1968) in educational research, it has been shown that teachers may form initial expectations for a student, and they then behave toward the student in accord with their expectations. In turn, students respond to the ways in which they are treated by teachers, and ultimately they may internalize these expectations, with the result that a self-fulfilling prophecy is evinced. While an initial teacher-held expectation may be erroneous in whole or in part, through numerous interactions students will come to behave in such a manner as to confirm the expectation, which then becomes more accurate. In the present context, we hypothesized that teachers who view playful students as a problem in the classroom may hold differential behavioral expectations consistent with their perception, and students may come to adopt these negative behaviors, and regard themselves in corresponding ways (Weinstein, 2002; Jussim et al., 2009; McKown et al., 2010). We thus endeavored to explore the extent to which more playful students were affected by the assessments held by their teachers, and how readily this might have transpired. We included students' self-assessments of the social and behavioral constructs evaluated by teachers in each grade to determine whether any transmittals occurred, and how quickly and to what extent. In concert with the Pygmalion effect and a self-fulfilling prophecy to which it may lead, we inquired as to whether playful children would come to perceive themselves to be disruptive in the classroom, consonant with the perceptions of their teachers.

The Clown in the Classroom

There are typically students in every classroom who are considered disruptive because they use humor in the form of jokes, gestures, and antics with the goal of amusing or entertaining other students, and they have been often been labeled "class clown." In comparisons to non-clowns, they have the common signature strength of generating and appreciating humor (Ruch et al., 2014), and teachers characterize them as more assertive, attention-seeking, and unruly (Ruch et al., 2014). They are almost always differentiated as interfering with the classroom climate, and are regarded by most of their teachers as presenting a disciplinary problem (Cohen and Fish, 1993; Hobday-Kusch and McVittie, 2002; Ruch et al., 2014). The bestowal of the "class clown" label and the negative attributes that accompany it can be of concern, in that research has found that boys who frequently clowned in the classroom received more

negative criticism from their teachers even when they weren't clowning (Yarrow et al., 1971). Students who exhibit "clowning" behavior are labeled "problematic" by their teacher because they require that a disproportionate amount of time and attention be devoted to them (Cohen and Fish, 1993; Ruch et al., 2014; Platt et al., 2016). In studies with primary grade teachers, children who teachers considered to be "aggressive," "impudent," "impulsive," "lazy," (Brophy and Good, 1974) and "non-conforming" (Helton and Oakland, 1977) were least preferred and were students who teachers most often nominated to leave their classroom. Teachers' overriding concern was that there would be a "behavioral contagion" or "ripple effect" originating from the unruly student's presence in the classroom (Safran and Safran, 1984). They will have doubts about whether these "difficult" students are able to participate and embrace the learning being offered, and will come to question whether they even belong in the classroom.

The descriptors seen as characterizing class clowns (Damico and Purkey, 1978; Ruch et al., 2014), and the behaviors that have been attributed to them, are virtually indistinguishable from those that have been empirically found to characterize the playful child. The use of humor has consistently shown to be a quality of playful children, as has their impulsivity, spontaneity, disobedience, verbosity, reactivity, and aggressiveness (Barnett, 1991b; Rogers et al., 1998). It seemed likely that the designation of "class clown" might be disproportionately attached to more playful children, and this was identified as an additional query in the study. We also considered the "class clown" label, and its association with playfulness, from the perspectives of the child, the teacher and of peers, wondering whether these perceptions might be divergent. This was also an important purpose of the study, as virtually all previous empirical studies have explored what it means to be the "class clown" from the perspective of older children and adolescents. We anticipated that if similar findings were discovered for younger children, they might stimulate longitudinal investigations as children aged developmentally.

Playful Boys and Girls in the Classroom

There are a number of indications, drawn from diverse literatures, that the constructs under scrutiny in the present study might yield differential outcomes for playful boys and girls. There is a voluminous body of literature demonstrating differences between boys and girls in a multitude of variables within and surrounding the school classroom. The amount and quality of teacher-student interactions, even in kindergarten, has been shown to be different for boys and girls (Jones and Dindia, 2004), and they are generally viewed as more difficult to manage (Matthews et al., 2009), and more disruptive in the classroom (Jones and Dindia, 2004) compared to girls. From the early grades on, teachers report, and have been consistently observed, to use significant amounts of negative feedback in the classroom to try to control boys' behaviors (Jones and Dindia, 2004).

Despite the plethora of research consistently documenting sex differences in play from birth through adolescence (Hughes, 2010), the literature investigating differences between boys and girls in their magnitude, scope, or expression of playfulness is sparse. The few studies of global playfulness with preschool and kindergarten-aged children have generally not detected any sex differences (Lieberman, 1977; Barnett, 1990, 1991a). However, some sex differences have been detected among the component playfulness dimensions (Barnett, 1991b). Boys were shown to exhibit heighted physical spontaneity, while girls surpassed them in social spontaneity, however, in the three other playfulness components no differences were observed. The high prevalence of boys, and not girls, among children labeled "class clown" (Yarrow et al., 1971; Damico and Purkey, 1978; Ruch et al., 2014; Platt et al., 2016) would also portend that sex differences would be evident in assigning this label to playful children. The presence of sex differences was thus a pervasive inquiry throughout the study in each of the different outcome variables and their relationships with playfulness.

METHODS

Participants

Participants were 278 children and their parents and teachers ($n = 43$) in kindergarten through third grades from six Midwestern public elementary schools. School records indicated that a large majority of the children were White (81%, Black = 15%, Hispanic = 1%, bi-racial = 2%, >1% of another race/ethnicity), and there were slightly more females than males (54%, $n = 150$). The age range of the children corresponded to their grade level at the time of testing (kindergarten: M = 5.7 years, SD = 0.42 years, $n = 299$; first grade: M = 6.6 years, SD = 0.61 years, $n = 295$; second grade: M = 7.6 years, SD = 0.64 years, $n = 289$; third grade: M = 8.7 years, SD = 0.70 years). Two-thirds of the children (68%, $n = 189$) were currently residing in two-parent homes at the time of testing, and 94% ($n = 261$) had at least one parent who was employed full-time. The sample, as well as individual classes and grades, had a normally distributed range in socioeconomic level, with the largest percentage (22%, $n = 61$) of family annual gross income falling within the $40,000 to $75,000 bracket (range = $15,000 to >$150,000).

A total of 43 teachers ($n_{\text{kindergarten}} = 12$, $n_{\text{1stgrade}} = 11$, $n_{\text{2ndgrade}} = 10$, $n_{\text{3rdgrade}} = 10$) provided data for the study, the vast majority of whom were female (95%, $n = 41$) and self-identified as White (93%, $n = 40$), with only a few others indicating Black (5%, $n = 2$) or bi-racial (2%, $n = 1$). There were no Hispanic/Latino teachers or Native Americans participating in the study. Teachers had been in the profession for an average of 14.90 years (range = 4–27), and in the focal school for 8.340 (range = 2–18) years. None of the teachers resided in a neighborhood below the median income level. Preliminary ANOVA tests detected no differences between teachers across or within grades on any of the demographic measures (all $p > 0.05$).

Measures

Playfulness

Children were measured on their degree of playfulness utilizing the Children's Playfulness Scale (CPS; Barnett, 1990), which has been validated for children between the ages of 27 and 68 months (Barnett, 1990, 1991a; Trevlas et al., 2003). The scale consists of 23 descriptive statements to which teachers (or parents) respond utilizing a 5-point scale with responses labeled "sounds exactly

like the child," "sounds a lot like the child," "sounds somewhat like the child," "sounds a little like the child," and "doesn't sound at all like the child." Responses to the items are summed after inverted coding on designated items, such that higher scores indicate a greater degree of playfulness. Factor analyses with preschool and kindergarten-aged children have shown that the 23-item scale is comprised of the five dimensions of "physical spontaneity" (e.g., "The child is physically active during play"), "social spontaneity" (e.g., "The child plays cooperatively with other children"), "cognitive spontaneity" (e.g., "The child invents his/her own games to play"), "manifest joy" (e.g., "The child demonstrates enthusiasm during play"), and "sense of humor" (e.g., "The child enjoys joking with other children"). In the present study, the internal consistency for the total CPS score for each grade was highly satisfactory (kindergarten: $\omega = 0.92$, first grade: $\omega = 0.91$, second grade: $\omega = 0.91$, third grade = 0.93), as were reliability coefficients for each dimension within each grade (ranges across dimensions for kindergarten: $\omega = 0.88$–0.95, ranges across dimensions for first grade: $\omega = 0.87$–0.93, ranges across dimensions for second grade: $\omega = 0.88$–0.94, ranges across dimensions for third grade: $\omega = 0.90$–0.93). [The omega statistic, as a measure of internal consistency, has been found to be more appropriate and preferable to Cronbach's coefficient alpha (Huysamen, 2007; Sijtsma, 2009; Dunn et al., 2014)]. In addition, confirmatory factor analysis replicated the 5-factor structure (available from the author) of the CPS and corroborated previous findings (Barnett, 1990, 1991a,b; Trevlas et al., 2003), validating its use with the present sample.

Peer-Rated Social Status

Children's social status was assessed using a peer-rating measure similar to that used by Asher et al. (1979) with children in kindergarten through third grades (Eisenberg et al., 2000). With the help of a research assistant, children rated classmates on a 4-point scale (4 = "You play with the child a lot—he or she is like a best friend" to 1 = "You do not play together because you don't want to"). Ratings by same-sex raters were averaged, as were ratings by other-sex raters, and these two scores were then averaged, following procedures utilized in previous research (Eisenberg et al., 2000). Children with parental consent were asked to rate all other children with consent, and those without consent were asked to rate their desire to play with the same number of characters from popular media familiar to their age group. A social status score for each authorized child was determined from the mean score across the peer raters. The internal consistency (ω) of the scale for the children with consent was 0.88 in first grade, 0.81 in second grade, and 0.84 in third grade.

Self-Rated Social Competence

Children's self-perceptions about how accepted or popular they were with their peers was assessed from their scores on the six items comprising the Social Competence subscale of the Self-Perception Profile for Children (SPPC; Harter, 1982, 1985). In each item pair, one statement depicted a child who was more socially accepted, and the contrasting statement portrayed a child who was less so (e.g., "Some kids find it hard to make friends" BUT "Other kids don't find it hard to make friends"). Once the child chose a statement, he or she was then asked to indicate the extent to which the statement was like him or her ("sort of true" or "really true"). Each statement was assigned a value between "1" and "4," with a higher score indicating more social competence. Following the recommendation of the scale's author, an assistant read each question to children at all grade levels. Reliability and validity of the scale and subscales with young children (early elementary school ages) have been well-documented (Muris et al., 2003; Harter, 2012a). The SPPC Social Competence subscale has been found to correlate significantly with ratings of children's acceptance by peers and teachers (Harter, 1985). Internal consistency reliability for the present sample was good in each grade (first grade: $\omega = 0.90$, second grade: $\omega = 0.86$, third grade: $\omega = 0.85$).

Teacher-Rated Social Competence

Teachers' perceptions of children's social competence were assessed by the teachers version of the SPCC scale (Harter, 1982, 1985, 2012a). The Social Competence subscale was utilized, and item ("This child finds it hard to make friends" vs. "For this child it's pretty easy to make friends") and response ("really true" and "sort of true"; 4-point scale) formats were identical to the children's version (see above). A total mean score reflecting the teacher's perception of the child's social competence was used in the analyses. For the sample of teachers, the ω values of this subscale were 0.90 for first grade, 0.86 for second grade, and 0.87 for third grade.

Peer-Rated Classroom Disruptive Behavior

A measure of perceptions of their classmates' disruptive classroom behaviors was developed from the DBR-SIS (see above). Specific behaviors that comprised the disruptive behavior category were generated from previous DBR-SIS research (Riley-Tillman et al., 2009; Christ et al., 2011), and wording of the items was simplified to be appropriate for the age of the children. The eight disruptive acts presented to children were: "gets out of his/her seat without permission," "talks or yells about things we're not working on," "makes sounds (like humming, laughing, whistling) that aren't allowed during class time," "talks to other kids when we're not allowed to," "calls out things to the teacher without permission to talk," "does or says things that interrupt what we're doing," "is rude or mean to the teacher," and "plays with things at his or her desk that don't have anything to do with our work." Children were asked to rate each of their peers on each behavior using a 3-point response scale of "never," "sometimes," and "a lot/always." The scale was administered to each child individually with the help of a graduate assistant and was completed over contiguous 3-day sessions. Responses to the items were summed, with a higher score indicating disruptive behaviors in the classroom were exhibited very often. The internal consistency reliability of the scale was satisfactory for all grades in the study (first grade: $\omega = 0.79$, second grade: $\omega = 0.82$, third grade: $\omega = 0.87$).

Self-Rated Classroom Disruptive Behavior

Children rated themselves on the extent to which they felt they exhibited disruptive behaviors in their classroom utilizing the same modified DBR-SIS scale used for peer assessments (see above). At the end of the second day after providing ratings for their classmates, the assistant asked the child "What about you—how much do you think YOU do this' for each of the eight disruptive behaviors. The same 3-point response scale was used, and a total score was calculated to indicate the child's perception of how often he or she exhibited disruptive classroom behaviors. The internal consistency for these self-ratings was acceptable for all grades (first grade: $\omega = 0.92$, second grade: $\omega = 0.89$, third grade: $\omega = 0.88$).

Teacher-Rated Disruptive Student Behavior

In each grade, teachers were asked to rate students' classroom behaviors using the DBR-SIS (Direct Behavior Rating—Single Item Scales; Riley-Tillman et al., 2009; Chafouleas, 2011) following four 2-h instructional sessions toward the end of the school year. At the end of each class period teachers were asked to estimate how often each student showed disruptive behavior on a 5-point scale (1 = "never/almost never," 3 = "sometimes," 5 = "always/almost always"). Disruptive behavior was defined as "student action that interrupts regular school or classroom activity, such as students getting out of their seat, fidgeting, and yelling" (Johnson et al., 2016, p. 43). The DBR-SIS was selected because of its favorable reliability and validity ratings for kindergarten through eighth grade children across different raters over time, as well as its ease of use (for a review see Johnson et al., 2016). In the present study, reliability across the four rating sessions (first grade: $\omega = 0.88$, second grade: $\omega = 0.90$, third grade = 0.87) was acceptable at each grade.

Peer Rating of Child as Class Clown

Children in each grade were seated with an assistant and read the following script: "Most classrooms have a few students who joke a lot and try to make others in the room laugh. Sometimes these students are funny and sometimes they are not really funny. Please tell me the names of students who clown around a lot of the time." If a child received 25% or more of students' nominations in a classroom he or she received a score of "4" indicating the "class clown" designation (Fang, 2001). If a child received 15–24% of classmates' nominations a score of "3" was assigned, indicating some but not consistent regard as a class clown. For 5–14% of peer nominations, a score of "2" was given, and students receiving 4% or fewer nominations were attributed a score of "1" indicating the child was not commonly regarded as a class clown.

Self-Rating as Class Clown

In addition to being asked about their classmates' clowning behaviors, the question was posed to children by an assistant: "What about you—do you think you clown around in your class a lot of the time, some of the time, or not at all?" The children's responses were scored from "1" to "3," with higher scores indicating self-perceptions of more frequent clowning behaviors in the classroom.

Teacher Rating of Child as Class Clown

Teacher ratings were obtained using a different format, with classroom children's names and a 4-point Likert-scale presented. At the top, the same initial statements were provided: "Most classrooms have a few students who joke a lot and try to make others in the room laugh. Sometimes these students are funny and sometimes they are not really funny." Teachers were then asked to rate each child utilizing the four response options of "child does this almost all of the time" (scored as "4"), "child does this a lot of the time" (scored as "3"), "child does this some of the time" (scored as "2"), and "child does this seldom or never" (scored as "1"). This rating scale was intended to be comparable to that provided by peers in that higher scores signified stronger perceptions of the child as a class clown. The mean score was utilized in statistical analyses.

Demographic Characteristics

Demographic information was obtained from a parent questionnaire, which contained questions including the age and sex of the child, number of months of preschool they attended, and age and sex of each sibling currently in the home. Each child's birth order was determined from this information, in recognition of early research detecting such differences in play with preschool and kindergarten aged children (Moore et al., 1974).

Procedures

Data Collection

Data were collected as part of a larger study exploring family and sibling interrelationships in school readiness, academic skills, and social competence in public elementary schools in the midwestern United States. After obtaining approvals from universities, school district administrations, principals, kindergarten through third grade teachers, and parents, assessments for teachers and questionnaires for parents were distributed. After three follow-up mailings, a response rate of 71% from parents, and 83% from teachers was obtained. Teachers were instructed that questionnaires could be completed in their free time when they were able to concentrate, following a typical school day (provided no special "incidents" occurred either in the classroom or school), and they should only respond to questions about children whom they felt they knew well. Teacher assessments were administered toward the end of each of the academic years so that they were based on numerous interactions with each child. Teachers were compensated for their time and thanked for their participation.

Children completed instruments individually toward the end of an academic year with the help of graduate research assistants who were blind to the purposes of the study. For each questionnaire they read the instructions and items aloud, provided examples of how a child might respond, and assisted with recording responses. Data was collected over a 2-week time period. Prior to testing, graduate assistants underwent training that included viewing and discussing videotapes of children (not in the final sample) on all assessments and rehearsing questions and situations that might occur. Coders were required to achieve at least 90% for both inter- and intra- rater reliability conducted

with a sampling of eight videotaped children on all measures. Children were provided with a gift card of their choice from either a local toy or book store, in consultation with their parent. Children whose parent either did not respond or who declined consent were provided with questions about their likes and dislikes in toys and play. They also received compensation following consent from a parent.

Missing Data

From one grade to the next, some missing data occurred because a focal child moved away or left school for other reasons. To examine whether sample attrition influenced the results, three groups of children with complete data (randomly chosen n of 40), those missing data at one time period ($n = 39$), and individuals missing data at two or more times ($n = 14$) were compared (utilizing 1-way multivariate F-tests for continuous data and chi-square tests for frequency data) on all outcome measures and control (demographic) variables. None of these comparisons showed a statistically significant difference (all $p > 0.05$); it was thus concluded that there were no differences in the demographics or outcome measures due to the study procedures, and hence generalizability was not likely affected. Only children with complete data at all grade levels ($n = 278$) were included in subsequent data analyses.

Data Analysis Strategy

The research design involved children and teachers who were nested in classrooms that were nested in schools. Initially, the 1,228 children who participated in the study were nested within four kindergarten classrooms within five schools, in first grade they were nested in three classrooms, in second grade they were in another three classrooms, and by grade three, these same children were in three other classrooms within the same school. No more than six children from the kindergarten class remained in the same first, second, and third grade classrooms together in any school.

Hierarchical Linear Modeling (HLM; Raudenbush and Bryk, 2002) was employed to account for the nesting of multiple observations per child, children in classrooms, and classrooms within schools. Model testing began with tests of simple unconditional models without the playfulness or sex predictors to determine intraclass correlations (ICCs) for all outcome variables. The ICC is an indication of the amount of variance that can be explained at each level, i.e., the extent to which children's outcome scores may be alike due to membership in the same classroom or school. Inspection of the ICC coefficients afforded decisions to be made about the number of levels that should be included in the final model.

Two conditional models with playfulness and sex predictors were tested, with the second adding an interaction between them. Comparisons between these models for all outcome measures indicated that for virtually all measures (the exception being first grade teachers' social competence ratings) the interaction term improved the model, as determined by it having the lowest Akaike Information Criterion value (AIC = 187.083; Vrieze, 2012), and hence the Playfulness x Sex interaction was retained. Initial conditional models included four potential covariates (experience in preschool, number of brothers and sisters, birth order) however none of these variables was a significant predictor of any of the outcome scores. Since these controls failed to reach statistical significance for all measures, they were excluded from final model construction to maintain parsimony.

The spline (or piecewise) extension of HLM (Raudenbush et al., 2011) allowed estimation of one model for each outcome at each grade and interval (thus assessing change from one grade to the next). As in typical piecewise regression frameworks, the specification allowed for separate slope estimates for each grade. To explore changes from one grade to the next we constructed the two time intervals of first to second grade (interval one), and second to third grade (interval two). The pattern of coefficients reflected in changes in each assessment during the interval of first to second grade and the interval from second to third grade were modeled as a function of the child's playfulness and/or sex.

At Levels 1 (within children) and 2 (between children), the predictor variables of playfulness (continuous) and sex (dichotomous) were person-centered and grand mean-centered, respectively (Raudenbush and Bryk, 2002). Person-centering afforded the opportunity of assessing change within the child from one grade to the next, as it reflects deviations from each child's own score. Grand mean-centering at Level 2 focused instead on how the child differed from other children on each outcome variable. Final models were tested for violations of the assumptions of HLM and none were found to deviate significantly (all $p > 0.05$).

In the event of a significant interaction, recommended procedures (Raudenbush and Bryk, 2002) involving dismantling the interaction into its component parts was followed to facilitate interpretation. Dummy coding sex (boys = 1) eliminated collinearity concerns so that significant interactions could be inspected for interpretation of main effects (Aiken and West, 1991). Procedures proposed by Hochberg (1988) were adopted, as they'd been shown to be the most appropriate for repeated measures designs with correlated outcome variables (Lix and Sajobi, 2010), and involved making an adjustment to conventional alpha levels (0.05, 0.01) in order to control for family-wise error rate. The 0.01 alpha level (Hochberg adjustment for 0.01 alpha = 0.0089) was set as the minimum for consideration of statistical significance in all HLM analyses.

RESULTS

Preliminary Findings

Descriptive Statistics

Calculations of skewness and kurtosis were conducted for all measures and inspected for deviations from normality. None of the outcome measures had skew or kurtosis indices that exceeded accepted values (skew ranged from 0.09 to 1.78; kurtosis ranged from −0.23 to 1.60), suggesting no significant departures, and hence no transformations were deemed necessary (Field, 2013). Means and standard deviations

for all outcome variables are shown in **Table 1** by sex and grade. Correlations between the total playfulness score and all measures were calculated for boys and girls separately across grades, and are provided in **Table 2**. They reveal a number of distinct differences between boys and girls in their perceptions of playfulness and their classroom clowning peers. Boys perceived the playful characteristic to relate positively to sociability and to awarding of the class clown label, and not to disruptiveness in the classroom. Girls, in contrast, recognized few associations between these variables and playfulness or class clown qualities. For teachers, playfulness was strongly correlated with all of the outcome variables, however, many showed inverse relationships, in contrast to the opinions held by the children. None of these significant interrelationships presented an impediment to statistical analyses, as HLM accommodates a lack of independence between variables (Raudenbush and Bryk, 2002).

Initial Differences in Playfulness

Kindergarten playfulness data was inspected for initial sex differences to provide some insight into whether the teachers evaluated playfulness differently in the boys and girls when they were of preschool age and began participation in this study. Results confirmed previous findings (Barnett, 1991b) in detecting no sex differences in total playfulness with children of kindergarten age [$t_{(276)} = 1.39, p > 0.05$].

Determining Levels for HLM Analyses

Preliminary analyses utilizing HLM were conducted to explore the influence of classroom- and school- level variability since

TABLE 1 | Means and standard deviations for boys and girls at each grade on all outcome variables.

	BOYS ($n = 135$)						GIRLS ($n = 143$)					
	First grade		Second grade		Third grade		First grade		Second grade		Third grade	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Social competence (SR)a	2.79	0.81	2.91	0.78	2.83	0.64	2.93	0.71	3.11	0.62	3.26	0.55
Social competence (TR)a	2.36	0.65	2.43	0.52	2.12	0.69	2.92	0.58	3.24	0.36	3.57	0.33
Social status (PR)[a]	2.33	0.28	2.49	0.37	2.28	0.54	2.37	0.32	2.41	0.43	2.69	0.64
Disruptive (SR)[b]	1.43	0.84	1.51	0.73	1.63	0.59	1.21	0.51	1.29	0.57	1.14	0.42
Disruptive (PR)[b]	1.18	1.09	1.22	0.97	2.13	0.66	1.03	0.53	1.17	0.79	1.21	0.51
Disruptive (TR)[c]	3.57	0.71	3.72	0.86	4.21	1.02	1.15	0.42	1.09	0.58	1.22	0.71
Class clown (SR)[b]	1.26	0.29	1.28	0.60	1.31	0.47	1.20	0.44	1.18	0.58	1.17	0.34
Class clown (PR)[a]	2.72	0.76	2.59	0.92	3.27	1.18	2.16	0.62	1.38	0.79	1.19	1.08
Class clown (TR)[a]	3.20	0.73	3.14	0.58	3.45	0.62	1.37	0.39	1.21	0.40	1.18	0.58

SR, self-rated; TR, teacher-rated; PR, peer-rated.
[a] *4-point scale.*
[b] *3-point scale.*
[c] *5-point scale.*

TABLE 2 | Partial correlations between kindergarten playfulness and outcome variables for boys (upper diagonal; $n = 135$) and girls (lower diagonal; $n = 143$) across grades.

	1	2	3	4	5	6	7	8	9	10
1. Playfulness	–	013	−001	406**	396**	−295**	289**	308**	361**	402**
2. Disruptive (SR)	009	–	298**	−299**	256*	−317**	043	041	019	283**
3. Disruptive (PR)	−010	225*	–	−258*	097	−306**	−307**	046	155*	362**
4. Disruptive (TR)	018	279**	−236*	–	362**	−405**	−328**	−015	313**	340**
5. Social competence (SR)	107	−105	−102	008	–	−326**	288**	−014	290**	−296**
6. Social competence (TR)	−031	−287**	−405**	−279**	284**	–	−264*	−023	279*	−311**
7. Social status (PR)	156*	−104	−089	−213*	062	−295**	–	088	275*	−288**
8. Class clown (SR)	019	007	061	245*	027	−039	068	–	296**	315**
9. Class clown (PR)	004	095	273**	256*	091	−088	310**	297**	–	−322**
10. Class clown (TR)	−023	276**	288**	278**	046	−294**	335**	023	−288**	–

Decimals omitted; covariates of experience in preschool, number of brothers and sisters, birth order controlled.
SR, self-rated; TR, teacher-rated; PR, peer-rated.
***$p < 0.001$; *$p < 0.01$.*

it can be argued that children enter school varying widely in their abilities, which may differ by their classroom or school (Christian et al., 2001; Raudenbush and Bryk, 2002). The intraclass correlation coefficients (ICC; the portion of the total variance allocated to differences between schools, and between classrooms) revealed they were considered "small" (Hox, 2002), ranging from 0.001 to 0.028 for schools, and 0.002 to 0.017 for classrooms (**Table 3**). In general, these coefficients divulged that at most 3 and 2% of the variance occurred between classrooms and schools, respectively. Therefore, for all outcomes, the vast portion of variance was attributed to individuals within classrooms (ranging from 95 to 99%). HLM analyses for all outcome variables thus proceeded without further testing for classroom or school differences, and these levels were eliminated in subsequent models.

TABLE 3 | Intra-class correlations (ICC) for individual, classroom, and school levels for all outcome measures ($N = 278$).

	Individuals	Classes	Schools
Social competence (SR)	0.987	0.002	0.011
Social competence (TR)	0.969	0.014	0.017
Social status (PR)	0.963	0.015	0.022
Disruptive (SR)	0.991	0.010	0.001
Disruptive (PR)	0.956	0.016	0.028
Disruptive (TR)	0.990	0.008	0.002
Class clown (SR)	0.988	0.010	0.002
Class clown (PR)	0.982	0.015	0.003
Class clown (TR)	0.974	0.017	0.019

SR, self-rated; TR, teacher-rated; PR, peer-rated.

Primary Findings

Social Competence and Status

Children's social competence and status were explored by examining their own assessments as well as those of their peers and teachers. The findings revealed that children's self-perceptions of their social competence were predicted by how playful they were in first and second grades, and sex was a further consideration in third grade (**Table 4A**). Playful boys and girls viewed themselves as more socially competent than their less playful counterparts in first and second grades, while in third grade more playful boys viewed themselves as least socially competent. As they progressed from first to second grade they didn't perceive there to be much change, however, the dramatic downturn was evident in moving from second to third grade for the playful boys.

Classmates provided another perspective on the popularity of more and less playful children. Peers were found to perceive more playful children in the first two grades as higher in social status compared to their less playful peers, with a particularly large distinction shown for boys (**Table 4B**). In third grade, a significant playfulness x sex interaction was found, with more playful boys viewed as lower in social status than all of their classmates. In contrast, no differences in social status as a function of the degree of playfulness were detected for girls in third grade. As children advanced from first to second grade, boys and girls who were regarded to be more playful continued to enjoy higher social status, however, this trend changed with promotion to third grade. In this latter progression, there was a decline in social status for the more playful boys, while their classmates showed an increase.

In stark contrast, teachers viewed more playful children as least socially competent in second and third grades but equivalent to their peers in first grade (**Table 4C**). *Post-hoc* tests for the significant interactions in the two upper grades revealed that more playful boys were judged by teachers as consistently lower in social competence compared to boys who were less playful and all girls. No such distinctions were found for playfulness in girls, and girls were consistently viewed by their teachers as more socially competent than boys in all grades. Promotion from first to second grade witnessed a significant increase in teachers' perceptions of social competence for all but the more playful boys, with a more substantial gain for all girls compared to less playful boys. For more playful boys, no change in teacher ratings of their social competence was evident from first to second grade, and they were the only children who were perceived as declining from second to third grades.

Disruptive Classroom Behavior

To address the first research question, we tested whether playfulness was related to each of the perceptions of the extent to which the child was seen as disruptive in the classroom. The analyses were conducted to examine whether the relationships differed as a function of the child's sex within and across grades (with the child characteristics of birth order, number of siblings, and preschool experience partialed out), and any changes in these scale means across grades. The HLM analyses indicated that self-rated disruptive behavior (**Table 5A**) was unrelated to playfulness, but boys regarded themselves as more disruptive than girls in in all three grades. Peers similarly saw no differences between more and less playful children in the first two grades, and they also thought boys showed more disruptive acts in second grade compared to girls (**Table 5B**). In third grade, however, they appeared attentive to the combination of playfulness and sex, in regarding playful boys as more disruptive than all other children. As children progressed from one grade to the next, this was the only relationship with playfulness noted by peers.

Teachers readily perceived differences in disruptive classroom behavior between more and less playful children, and between boys and girls, in all three grades (**Table 5C**). They consistently viewed less playful boys and all girls as least disruptive, and by third grade this tendency became more pronounced. As children moved from first to second grade, more playful boys were regarded as more disruptive by their teachers compared to their female counterparts and less playful others, whose assessments instead showed no significant change. When progressing to third grade, decreases in teachers' ratings of classroom disruption for almost all children were shown, the exception being playful boys whose ratings continued to increase even more sharply from second to third grades.

TABLE 4 | Hierarchical linear modeling results for social competence and status ($N = 278$).

	First grade *b* (SE)	Second grade *b* (SE)	Third grade *b* (SE)
A. SOCIAL COMPETENCE: SELF-RATED			
Grade (Intercept)			
Mean initial score	12.42** (0.214)	13.79** (0.117)	20.41** (0.082)
Sex	0.00 (0.131)	0.01 (0.028)	0.15* (0.121)
Playfulness	0.16* (0.018)	0.17* (0.070)	0.01 (0.098)
Playfulness x Sex	0.01 (0.112)	0.02 (0.022)	0.14* (0.037)
Change From First to Second Grade (Slope)			
Change in average score	1.51 (0.079)		
Sex	0.00 (0.108)		
Playfulness	0.05 (0.059)		
Playfulness x Sex	0.04 (0.046)		
Change From Second to Third Grade (Slope)			
Mean change in average score	5.15* (0.141)		
Sex	0.00 (0.158)		
Playfulness	0.02 (0.114)		
Playfulness x Sex	0.17* (0.059)		
B. SOCIAL STATUS: PEER-RATED			
Grade (Intercept)			
Initial score	14.18* (0.203)	13.55* (0.099)	13.67* (0.051)
Sex	0.01 (0.098)	0.00 (0.103)	0.14* (0.108)
Playfulness	0.27** (0.060)	0.34** (0.051)	0.02 (0.099)
Playfulness x Sex	0.00 (0.082)	0.02 (0.087)	0.15* (0.031)
Change From First to Second Grade (Slope)			
Mean change in average score	0.97 (0.027)		
Sex	0.02 (0.005)		
Playfulness	0.00 (0.038)		
Playfulness x Sex	0.03 (0.071)		
Change From Second to Third Grade (Slope)			
Mean change in average score	16.24** (0.019)		
Sex	0.02 (0.081)		
Playfulness	0.03 (0.072)		
Playfulness x Sex	0.15* (0.035)		
C. SOCIAL COMPETENCE: TEACHER-RATED			
Grade (Intercept)			
Initial score	22.59** (0.081)	20.34** (0.029)	17.38** (0.067)
Sex	−0.40** (0.114)	−0.09* (0.078)	−0.21** (0.164)
Playfulness	−0.01 (0.032)	−0.11 (0.053)	−0.02 (0.056)
Playfulness x Sex	−0.01 (0.089)	−0.38** (0.042)	−0.37** (0.098)
Change From First to Second Grade (Slope)			
Mean change in average score	13.54* (0.223)		
Sex	−0.01 (0.134)		
Playfulness	−0.02 (0.040)		
Playfulness x Sex	−0.16* (0.072)		
Change From Second to Third Grade (Slope)			
Mean change in average score	19.88** (0.175)		
Sex	−0.15* (0.067)		
Playfulness	−0.16* (0.052)		
Playfulness x Sex	−0.22** (0.059)		

$^{**}p < 0.001$; $^{*}p < 0.01$.

"Class Clown" Designation

Children appeared reluctant to assign the label of "class clown" to themselves, regardless of how playful they were or their sex. There was no distinction found between more and less playful children, or boys and girls, in children viewing themselves as the "class clown" in any grade (**Table 6A**). Playfulness, however, did influence peer perceptions of being a "class clown" for both boys and girls in first grade, but only for boys in second and third grades (**Table 6B**). After first grade, no such relationships were found between girls' playfulness and being seen as a "class clown" and there was a steep decline in girls being regarded as the class clown as children moved through the grades. In second and third grades there was an increasing tendency for boys to be viewed as a class clown compared to girls, particularly those who were more playful.

Teacher designations of playful children as "class clown" were apparent for boys but absent for girls (**Table 6C**). At all grades, more playful boys were predictive of higher teacher scores as a class clown, while no relationship was found with playfulness for girls at any grade. Teachers consistently assigned the class clown moniker to boys more than girls, and to particularly playful ones compared to those who were less playful.

DISCUSSION AND CONCLUSIONS

Summary of Findings

The data are compelling in revealing that playful children are perceived by their teachers and peers very differently than their less playful classmates. In first and second grades, children who were more playful were seen by their classmates as desired playmates, inclined to be ascribed the label of class clown, but not seen as disruptive to themselves or their classroom decorum. In these same grades, the children perceived themselves to be popular among their peers, and adept in social skills. They did not see any of their playful antics as disturbances in the classroom, although they were less hesitant to assign the class clown moniker to themselves. Children did not see playful boys and girls as very different, viewing them all as more preferred play partners to their less playful peers.

In third grade, however, things took a dramatic turn. While children continued to view more and less playful children differently, they now paid careful attention to their gender, and constructed a sharp distinction between playful boys and playful girls. Most significantly, their views of boys who were very playful completely reversed, in that they now came to view them as least preferred playmates with lowest social status. And while they continued to assign them the label of class clown, peers came to view their associated clowning behaviors as disruptive activities in their classroom. In third grade, more playful girls were not any different than girls who were less playful, although the subgroup of playful boys took on their own persona, which was now predominantly negative and contrasted dramatically with how they were seen in the two prior years. The most startling (and alarming) finding was that the children themselves—most notably the playful boys—who shifted to hold increasingly negative perceptions of themselves as well by third grade. Like their peers, they came to view themselves as unpopular, and less

TABLE 5 | Hierarchical linear modeling results for disruptive classroom behaviors ($N = 278$).

	First grade *b* (SE)	Second grade *b* (SE)	Third grade *b* (SE)
A. DISRUPTIVE: SELF-RATED			
grade (intercept)			
Initial score	1.14 (0.083)	2.06 (0.114)	1.81 (0.101)
Sex	0.21* (0.026)	0.19* (0.054)	0.16* (0.027)
Playfulness	0.00 (0.017)	0.03 (0.066)	0.01 (0.089)
Playfulness x Sex	0.04 (0.039)	0.01 (0.109)	0.02 (0.054)
Change From First to Second Grade (Slope)			
Change in average score	1.76 (0.127)		
Sex	0.03 (0.058)		
Playfulness	0.02 (0.043)		
Playfulness x Sex	0.02 (0.067)		
Change From Second to Third Grade (Slope)			
Change in average score	1.95 (0.223)		
Sex	0.01 (0.056)		
Playfulness	0.03 (0.074)		
Playfulness x Sex	0.03 (0.022)		
B. DISRUPTIVE: PEER-RATED			
Initial score	1.42 (0.088)	0.99 (0.071)	1.13 (0.045)
Sex	0.01 (0.029)	0.14* (0.013)	0.15* (0.017)
Playfulness	0.02 (0.041)	0.03 (0.019)	0.02 (0.029)
Playfulness x Sex	0.01 (0.098)	0.10 (0.025)	0.27* (0.021)
Change From First to Second Grade (Slope)			
Change in average score	1.84 (0.036)		
Sex	0.00 (0.107)		
Playfulness	0.01 (0.053)		
Playfulness x Sex	0.01 (0.044)		
Change From Second to Third Grade (Slope)			
Change in average score	13.63* (0.027)		
Sex	0.06* (0.119)		
Playfulness	0.01 (0.065)		
Playfulness x Sex	0.17* (0.058)		
C. DISRUPTIVE: TEACHER-RATED			
Initial score	24.25** (0.037)	24.88** (0.019)	35.47** (0.008)
Sex	0.37** (0.065)	0.41** (0.051)	0.39* (0.022)
Playfulness	0.02 (0.079)	0.02 (0.068)	0.03 (0.089)
Playfulness x Sex	0.43** (0.056)	0.27** (0.061)	0.44** (0.073)
Change From First to Second Grade (Slope)			
Change in average score	1.29 (0.004)		
Sex	0.23** (0.089)		
Playfulness	0.01 (0.091)		
Playfulness x Sex	0.30** (0.083)		
Change From Second to Third Grade (Slope)			
Change in average score	13.76* (0.037)		
Sex	0.02 (0.034)		
Playfulness	0.00 (0.085)		
Playfulness x Sex	0.41** (0.098)		

$^{**}p < 0.001$; $^{*}p < 0.01$.

TABLE 6 | Hierarchical linear modeling results for class clown ratings ($N = 278$).

	First grade *b* (SE)	Second grade *b* (SE)	Third grade *b* (SE)
A. CLASS CLOWN: SELF-RATED			
Grade (Intercept)			
Initial score	0.43 (0.089)	0.91 (0.021)	0.28 (0.060)
Sex	0.01 (0.044)	0.00 (0.019)	0.01 (0.022)
Playfulness	0.01 (0.080)	0.02 (0.042)	0.01 (0.058)
Playfulness x Sex	0.00 (0.074)	0.01 (0.011)	0.03 (0.082)
Change From First TO Second Grade (Slope)			
Change in average score	1.56 (0.104)		
Sex	0.00 (0.106)		
Playfulness	0.02 (0.088)		
Playfulness x Sex	0.00 (0.039)		
Change From Second to Third Grade (Slope)			
Change in average score	1.93 (0.009)		
Sex	0.02 (0.078)		
Playfulness	0.01 (0.061)		
Playfulness x Sex	0.01 (0.047)		
B. CLASS CLOWN: PEER-RATED			
Initial score	14.88* (0.067)	16.90* (0.049)	25.61** (0.012)
Sex	0.15* (0.077)	0.16* (0.018)	0.15* (0.023)
Playfulness	0.29** (0.032)	0.15* (0.026)	0.14* (0.076)
Playfulness x Sex	0.03 (0.039)	0.15* (0.019)	0.33** (0.038)
Change From First to Second Grade (Slope)			
Change in average score	2.32 (0.034)		
Sex	0.00 (0.089)		
Playfulness	0.01 (0.071)		
Playfulness x Sex	0.00 (0.015)		
Change From Second to Third Grade (Slope)			
Change in average score	2.38 (0.019)		
Sex	0.03 (0.060)		
Playfulness	0.01 (0.107)		
Playfulness x Sex	0.16* (0.018)		
C. CLASS CLOWN (TEACHER-RATED)			
Initial score	13.09* (0.112)	14.22* (0.065)	27.96** (0.089)
Sex	0.32** (0.067)	0.37** (0.030)	0.36** (0.049)
Playfulness	0.31** (0.031)	0.32** (0.076)	0.44** (0.072)
Playfulness x Sex	0.28** (0.069)	0.31** (0.022)	0.29** (0.028)
Change From First to Second Grade (Slope)			
Change in average score	1.18 (0.053)		
Sex	0.01 (0.103)		
Playfulness	0.04 (0.067)		
Playfulness x Sex	0.02 (0.039)		
Change From Second to Third Grade (Slope)			
Change in average score	1.92 (0.114)		
Sex	0.01 (0.023)		
Playfulness	0.01 (0.078)		
Playfulness x Sex	0.03 (0.045)		

$^{**}p < 0.001$; $^{*}p < 0.01$.

socially skilled, compared to their classmates. Their perceptions of their classroom behaviors transformed as well, so that they now regarded them to be problematic, which had not been their perspective previously. We strongly suspect that the cause of such a substantial/considerable turnaround is rooted in the eventual influence exerted by teachers, directly and indirectly, on playful boys' self-perceptions and those of their classmates.

Beginning in first grade, teachers showed their distaste for playful boys, consistently viewing them as disruptive in the classroom and as least socially skilled, and assigning them the label of class clown. These perceptions strengthened as children progressed through their three years of school, and while most children were seen as becoming more socially competent across time, playful boys were actually regarded as declining as they approached third grade. In all grades, teachers did not view playful girls as distinct from other children—it was only playful boys that were the focus of their negative perceptions—created and continuing from the earliest school years.

Discussion

Children's Playfulness

One of the most significant discoveries of the study was the antipathy held by teachers for playful boys from the earliest primary grade. In all grades, teachers viewed playful boys as the most disruptive in the classroom, consistently more so than less playful boys, and all girls. At first glance, these results reinforce several streams of research conducted in school settings. In general, female teachers report a closer relationship with girls in their classroom compared to boys (Koepke and Harkins, 2008; Spilt et al., 2012) and primary school teachers have been shown to have more negative and conflictual relationships with boys in their classroom (Hamre and Pianta, 2001; Spilt et al., 2012). At all grades, boys are generally regarded by teachers as disruptive (Jones and Dindia, 2004; Esturgó-Deu and Sala-Roca, 2010; Spilt et al., 2012) and off task more frequently (Kean, 1995) than girls, which might at least partially explain their more antagonistic assessments. In addition, studies have also observed that disruptive behaviors by younger school-aged children are largely directed at teachers (Hall and Hayden, 2007), so that teachers are thus more likely to perceive playful behaviors as distracting and irritating, and in need of intercession. The finding that teachers and children differed in their views of disruptive actions in the first two grades is consistent with studies demonstrating that they often differ in their perceptions of what constitutes disruptive classroom behaviors, and the extent to which they represent a serious intrusion (Mitchell et al., 2010). A number of studies have shown that when they are aware, children don't necessarily regard disruptive classroom behaviors as undesirable or disturbing to others, and in fact might instead view them as engaging or amusing (Huesmann and Guerra, 1997). This could explain the high social status attributed to playful children by their peers in first and second grades.

The finding that it was not just boys, but rather more playful boys in particular, that incurred reactions and stigma from teachers, provides insight into the nature of playfulness for children under constricting conditions. Studies have questioned primary school teachers about the types of students they found to be problematic and those they hoped would leave school (Brophy and Good, 1974). Children who were characterized by their teachers as "aggressive," "impulsive," "impudent," or "lazy," were most frequently identified, and when asked what behaviors typify "difficult" or "problem" students, the vast majority were categorized as disruptive behaviors, with only a few related in any way to learning difficulties (Brophy and Rohrkemper, 1981). Observations in primary classrooms sought to characterize children who were most likely to extract sentiments of preference, concern, or rejection from their grade school teachers (Helton and Oakland, 1977). Results showed that teachers preferred "rigid," "conforming," "orderly," "passive," and "dependent" children, and were much more likely to reject those who were "non-conforming" or "aggressive." Teachers are more likely to condemn behaviors directed toward themselves or other students ("aggressive," "impudent"), particularly incessant or disruptive talking or chattering, disturbing other students, making unnecessary noise, wandering around without permission, avoiding school work, physical aggression against fellow students, and exhibiting rough or wild behavior (Safran and Safran, 1985; Hall and Hayden, 2007). The vast majority of these attributes and behaviors have been found to describe playful young children, and to uniquely distinguish them from those who are less playful (Barnett, 1991b). The qualities that have been found to be discriminating, and the constituent playfulness dimensions that emphasize the physically active, sociable, joking, impulsive, and exuberant predispositions that predominate in young playful children (Lieberman, 1977; Barnett, 1990, 1991a), appear strikingly similar to many of those found to be objectionable or intolerable by teachers. It is thus not surprising that teachers perceived more playful boys to also be more disruptive.

The perceptions of teachers that playful boys were disruptive to classroom tenor and have inferior social skills may forebode a longer-term negative trajectory for them as they move through their formal school years. Research has shown that positive student-teacher relationships relate to fewer disruptive behaviors (Wang et al., 2013), and when interactions with teachers become increasingly negative, classroom disruptions may become more frequent. Positive teacher-child relationships have also been found to be vital for children's feelings of well-being, and academic engagement and performance (Hamre and Pianta, 2006; Hughes et al., 2008). Children who experience supportive, amicable relationships with their teachers have more effectual current and future academic and social outcomes (Hamre and Pianta, 2006). Thus, playful boys who perceive negative affect or criticism by teachers may be at risk. Teachers' perceptions that playful boys have lower social competence, and a lagging rate of social development, may be communicated, and in turn impact their peer relationships (De Laet et al., 2014) and acceptance (Hughes et al., 2001; Hamre and Pianta, 2006). To the extent that this influence becomes internalized by the playful boys or their peers, a negative trajectory might be imminent with the dire longer-term outcome might be that increased problem behaviors will be observed, including delinquency and aggression (Newcomb et al., 1993) with associated feelings of loneliness, depression, and anxiety (Ladd, 2006). Thus,

teachers' negative assessments of playful boys may pose ominous potential consequences for these children's social and academic development and success.

The findings in the present study that peers' assessments of playful boys were largely positive in first and second grades on all measures and then abruptly reversed to a negative course, incites disquieting concerns. These results, coupled with the observation that playful boys regarded themselves as socially skilled in first and second grades and then as socially deficient in third grade, identical to the perceptions of teachers and peers, is strongly suggestive of a social referencing transformation (Hendrickx et al., 2017) taking place. Teachers who hold and project negative attitudes toward playful boys may influence peers to embrace similar opinions by their remarks, gestures, and behaviors (Maas and Meijnen, 1999). The inverted ratings of social status, disruptive classroom behavior, and class clown branding of playful boys by peers in third grade could be construed as evidence that the classroom teacher is a dominant socializing agent who affects children's peer perceptions and relationships, particularly for those of younger age (Farmer et al., 2011). These results support and extend the literature demonstrating that students' observations of the relationships that teachers have with their classmates influences their own perceptions, affective appraisals and responses (De Laet et al., 2014; Hughes and Im, 2016).

The children's own transformed assessment of their social competence may insinuate the presence of a Pygmalion effect (Rosenthal and Jacobson, 1968) in furtherance of the literature revealing the potential that teachers have to influence children based on their beliefs about their attributes and abilities. This "invisible curriculum" (Farmer et al., 2011) focuses on the interactions between students and their teachers, and the potency of teachers' expectations for socializing students as to their behavioral conduct and performance. The essence of this proposition is that teachers form expectations for students based upon their impressions, which may or may not be accurate, and create a self-fulfilling prophesy (Jussim and Harber, 2005). While their impression might be based on a number of factors, one common and potent source is students' behavioral conduct in the classroom (Dusek and Joseph, 1983). Teachers then begin to behave differently toward certain students through the use of consistent verbal and non-verbal cues, in accord with their expectancy (Brophy and Good, 1974). Students will come to respond to this differential treatment and coordinate their behaviors accordingly, eventually internalizing their teachers' expectations (Jussim et al., 2009). This process can function directly on the student in this way, or it may be more indirect if classmates observe the distinctive feedback rendered to playful boys by the teacher. Mirroring the teacher, their classmates may treat them differently as well, and the playful boys will respond to this influence to interpret their own behavior (Brown, 2012) and to guide their future actions (Armstrong, 2011). The finding that it was not until third grade that playful boys acutely perceived and reacted to social changes in how they were viewed by their peers may be attributable to enhanced social-cognitive and socio-emotional abilities as children move through middle childhood. Studies of third graders (and older ages) have shown that children's self-understanding becomes more differentiated and they are more attuned to their social self, and receptive to social information and social comparisons with their peers (Harter, 1998, 2012b; Marsh et al., 1998). Their increasing social knowledge and the predominance assigned to peer relations have endowed them with insight into the social cues, predilections, and actions of their classmates (Rudolph et al., 1995; Harter, 2012b). Advances in perspective-taking ability would also meld so that sensitivity to peers' thoughts and feelings would be heightened with increasing age (Selman, 1980). Perhaps with these heightened abilities, peers are more likely to be aware of playful boys' behaviors and to interpret them as aberrant and/or problematic (at least in the classroom), and playful boys are simultaneously more able to process and assimilate information about peer relations.

Class Clowns

The coincident conscription of the "class clown" label exclusively to playful boys by their teachers strongly suggests that being playful may well be maladaptive in the school classroom. The few studies that have chronicled the behaviors of class clowns have found that they are almost universally perceived by teachers as distracting and problematic, whose behaviors must be managed, shaped, or extinguished (Hobday-Kusch and McVittie, 2002; Ruch et al., 2014; Platt et al., 2016). While the findings of the study are admittedly correlational, we can posit that they are strongly indicative of a linkage between the class clown bestowal and perceptions by classroom teachers that the behaviors of playful boys are unwelcome and objectionable. The further finding that it was exclusively playful boys who were the recipient of the class clown moniker further supports the literature showing a pervasive gender disparity in ascribing this label to youth and adolescents (Fang, 2001; Platt et al., 2016). The concomitant results that playfulness in girls leads to few difficulties with teachers extends this literature to school-aged children. These data do not provide any substantive explanations as to what it is about playful boys or their specific characteristic behaviors that individualizes them from their peers—an issue that awaits empirical study. It is possible that girls respond more readily to early teacher conditioning of appropriate classroom behavior and become more adept at controlling their playful urges in comparison to boys. There is some speculative evidence from this data that this "acclimatizing" may occur to some extent in early primary grades, or to a greater extent as the child enters third grade. It is also possible that boys are more resistant to these efforts, or that their impulse or emotional control is less well developed than girls (Hines, 2004). It is also probable that teachers may shape the behavior of girls as to what is classroom-appropriate before boys at an earlier age (Sax, 2005). These may all be credible explanations for these data—rather than speculation that there are gender differences in the ability to control playful impulses.

Teachers' and peers' ascription of the "class clown" label to playful boys is worrisome in that studies have shown that the way children are labeled comes to demarcate who they are and is a

strong determinant of how they feel about themselves (Becker, 1963). Labels can have a powerful effect on the behaviors and socialization of children, and if this marker is a negative one it can result in the playful boy detaching himself from his peers. If peers come to hold negative beliefs about a playful boy or about being a class clown, they may come to view the playful boy as deviating from the normative social group and exert pressure on him to either conform or conceal his playful attributes (Crocker et al., 1998). Classmates may treat the playful boy differently, or hold inaccurate expectations of him, which could lead to him experiencing social anxiety or isolation (Hirschi, 1969). In this way, the characterization of being "playful" or the designation as a "class clown" has the potential to alter the life course of a child.

Conclusions

This longitudinal study investigated how young children's playfulness, assessed in kindergarten, was predictive of their subsequent social competence, disruptive classroom behaviors, and the designation of "class clown" from first through third grades, viewed from the lens of teachers, classmates, and themselves. The findings enlighten our understanding about playfulness in children in several ways. They extend our knowledge about playfulness to school-aged children, a developmental stage about which there is a paucity of information, and explore its predictive power with data collected over a 3-year time span. The longitudinal design of the study allowed us to reveal the fragility of playfulness in young children as the setting became increasingly rule- and adult- governed, and hence, progressively antithetical to playful expressions. Further, the initial diverging perceptions of playfulness held by teachers and peers eventually converged such that playfulness came to be regarded as deleterious to boys' social relationships and classroom behavior.

The results of this research also contribute to the "class clown" literature in a number of significant ways. The study is the first to directly link the construct of playfulness in children with the existence of the "class clown" marking in the school classroom. Several investigations have utilized one as a part of the definitional criteria for the other, yet without empirical evidence to support and explore their association. By demonstrating their synchronized appearance with the children in the study, we proffer that we have taken a first step toward this end, and in so doing hope to stimulate others to chart a sequential scientific course. In addition, to date the majority of the class clown literature has explored the application of the label to older children and adolescents, with only a few studies conducted with younger aged children. The data assessing awareness and attribution by teachers, peers, and the children themselves revealed that it was viewed in different ways and that a range from positive to negative attitudes were operative. While the "class clown" moniker was evident in all three grades, its valence changed from a positive to negative one, demonstrating its susceptibility to consequences and dissuasion. Lastly, the data reinforced the engendered nature of the class clown branding, with its increasingly exclusive application to boys in progressing from the first to third grade school years.

Limitations and Suggestions for Future Research

This study is one of the first to adopt a longitudinal view of the outcomes of being playful, investigating some select social and behavioral consequences. Previous research on playfulness has identified temperament and personality characteristics of the preschool child, correlates and constitutive dimensions, and parenting styles and demographics particular to the home environment (Barnett and Kleiber, 1982, 1984; Barnett, 1990, 1991a,b; Rogers et al., 1998), yet there has been a lack of research delving into the playful predisposition in school aged children and in different types of settings. While this study has demonstrated that kindergarten playfulness is predictive of social and disruptive classroom behavior in first through third grades, caution must be exercised in adopting a causal interpretation. The implementation of an experimental design in which the playful quality could be facilitated, directed, or discouraged, and resulting social and behavioral effects detected, would enable a causal argument to be made. While the question of whether playfulness can be taught has not been resolved, much less approached, the possibility that it might be susceptible to environmental mediation is plausible. The proposition put forth that through play children's self-regulation and executive function skills are supported and enhanced (Barker et al., 2014), and that class clowns can be taught when their behavior is appropriate and when it is not (Cohen and Fish, 1993), inspires experimental research on playfulness. How playfulness can be encouraged and productively channeled in the classroom is an important question to address, as this research has begun to demonstrate the consequences of being playful for young children.

Future research should objectively chronicle the behaviors of the children in the classroom to determine whether the more playful boys were indeed acting differently than more playful girls or other boys. The negative ratings of classroom behavior for playful boys, viewed by teachers as disruptive, could be verified by objective observational assessments. The finding that peers did not initially regard the classroom behavior of playful peers in the same way as teachers, suggests that teacher expectations for classroom conduct may not have been adequately communicated to students, or that their negative perceptions were not based on tangible readily observable disobedient or mischievous actions. It would be important to determine if playfulness was viewed so negatively by teachers based on actual conduct problems or on stringent behavior expectations that differed for boys and girls. In addition, the assessment of playfulness occurred before the transition to formal schooling, so that observations of the type and extent of playful expression are essential. Evidence to suggest that playfulness has temporal or situational stability is absent, and while the data uncovered relationships with several of the outcome measures, it is crucial to be able to describe what playful behaviors are actually attempted or emitted. The ability to delineate both playful actions and disruptive behaviors in the classroom would advance this line of research considerably.

It is ardently recommended that future research on class clowns consider our procedures in response to the recommendation by Ruch et al. (2014) and Platt et al. (2016) that the construct is best defined and assessed on a continuum, rather

than dichotomously as has been typical in almost all earlier research. The lack of consistency among raters in perceptions of definitional characteristics of class clowns, and/or whether a peer fits those criteria, implies that we adopt new methods and procedures. In addition to measurement issues, in the present study no consideration was given to what these divergent class clown qualities might be, which is an important avenue for further study in that some might be to be more disruptive, or asocial, or aggressive than others. For example, in their study with older children and adolescents, Ruch et al. (2014; Platt et al., 2016) identified disparate types of adolescent class clown behaviors, and clustered these into four types. Only one corresponded to generating disturbances in the classroom, and not all were rated as equally disturbing by teachers. Platt et al. (2016) speculated that class clowning behaviors could differ at different ages, and advocated for longitudinal research to explore how these behaviors might change over time. It would also be enlightening to conduct longer-term studies that follow boys identified early on as class clowns through their school years, to determine how they are able to "survive" efforts to suppress or extinguish their behaviors, and whether those who persist have suffered the ill fates we've hazarded.

The results found in the present study, as well as in others (Farnetti and Palloni, 2010), that class clowns, and perhaps more playful children, are habitually disruptive to school settings, also requires additional detailed scrutiny. While we defined what we meant by this term when the question was posed to teachers and children, it would be informative to discern whether the playful boys who were regarded as such were affected by any conditions that resulted in this perception other than a high playfulness score. For example, children with certain subtypes of ADHD (Cordier et al., 2010), problems self-regulating their behavior (McClelland et al., 2007), and those with poor inhibitory control (Ponitz et al., 2009) have also been consistently identified as disruptive and impulsive in the classroom setting. Our ability to disentangle playfulness from other comorbid conditions is paramount in continuing to hypothesize the existence of this predilection in children.

A further limitation of the study is the inadequacy of the research design to enable exploration of the role of culture, race, or ethnicity in both teacher assessments of playfulness or in classroom disruptiveness in the three primary grades. The literature is compelling in revealing cultural differences in play (Farver and Howes, 1993; Roopnarine et al., 1994; Farver et al., 1995; Farver and Shin, 1997), and in parental beliefs about children's play (cf. Lancy, 2002; Fogle and Mendez, 2006). In the current study, as in previous ones (Lieberman, 1977; Barnett, 1990, 1991b), the small number of children from any non-White ethnic group (42 Black, 3 Hispanic, 9 bi-racial) precluded any statistical testing. In addition to playfulness, there is a substantial and growing body of research about inherent biases of elementary school teachers in viewing their students' classroom behaviors. Studies have demonstrated that White teachers scrutinize Black students more than White students (Gilliam et al., 2016), they rate them as more problematic (Skiba et al., 2011; Gilliam et al., 2016) and disruptive (Thomas et al., 2008), and they impose harsher sanctions (Skiba et al., 2002, 2011; Tenenbaum and Ruck, 2007). The findings that these biases can be seen as early as preschool age (Downer et al., 2016; Gilliam et al., 2016) beseeches playfulness researchers to consider the race of the student and of the teacher in subsequent studies, and to intentionally provide for the inclusion and necessary sample sizes to investigate cultural (disentangled from social class) differences. While acknowledging the small number of Black students in this study, it remains an open question as to whether any differential perceptions of these children existed in teacher ratings of their playfulness, disruptive behaviors, or ascribing the "class clown" descriptor.

In addition to consideration of ethnicity as a salient (and potentially influential) child characteristic, there are critical teacher and classroom qualities that could also play a role and hence would be important to study. Examination of the effects of playfulness on social and behavioral outcomes did not investigate other specific types and levels of influence such as the climate of the classroom, methods of instruction, and personality of the teacher (Bierman, 2011). As there were different teachers both within and across grades, it is likely that the teacher-student relationship varied along with expectations for conduct and how explicit a hierarchy between teacher and children was in place and communicated. Hence, children's degree of classroom disruptiveness or perceptions of social competence may have varied with the characteristics of the setting (teacher, classroom, other students) such that the "antics" of more playful children, and the extent to which gender expectations were in force, may be important considerations. Children might also have been affected by the teacher's warmth or characteristic tendency to show or elicit positive affect, as has been shown to be instrumental in classroom studies (Sabol and Pianta, 2012). Future research more systematically investigating these different levels of influence, and their interaction, is needed to shed additional light on the ways in which playfulness manifests through teacher and peer perceptions and through environmental conditions, or both.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of name of guidelines, name of committee with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Institutional Review Board.

AUTHOR CONTRIBUTIONS

LB designed the study, conducted and supervised data collection and analysis, wrote the manuscript.

ACKNOWLEDGMENTS

Grateful appreciation is extended to the children, parents, teachers, staff, and administrators of the schools that participated in this study, and to the research team members of the latest round of the Families Readiness project.

REFERENCES

Aiken, L. S., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: Sage.

Armstrong, C. (2011). The impact of self and peer perceptual differences on student social behaviour. *Br. J. Educ. Psychol.* 81, 355–368. doi: 10.1348/000709910X522393

Asher, S. R., Singleton, L. C., Tinsley, B. R., and Hymel, S. (1979). A reliable sociometric measure for preschool children. *Dev. Psychol.* 15, 443–444. doi: 10.1037/0012-1649.15.4.443

Barker, J. E., Semenov, A. D., Michaelson, L., Provan, L. S., Snyder, H. R., and Munakata, Y. (2014). Less structured time in children's daily lives predicts self-directed executive functioning. *Front. Psychol.* 5:593. doi: 10.3389/fpsyg.2014.00593

Barnett, L. A. (1990). Playfulness: definition, design, and measurement. *Play Cult.* 3, 319–336.

Barnett, L. A. (1991a). The playful child: measurement of disposition to play. *Play Cult.* 4, 51–74.

Barnett, L. A. (1991b). Characterizing playfulness: correlates with individual attributes and personality traits. *Play Cult.* 4, 371–393.

Barnett, L. A., and Kleiber, D. A. (1982). Concomitants of playfulness in early childhood: cognitive abilities and gender. *J. Genet. Psychol.* 141, 115–127. doi: 10.1080/00221325.1982.10533462

Barnett, L. A., and Kleiber, D. A. (1984). Playfulness and the early play environment. *J. Genet. Psychol.* 144, 153–164. doi: 10.1080/00221325.1984.9923421

Becker, H. S. (1963). *Outsiders: Studies in the Sociology of Deviance*. New York, NY: Free Press of Glencoe.

Bierman, K. L. (2011). The promise and potential of studying the 'invisible hand' of teacher influence on peer relations and student outcomes: a commentary. *J. Appl. Dev. Psychol.* 32, 297–303. doi: 10.1016/j.appdev.2011.04.004

Blair, C. (2002). School readiness: integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *Am. Psychol.* 57, 111–127. doi: 10.1037/0003-066X.57.2.111

Brophy, J. E., and Good, T. (1974). *Teacher-Student Relationships: Causes and Consequences*. New York, NY: Holt, Rinehart, & Winston.

Brophy, J. E., and Rohrkemper, M. M. (1981). The influence of problem ownership on teachers' perceptions of and strategies for coping with problem students. *J. Educ. Psychol.* 73, 295–311.

Brown, J. D. (2012). Understanding the better than average effect: motives (still) matter. *Pers. Soc. Psychol. Bull.* 38, 209–219. doi: 10.1177/0146167211432763

Chafouleas, S. M. (2011). Direct behavior rating: a review of the issues and research in its development. *Educ. Treat. Child.* 34, 575–591. doi: 10.1353/etc.2011.0034

Christ, T. J., Riley-Tillman, T. C., Chafouleas, S. M., and Jaffery, R. (2011). Direct behavior rating: an evaluation of alternate definitions to assess classroom behaviors. *School Psych. Rev.* 40, 181–199.

Christian, K., Bachman, H., and Morrison, F. (2001). "Schooling and cognitive development," in *Environmental Effects on Cognitive Abilities*, eds R. J. Sternberg and R. I. Grigorenko (Mahwah, NJ: Lawrence Erlbaum), 287–335.

Cohen, J. J., and Fish, M. C. (1993). *Handbook of School-Based Interventions: Resolving Student Problems and Promoting Healthy Educational Environments*. San Francisco, CA: Jossey-Bass.

Cordier, R., Bundy, A., Hocking, C., and Einfeld, S. (2010). Comparison of the play of children with attention deficit hyperactivity disorder by subtypes. *Aust. Occup. Ther. J.* 57, 137–145. doi: 10.1111/j.1440-1630.2009.00821.x

Corsaro, W. A., and Eder, D. (1990). Children's peer cultures. *Annu. Rev. Sociol.* 16, 197–220.

Crocker, J., Major, B., and Steele, C. (1998). "Social stigma," in *Handbook of Social Psychology*, *Vol. 2*, eds S. Fiske, D. Gilbert, and G. Lindzey (Boston, MA: McGraw Hill), 504–553.

Damico, S. B., and Purkey, W. W. (1978). Class clowns: a study of middle school students. *Am. Educ. Res. J.* 15, 391–398. doi: 10.3102/00028312015003391

De Laet, S., Doumen, S., Vervoort, E., Colpin, H., Van Leeuwen, K., Goossens, L., et al. (2014). Transactional links between teacher-child relationship quality and perceived versus sociometric popularity: a three-wave longitudinal study. *Child Dev.* 85, 1647–1662. doi: 10.1111/cdev.12216

Diener, M. L., and Kim, D.-Y. (2004). Maternal and child predictors of preschool children's social competence. *J. Appl. Dev. Psychol.* 25, 3–24. doi: 10.1016/j.appdev.2003.11.006

Downer, J. T., Goble, P., Myers, S. S., and Pianta, R. C. (2016). Teacher-child racial/ethnic match within pre-kindergarten classrooms and children's early school adjustment. *Early Child. Res. Q.* 37, 26–38. doi: 10.1016/j.ecresq.2016.02.007

Dunn, T. J., Baguley, T., and Brunsden, V. (2014). From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br. J. Psychol.* 105, 399–412. doi: 10.1111/bjop.12046

Dursley, L., and Betts, L. (2015). Exploring children's perceptions of the perceived seriousness of disruptive classroom behaviours. *Educ. Psychol.* 35, 416–429. doi: 10.1080/01443410.2014.907556

Dusek, J. B., and Joseph, G. (1983). The bases of teacher expectancies: a meta-analysis. *J. Educ. Psychol.* 75, 327–346. doi: 10.1037/0022-0663.75.3.327

Eisenberg, N., Fabes, R. A., Guthrie, I. K., and Reiser, M. (2000). Dispositional emotionality and regulation: their role in predicting quality of social functioning. *J. Pers. Soc. Psychol.* 78, 136–157. doi: 10.1037/0022-3514.78.1.136

Esturgó-Deu, M. E., and Sala-Roca, J. (2010). Disruptive behaviour of students in primary education and emotional intelligence. *Teach. Teach. Educ.* 26, 830–837. doi: 10.1016/j.tate.2009.10.020

Fang, D. (2001). *Class Clowns' Behavior Profiles and Creativity across Sociometric Status*. Unpublished doctoral dissertation, University of Georgia, Athens, GA.

Farmer, T. W., McAuliffe, M., and Hamm, J. V. (2011). Revealing the invisible hand: the role of teachers in children's peer experiences. *J. Appl. Dev. Psychol.* 32, 247–256. doi: 10.1016/j.appdev.2011.04.006

Farnetti, A., and Palloni, F. (2010). Clowning: its effects on self-imagery and interpersonal relationships in nursery school. *Proced. Soc. Behav. Sci.* 5, 1527–1532. doi: 10.1016/j.sbspro.2010.07.320

Farver, J. A., and Howes, C. (1993). Cultural differences in American and Mexican mother-child pretend play. *Merrill Palmer Q.* 39, 344–358.

Farver, J. A., and Shin, Y. L. (1997). Social pretend play in Korean- and Anglo-American preschoolers *Child Dev.* 68, 544–556. doi: 10.2307/1131677

Farver, J. A., Kim, Y. K., and Lee, Y. (1995). Cultural differences in Korean- and Anglo-American preschoolers' social interaction and play behaviors. *Child Dev.* 66, 1088–1099.

Field, A. (2013). *Discovering Statistics Using SPSS, 4th Edn*. London: Sage.

Fogle, L. M., and Mendez, J. L. (2006). Assessing the play beliefs of African American mothers with preschool children. *Early Child. Res. Q.* 21, 507–518. doi: 10.1016/j.ecresq.2006.08.002

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., and Shic, F. (2016). *Do Early Educators' Implicit Biases Regarding Sex and Race Relate to Behavior Expectations and Recommendations of Preschool Expulsions and Suspensions?* New Haven, CT: Yale University Child Study Center.

Hall, N., and Hayden, C. (2007). Is 'hate crime' a relevant and useful way of conceptualizing some forms of school bullying? *Int. J. Violence Sch.* 3, 3–24.

Hamre, B. K., and Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child Dev.* 72, 625–638. doi: 10.1111/1467-8624.00301

Hamre, B. K., and Pianta, R. C. (2006). "Student-teacher relationships," in *Children's Needs III: Development, Prevention, and Intervention*, eds G. G. Bear and K. M. Minke (Bethesda, MD: National Association of School Psychologists), 59–72.

Harter, S. (1982). The perceived competence scale for children. *Child Dev.* 53, 87–97. doi: 10.2307/1129640

Harter, S. (1985). *Manual for the Perceived Competence Scale for Children.* Denver, CO: University of Denver Press.

Harter, S. (1998). "The development of self-representations," in *Social, Emotional, and Personality Development, 5th Edn, Vol. 3*, ed N. Eisenberg (New York, NY: Wiley), 553–618.

Harter, S. (2012a). Self-*Perception Profile for Children: Manual and Questionnaires*. Denver, CO: University of Denver.

Harter, S. (2012b). *The Construction of the Self, 2nd Edn*. New York, NY: Wiley.

Helton, G. B., and Oakland, T. D. (1977). Teachers' attitudinal responses to differing characteristics of elementary school students. *J. Educ. Psychol.* 69, 261–265.

Hendrickx, M.M.H.G., Mainhard, T., Boor-Klip, H. J., and Brekelmans, M. (2017). Our teacher likes you, so I like you: a social network approach to social referencing. *J. Sch. Psychol.* 63, 35–48. doi: 10.1016/j.jsp.2017.02.004

Hines, M. (2004). *Brain Gender*. New York, NY: Oxford University Press.

Hirschi, T. (1969). *Causes of Delinquency*. Berkeley, CA: University of California Press.

Hobday-Kusch, J., and McVittie, J. (2002). Just clowning around: classroom perspectives on children's humour. *Can. J. Educ.* 27, 195–210. doi: 10.2307/1602220

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75, 800–802. doi: 10.1093/biomet/75.4.800

Hox, J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Erlbaum.

Huesmann, L. R., and Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behaviour. *J. Pers. Soc. Psychol.* 72, 408–419.

Hughes, F. P. (2010). *Children, Play, and Development, 4th Edn*. Thousand Oaks, CA: Sage.

Hughes, J. N., and Im, M. H. (2016). Teacher-student relationship and peer disliking and liking across grades 1-4. *Child Dev.* 87, 593–611. doi: 10.1111/cdev.12477

Hughes, J. N., Cavell, T. A., and Wilson, V. (2001). Further support for the developmental significance of the quality of teacher-student relationships. *J. Sch. Psychol.* 39, 281–301. doi: 10.1016/S0022-4405(01)00074-7

Hughes, J. N., Im, M. H., and Wehrly, S. E. (2014). Effect of peer nominations of teacher student support at individual and classroom levels on social and academic outcomes. *J. Sch. Psychol.* 52, 309–322. doi: 10.1016/j.jsp.2013.12.004

Hughes, J. N., Luo, W., Kwok, O., and Loyd, L. (2008). Teacher–student support, effortful engagement, and achievement: a three-year longitudinal study. *J. Educ. Psychol.* 100, 1–14. doi: 10.1037/0022-0663.100.1.1

Huysamen, G. (2007). Coefficient alpha: unnecessarily ambiguous; unduly ubiquitous. *J. Indus. Psychol.* 32, 34–40. doi: 10.4102/sajip.v32i4.242

Johnson, A. H., Miller, F. G., Chafouleas, S. M., Welsh, M. E., Chris Riley-Tillman, T., and Fabiano, G. (2016). Evaluating the technical adequacy of DBR-SIS in tri-annual behavioral screening: a multisite investigation. *J. Sch. Psychol.* 54, 39–57. doi: 10.1016/j.jsp.2015.10.001

Jones, S. M., and Dindia, K. (2004). A meta-analytic perspective on sex equity in the classroom. *Rev. Educ. Res.* 74, 443–471. doi: 10.3102/00346543074004443

Jussim, L. (1986). Self-fulfilling prophecies: a theoretical and integrative review. *Psychol. Rev.* 93, 429–445. doi: 10.1037/0033-295X.93.4.429

Jussim, L., and Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: knowns and unknowns, resolved and unresolved controversies. *Pers. Soc. Psychol. Rev.* 9, 131–155. doi: 10.1207/s15327957pspr0902_3

Jussim, L., Robustelli, S. L., and Cain, T. R. (2009). "Teacher expectations and self-fulfilling prophesies," in *Handbook of Motivation at School*, eds K. Wenzel and A. Wigfield (New York, NY: Routledge/Taylor & Francis), 349–380.

Kean, J. (1995). Teacher ethnotheories and child temperament: impact on classroom interactions. *Austr. J. Early Educ.* 22, 13–18.

Koepke, M. F., and Harkins, D. A. (2008). Conflict in the classroom: gender differences in the teacher-child relationship. *Early Educ. Dev.* 19, 843–864. doi: 10.1080/10409280802516108

Ladd, G. W. (2006). Peer rejection, aggressive or withdrawn behavior, and psychological maladjustment from ages 5 to 12: an examination of four predictive models. *Child Dev.* 77, 822–846. doi: 10.1111/j.1467-8624.2006.00905.x

Lancy, D. F. (2002). "Cultural constraints on children's play," in *Conceptual, Social-cognitive and Contextual Issues in the Fields of Play*, ed J. L. Roopnarine (Westport, CT: Ablex Publishing), 53–60.

Lieberman, J. N. (1966). Playfulness: an attempt to conceptualize a quality of play and of the player. *Psychol. Rep.* 19:1978. doi: 10.2466/pr0.1966.19.3f.1278

Lieberman, J. N. (1977). *Playfulness: Its Relationship to Imagination and Creativity*. New York, NY: Academic Press.

Lix, L. M., and Sajobi, T. (2010). Testing multiple outcomes in repeated measures designs. *Psychol. Methods* 15, 268–280. doi: 10.1037/a0017737

Maas, C. J. M., and Meijnen, G. W. (1999). Problem students: a contextual phenomenon? *Soc. Behav. Pers.* 27, 387–406. doi: 10.2224/sbp.1999.27.4.387

Marsh, H. W., Craven, R., and Debus, R. (1998). Structure, stability, and development of young children's self-concepts: a multicohort-multi-occasion study. *Child Dev.* 69, 1030–1053.

Matthews, J. S., Ponitz, C. C., and Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *J. Educ. Psychol.* 101, 689–704. doi: 10.1037/a0014240

McClelland, M. M., Acock, A. C., and Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Child. Res. Q.* 21, 471–490. doi: 10.1016/j.ecresq.2006.09.003

McClelland, M. M., Cameron, C. E., Wanless, S., and Murray, A. (2007). "Executive function, behavioral self-regulation, and social-emotional competence: Links to school readiness," in *Contemporary Perspectives in Early Childhood Education: Social Learning in Early Childhood Education*, *Vol. 7*, eds O. N. Saracho and B. Spodek (Greenwich, CT: Information Age), 113–137.

McClelland, M. M., Morrison, F. J., and Holmes, D. L. (2000). Children at risk for early academic problems: the role of learning-related social skills. *Early Child. Res. Q.* 15, 307–329. doi: 10.1016/S0885-2006(00)00069-7

McKown, C., Gregory, A., and Weinstein, R. S. (2010). "Expectations, stereotypes, and self-fulfilling prophecies in classroom and school life," in *Handbook of Research on Schools, Schooling, and Human Development,* eds J. Meece and J. Eccles (New York, NY: Routledge/Taylor & Francis), 256–274.

Mitchell, M. M., Bradshaw, C. P., and Leaf, P. J. (2010). Student and teacher perceptions of school climate: a multilevel exploration of patterns of discrepancy. *J. Sch. Health* 80, 271–279. doi: 10.1111/j.1746-1561.2010.00501.x

Moore, N. V., Evertson, C. M., and Brophy, J. E. (1974). Solitary play: some functional reconsiderations. *Dev. Psychol.* 10, 830–834. doi: 10.1037/h0037257

Muris, P., Meesters, C., and Figen, P. (2003). The Self-perception profile for children: further evidence for its factor structure, reliability, and validity. *Pers. Individ. Dif.* 35, 1791–1802. doi: 10.1016/S0191-8869(03)00004-7

Newcomb, A. F., Bukoswki, W. M., and Pattee, L. (1993). Children's peer relations: a meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychol. Bull.* 113, 99–128.

Perry, K. E., and Weinstein, R. S. (1998). The social context of early schooling and children's school adjustment. *Educ. Psychol.* 33, 177–194.

Petriwskyj, A., Thorpe, K., and Tayler, C. (2005). Trends in construction of transition to school in three western regions, 1990–2004. *Int. J. Early Years Educ.* 13, 55–69. doi: 10.1080/09669760500048360

Platt, T., Wagner, L., and Ruch, W. (2016). The association between class clown dimensions, school experiences and accomplishment. *Learn. Individ. Differ.* 51, 327–332. doi: 10.1016/j.lindif.2016.08.036

Ponitz, C. C., McClelland, M. M., Matthews, J. S., and Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Dev. Psychol.* 45, 605–619. doi: 10.1037/a0015365

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models: Application and Data Analysis Methods*. Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., and du Toit, M. (2011). *HLM 7: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International.

Riley-Tillman, T. C., Chafouleas, S. M., Christ, T. J., Briesch, A. M., and LeBel, T. J. (2009). The impact of item wording and behavioral specificity on the accuracy of Direct Behavior Rating (DBR). *Sch. Psychol. Q.* 24, 1–12. doi: 10.1037/a0015248

Rogers, C. S., Impara, J. C., Frary, R. B., Harris, T., Meeks, A., Semanic-Lauth, S., et al. (1998). Measuring playfulness: development of the Child Behavior Inventory of Playfulness. *Play Cult.* 4, 121–135.

Roopnarine, J., Johnson, J., and Hopper, H. (1994). *Children's Play in Diverse Cultures*. Albany, NY: State University of New York Press.

Rosenthal, R., and Jacobson, L. (1968). *Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development*. New York, NY: Rinehart and Winston.

Rothbart, M. L., and Bates, J. E. (2006). "Temperament," in *Handbook of Child Psychology, Vol. 3, Social, Emotional, and Personality Development*, eds W. Damon and R. M. Lerner (Hoboken, NJ: John Wiley & Sons, Inc.), 99–166.

Rubin, K. H., Bukowski, W. M., and Parker, J. G. (2006). "Peer interactions, relationships, and groups," in *Handbook of Child Psychology:* Vol. 3, *Social, Emotional, and Personality Development*, eds. W. Damon and R. M. Lerner (Hoboken, NJ: John Wiley & Sons), 571–645.

Ruch, W., Platt, T., and Hofmann, J. (2014).The character strengths of class clowns. *Front. Psychol.* 5:1075. doi: 10.3389/fpsyg.2014.01075

Rudolph, K. D., Hammen, C., and Burge, D. (1995). Cognitive representations of self, family, and peers in school-age children: links with social competence and sociometric status. *Child Dev.* 66, 1385–1402. doi: 10.2307/1131653

Sabol, T. J., and Pianta, R. C. (2012). Recent trends in research on teacher-child relationships. *Attach. Hum. Dev.* 14, 213–231. doi: 10.1080/14616734.2012.672262

Safran, S. P., and Safran, J. S. (1984). Elementary teachers' tolerance of problem behaviors. *Elem. Sch. J.* 85, 237–243.

Safran, S. P., and Safran, J. S. (1985). Classroom context and teachers' perceptions of problem behaviors. *J. Educ. Psychol.* 77, 20–28.

Sax, L. (2005). *Why Gender Matters*. New York, NY: Broadway Books.

Selman, R. L. (1980). *The Growth of Interpersonal Understanding*. New York, NY: Academic Press.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 74, 107–120. doi: 10.1007/s11336-008-9101-0

Singer, J. L., Singer, D. G., and Sherrod, L. (1980). A factor analytic study of preschooler's play behavior. *Am. Psychol. Bull.* 2, 143–156.

Skiba, R. J., Horner, R. H., Chung, C. G., Rausch, M., May, S. L., and Tobin, T. (2011). Race is not neutral: a national investigation of African American and Latino disproportionality in school discipline. *Sch. Psychol. Rev.* 40, 85–107.

Skiba, R. J., Michael, R. S., Nardo, A. C., and Peterson, R. L. (2002). The color of discipline: sources of racial and gender disproportionality in school punishment. *Urban Rev.* 34, 317–342. doi: 10.1023/A:1021320817372

Spilt, J. L., Koomen, H., and Jak, S. (2012). Are boys better off with male and girls with female teachers? A multilevel investigation of measurement invariance and gender match in teacher–student relationship quality. *J. Sch. Psychol.* 50, 363–378. doi: 10.1016/j.jsp.2011.12.002

Stipek, D. J. (1981). Children's perceptions of their own and their classmates' ability. *J. Educ. Psychol.* 73, 404–410.

Stipek, D. J., and Tannatt, L. M. (1984). Children's judgments of their own and their peers' academic competence. *J. Educ. Psychol.* 76, 75–84.

Suchodoletz, A. V., Trommsdorff, G., Heikamp, T., Wieber, F., and Gollwitzer, P. (2009). Transition to school: the role of kindergarten children's behavior regulation. *Learn. Individ. Differ.* 19, 561–566. doi: 10.1016/j.lindif.2009.07.006

Tenenbaum, H. R., and Ruck, M. D. (2007). Are teachers' expectations different for racial minority than European American students? A meta-analysis. *J. Educ. Psychol.* 99, 253–273. doi: 10.1037/0022-0663.99.2.253

Thomas, D. E., Bierman, K. L., Thompson, C., and Powers, C. J. (2008). Double jeopardy: Child and school characteristics that predict aggressive-disruptive behavior in first grade. *Sch. Psychol. Rev.* 37, 516–532.

Trentacosta, C. J., and Izard, C. E. (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion* 7, 77–88. doi: 10.1037/1528-3542.7.1.77

Trevlas, E., Grammatikopoulos, V., Tsigilis, N., and Zachopoulou, E. (2003). Evaluating playfulness: construct validity of the Children's Playfulness Scale. *Early Childhood Res. J.* 31, 33–39.

Vitaro, F., Brendgen, M., Larose, S., and Tremblay, R. E. (2005). Kindergarten disruptive behaviors, protective factors, and educational achievement by early adulthood. *J. Educ. Psychol.* 97, 617– 629. doi: 10.1037/0022-0663.97.4.617

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol. Methods* 17, 228–243. doi: 10.1037/a0027127

Wang, M.-T., Brinkworth, M., and Eccles, J. (2013). Moderating effects of teacher–student relationship in adolescent trajectories of emotional and behavioral adjustment. *Dev. Psychol.* 49, 690–705. doi: 10.1037/a0027916

Weinstein, R. (2002). *Reaching Higher: The Power of Expectations in Schooling*. Cambridge, MA: Harvard University Press.

Weinstein, R. S., Marshall, H. H., Sharp, L., and Botkin, M. (1987). Pygmalion and the student: age and classroom differences in children's awareness of teacher expectations. *Child Dev.* 58, 1079–1093. doi: 10.2307/1130548

Yarrow, M. R., Waxler, C. Z., and Scott, P. M. (1971). Child effects on adult behavior. *Dev. Psychol.* 5, 300–311. doi: 10.1037/h0031434

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Barnett. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 19 December 2017
doi: 10.3389/fpsyg.2017.02214

The Relation between Teachers' and Children's Playfulness: A Pilot Study

*Shulamit Pinchover**

The Paul Baerwald School of Social Work and Social Welfare, Hebrew University of Jerusalem, Jerusalem, Israel

Young children spend considerable time in educational settings, in which traditionally, their primary occupation is play. A playful preschool environment has been related to better cognitive, social and emotional development. Although it is assumed that teachers' playful behaviors are important in creating a playful school environment, empirical knowledge on this subject is lacking. The current study pilot examines the relation between teachers' and children's playfulness. Thirty-one teacher–child dyads participated. The teachers were asked to complete the Adult Playfulness Scale (APS). Thirty-minute videotapes of teacher–child play interactions were used to evaluate the child's playfulness using the Test of Playfulness. A positive relation was found between two of the APS subscales (spontaneity and silliness) and child playfulness. Teacher silliness mediated the relation between children's age and playfulness. This study is the first to show that teachers' playfulness aspects are related to higher playfulness in children. Promoting teachers' playful behaviors can be related to better teacher–child playful interactions, thereby enhancing children's playfulness.

Keywords: playfulness, teachers, children, early childhood, teacher–child interaction

OPEN ACCESS

Edited by:
René T. Proyer,
Martin Luther University of Halle-Wittenberg, Germany

Reviewed by:
Lynn A. Barnett,
University of Illinois at Urbana–Champaign, United States
Nicola Whitton,
Manchester Metropolitan University, United Kingdom

***Correspondence:**
Shulamit Pinchover
shulamit.pinchover@mail.huji.ac.il

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *05 September 2017*
Accepted: *06 December 2017*
Published: *19 December 2017*

Citation:
Pinchover S (2017) The Relation between Teachers' and Children's Playfulness: A Pilot Study. Front. Psychol. 8:2214. doi: 10.3389/fpsyg.2017.02214

INTRODUCTION

Young children spend considerable time in educational settings, in which traditionally, their primary occupation is play (Wong and Logan, 2016; Pyle et al., 2017). Play and playfulness are considered basic features of early childhood education (ECE) and have been related to social, emotional, and cognitive development (Fisher, 1992; Frost et al., 2001; Youell, 2008). A playful school environment has been related to higher child involvement in play, and to improved learning and development (Jones and Reynolds, 2015). However, despite the extensive literature on how to create a playful educational environment (Jones and Reynolds, 2015), and the common intuitive assumption regarding the importance of teachers' playful behaviors in promoting a playful climate in class and developing children's play and playfulness, empirical knowledge on the subject is lacking (Singer, 2013). First, adults' and specifically teachers' playfulness is less studied compared to children's (Proyer, 2012). Additionally, the relationship between teachers' and children's playfulness has hardly been examined empirically, to the best of our knowledge.

The current study is part of a broader research project on play interactions between children and adults at home and in educational settings. Specifically, it aims to examine the relation between aspects of teachers' perceived playfulness (including spontaneous, expressive, fun, creative, and silly behaviors) and children's observed playfulness in ECE settings.

Play in Early Childhood Education

Over the years, several attempts have been made to understand the role of play in ECE, based on extensive theoretical and empirical literature supporting its significance for young children's development and well-being (for a review, see Johnson et al., 2013). There is substantial evidence that through play and playfulness children demonstrate improved verbal and social communication, high levels of interaction skills, creativity, imagination, divergent thinking, and problem-solving skills (see review at Wood and Attfield, 2005). Playful educational environment was found to foster creativity and imagination along with academic achievements (Kangas, 2010; Kangas and Ruokamo, 2012). Specifically, playful ECE settings are considered helpful in developing young children's play. Although it is almost axiomatic that play is a cornerstone in ECE (Pyle et al., 2017), in recent years it seems that play has been sidelined by early learning standards and assessments of academic attainments (Roskos and Christie, 2007; Bodrova, 2008; Wisneski and Reifel, 2012; Dickey et al., 2016). This is increasingly so as the children grow older. This trend, at least in the developed world, is so acute, in fact, that lately play interventions have been increasingly attempted in order to teach teachers and children to create a playful and imaginative world together (Lobman, 2003, 2006).

The teacher is considered to play a significant role in creating a playful environment and developing children's play (Pramling Samuelsson and Johansson, 2009; Johnson et al., 2013). Teachers' role includes planning of the setting for play, using a playful pedagogic approach, and engaging with children in play (Wood, 2008). Teachers' involvement in play interactions can increase the frequency, duration, and complexity of children's play (McAfee and Leong, 2010). Vygotsky (1978) emphasizes the active role teachers have in children's play. However, there is a debate on how the teacher should be involved: as a co-player, an instructor or a supervisor (Jones and Reynolds, 2015). Most scholars would agree that teachers need to create a playful classroom environment and allow spontaneous child–child play activity, but also be skillful play partners themselves (Ashiabi, 2007; Childress, 2010). For example, a study with Dutch teachers found that teachers' higher involvement in play interaction is related to children's higher involvement in play (Singer et al., 2014). In real life, however, teachers are often too busy with classroom management to be available to promote play, or simply do not know how to do that (Aras, 2016; Elliott and Jarneman, 2017).

Playfulness

Playfulness is defined as the disposition to engage in play (Barnett, 1991), and is considered a personality trait that exists and is expressed across the lifespan (Lieberman, 1977). According to Lieberman (1977), each child or person has a different playfulness style, affected by personality as well as environmental characteristics, such as those of home and educational setting. Proyer (2012) found it to have significant role in both children and adults' lives. Recent studies have examined the lifetime flexibility of playfulness and the ability to enhance it using various interventions (Bundy et al., 2008). Results indicate that it is responsive to intervention and can be change over time (Okimoto et al., 2000; Case-Smith, 2013; Fabrizi, 2014).

Children's Playfulness

The way children play can be captured in different ways; one of them is through playfulness (Bundy, 1997; Cornelli Sanderson, 2010). Playfulness not only captures the mechanism of play, but also addresses the child's general approach to play (Bundy, 1997). Playfulness is composed of four dimensions: (1) the child's internal motivation independently of external expectations; (2) internal control – the child's ability to determine or direct the play action; (3) the freedom to suspend reality in play; and (4) framing – the child's ability to communicate and interpret social cues. Playfulness has been related to children's social, emotional and cognitive development and well-being (Youell, 2008). For example, it is significantly related to active coping, affective regulation, and willingness to express emotions (Christian, 2012).

When children play, they learn about reality and ways of affecting and manipulating it. Being playful means being free to create roles and activities, regardless of external constraints. Children's playful behavior is guided by an internal motivation for a process with self-imposed goals, with a tendency to attribute their own meanings to objects and behaviors (Rubin et al., 1983). Those characteristics of playfulness help children learn, be creative and cope with difficulties (Youell, 2008).

Teachers' Playfulness

While the literature on adults' playfulness has been growing in recent years, it is still understudied compared to children's playfulness (Proyer, 2012). Individuals who are playful are typically funny, humorous, spontaneous, and are more likely to act in a playful manner by joking, teasing, clowning, and being silly. Existing studies have shown that adults' playfulness is related to well-being, sense of happiness, relationship satisfaction, and higher self-estimates of ingenuity and creativity (Proyer, 2012, 2013, 2014; Bateson et al., 2013; Yue et al., 2016). Specifically, research indicates that playfulness is positively related to both job satisfaction and job performance (Yu et al., 2007). Two recent studies have focused on parental playfulness. A study from 32 young adult's perspective reported a positive relation between parents' playfulness and children's adoptive behavior (Shen et al., 2017). Similarly, Menshe-Grinberg and Atzaba-Poria (2017) found that parental playfulness moderate the relation between parental behaviors and child's negativity.

However, the playfulness of educational professionals has rarely been investigated. Based on a small exploratory study of 16 teachers (Lieberman, 1974), and on other qualitative research observations in a natural environment, Lieberman (1977) emphasized the importance of teachers' playfulness in relation to children's play, playfulness and divergent thinking. She concluded that through playfulness behaviors, the teacher could create an environment that enables children to express greater joy and be more creative and flexible in play. The few other studies available show that teacher playfulness is related to teacher–child interaction and relationship (Graham et al., 1989; Tegano et al., 1999; McMillan, 2017). For example, a qualitative case study of two ECE teachers showed that teachers' playful

behavior could alleviate toddlers' emotional distress and help children in transitions and that playfulness is a constructive way to build a secure relationship between toddlers and teachers (Jung, 2011). Another study examined Lithuanian and Greek teachers' perceptions regarding playfulness in their kindergarten class and found that different teacher found different ways to engage in playful behaviors in class. While teachers in Lithuania believed that their role as adults was to promote playfulness and a playful atmosphere, Greek teachers paid more attention to modeling playfulness themselves (Synodi et al., 2015). Teachers' engagement in playful learning environment was also found to be related to student satisfaction from learning (Kangas et al., 2017). Finally, a longitudinal evaluation of playful curricula found that absence of teacher's playfulness was often associated with lower levels of child engagement in play and activities in general (Walsh et al., 2011).

Following these studies, we assume that teachers' playfulness influences preschoolers' play behavior. The current pilot study addresses the empirical gap in our knowledge on the relation between teachers' and children's playfulness. The research hypothesis is that a positive relation will be found between teachers' and children's observed playfulness.

MATERIALS AND METHODS

Sample

Thirty-one teacher–child dyads participated in the study. The children were all typically developed and their ages ranged between 40 and 72 months (M = 54.47, SD = 10.01). Four girls and 27 boys participated [this is due to the nature of the sample, being part of a larger research project that included children with typical development and with autistic spectrum disorder (ASD), which is more common in boys]. The children were all from mid-upper socioeconomic class families. The teachers – all women – were aged 24 to 57 (M = 40.97, SD = 10.43). All teachers had a degree in education and a teaching certificate. All children learned in urban public preschools and kindergartens supervised by of the Israeli Ministry of Education. All teachers knew the children they were playing with at least 3 months before data collection.

Measures

Background Characteristics

Teachers were asked to report their age and education level. In addition, they were asked to report the child's age and their length of acquaintance (in months).

Developmental Assessment

In order to make sure all children were typically developed, they passed a developmental assessment using the Mullen Scales of Early Learning (MSEL; Mullen, 1995) or the Wechsler Preschool and Primary Scales of Intelligence-Revised (WPPSI-R; Wechsler, 1989), depending on their chronological age. The MSEL is a standardized developmental test for children aged 3 to 68 months, consisting of five subscales: gross motor, fine motor, visual reception, expressive language, and receptive language. It provides separate standard verbal and non-verbal summary mental age (MA) scores. Commonly used and well-validated, the MSEL was translated and standardized in Israel (Ben-Sasson et al., 2007). The WPPSI-R is an intelligence test designed for children aged 2.5 to 7.25 years. It provides verbal, performance, and full-scale IQ scores, converted in the current study into an MA score. It is a well-known measure translated and normed in Israel (Pilowsky et al., 1998).

Teachers' Playfulness

Teachers were asked to complete a background questionnaire as well as the Adult Playfulness Scale (APS; Glynn and Webster, 1992, 1993). The APS is a list of 32 adjectives which are scored on a seven-point scale; five additional facets of adult playfulness may be evaluated (Bozionelos and Bozionelos, 1999; Proyer, 2011). These facets are spontaneous (the alpha-coefficient in this sample was 0.74), expressive (e.g., bouncy vs. staid; α = 0.76), fun (e.g., bright vs. dull; α = 0.75), creative (e.g., imaginative vs. unimaginative; α = 0.78), and silly (e.g., childlike vs. mature; α = 0.79). Glynn and Webster (1992) report satisfactory internal consistencies and test–retest correlations, and a robust factor solution for their instrument. The APS was translated into Hebrew and retranslated for the current study.

Children's Playfulness

The Test of Playfulness (ToP; Bundy, 1997) was used to evaluate children's playfulness level. The ToP is an observation-based assessment of the playfulness of children between the ages of 6 months and 18 years. It consists of 29 items (e.g., "Engaged in social play"; "Incorporates objects or other people into play in unconventional or variable ways") scored on three Likert scales: (1) Extent (0 = *rarely or never*, 3 = *almost always*); (2) Intensity (0 = *not*, 3 = *highly*); and (3) Skill (0 = *unskilled*, 3 = *highly skilled*). In our study, one of the items ("Enters a group already engaged in an activity") was eliminated because it was inappropriate for dyadic play. A complete list of ToP items and examples may be found in Pinchover et al. (2016).

Scoring the ToP utilizes a test-specific keyform, which plots the relative difficulty of each item against the means and standard deviations for all items, and produces a total score ranging from 7 to −7, subsequently translated into a 0–3 score (Bundy et al., 2001).

In addition, Rasch analysis was performed using a "fit-the-model" methodology (Andrich, 1988). Rasch analysis uses a probability model to estimate personal "ability" and item "difficulty" by "comparing the response patterns of individuals to the entire sample" (Duncan et al., 2003, p. 951). Called "logits," these equal-interval measures reflect the participant's ability to perform a particular task.

All videotapes were coded by three trained and reliable coders. Inter-coder reliability was established for 15% of the videotapes using inter-class correlations (Koch, 1982), and ranged between 0.70 and 0.79.

Procedure

As mentioned, the data were collected as part of a larger study that investigated play interactions of children with and without

ASD at home and educational settings. The study was approved by the Ethics Committee of the Hebrew University of Jerusalem and the Israel Ministry of Education.

First, the research team contacted teachers in preschools and kindergartens for typically developed children that are supervised by the ministry. Teachers who agreed to participate were asked to write to all the parents in the preschool or kindergarten, explaining about the research and asking their permission for their children to participate. The children who were permitted to participate and who met our age and gender criteria were included in this study. Each teacher was paired to the first child/two children from her classroom included in the study.

Data were collected during the second and third quarters of the school year, in order to let the teachers get to know the children and establish their relationship before the research. A certified psychologist administered the developmental assessment to each child at their home. Parent who were interested in getting the test result, could receive a formal report and meet with the head psychologist for discussion. Each child–teacher dyad was videotaped in a 30-min play interaction at school, playing "as you usually do." The dyad was offered to play with developmentally appropriate and attractive toys (e.g., books, blocks, puzzles, dolls, cars) provided by the researchers. Finally, teachers were asked to complete a demographic questionnaire and the APS.

Data Analysis

First, descriptive statistics and bivariate Pearson's correlations between research variables were calculated. Next, a mediation model was tested to investigate the indirect link between children's chronological age (CA) and playfulness as mediated by teachers' playfulness aspects. The mediation analysis was conducted using PROCESS macro for SPSS (Models 4; Preacher and Hayes, 2008; Hayes, 2009), which enables examination of mediation models on small samples, using the bootstrapping method. The analysis provides bootstrapped confidence intervals (CIs) for the conditional effects; when the model is significant, 0 will not be included in the CI, and the CI will be 95%.

RESULTS

Descriptive and Bivariate Statistics

Table 1 presents means, standard deviations, and inter-correlations for all variables. Research hypotheses were examined using *t*-tests and Pearson correlations. Child's gender and child's playfulness as well as teacher's playfulness existed independently from each other ($U = 35.00$, $p = 0.35$ and $U = 38.00$, $p = 0.17$, respectively). No relationship was found between children's chronological or mental age (CA/MA) and their level of playfulness ($r = 1.63$, $p = 0.38$ and $r = 0.05$, $p = 0.78$, respectively). However, a negative correlation was found between children's CA and teachers' perceived silliness ($r = 0.41$, $p < 0.05$), indicating that teachers of younger children perceived themselves as more "silly." A positive relation was found between two of the APS subscales and children's playfulness. Specifically, teachers' spontaneity and silliness were both positively related to higher levels of child playfulness ($r = 0.38$, $p < 0.05$ and $r = 0.35$, $p < 0.05$, respectively). No significant relation was found with the expressive, fun and creative subscales. A positive, non-significant, correlation was found between the teachers' overall APS score and child playfulness ($r = 0.24$, $p = 0.07$).

Mediation Model

Given the correlation found between children's age and teachers' silliness, as well as between teachers' silliness and children's playfulness, a mediation model was run to examine whether the relationship between children's age (independent variable) and playfulness (dependent variable) was mediated by teachers' silliness (mediator variable). According to Hayes (2009), a significant association between independent and dependent variables is not necessary for testing and establishing mediation: "a failure to test for indirect effects in the absence of a total effect can lead to you miss some potentially interesting, important, or useful mechanisms by which X exerts some kind of effect on Y" (p. 414). dependent variables is not necessary for testing and establishing mediation, and A significant mediation effect was found (indirect effect = 0.017, $SE = 0.12$; 95% confidence interval: LLCI = −0.054, ULCI = −0.003). Thus, the relationships between a child's age and playfulness was mediated by teacher's silliness.

DISCUSSION

This pilot study is one of the first to demonstrate that aspects of teachers' playfulness are positively related to higher levels of children's playfulness. The study showed that teacher spontaneity and silliness were positively related to child playfulness. In addition, the relation between teachers' overall playfulness and children's playfulness was close to significant, and should be reexamined on a bigger sample. Those results are some of the first that empirically support the hypothesis that aspects of teachers' playfulness in ECE, and specifically in teacher–child play interactions, are related to higher level of child playfulness. Note, however that this relation can be bidirectional nature, and that children's classroom behavior can also affect teachers' playfulness.

Our findings confirm Lieberman's (1977) early intuition regarding the importance of teachers' playful behavior in ECE: a teacher who knows how to act playfully by joking, teasing, clowning, and acting silly playful teacher is more likely to facilitate playful behaviors in her students. Specifically, in the current study, teachers' spontaneity and silliness were found to be significant in that regard. Spontaneity has a major role in Lieberman's (1977) definition of playfulness, referring to the individual's ability to be flexible. Glynn and Webster (1992) described spontaneity as the ability to be free-spirited and less disciplined. It is possible that being more spontaneous allow teachers to concentrate more on play and less on discipline in their classrooms, which in turn gives the children more opportunities to be playful. However, it is also possible that teacher's spontaneity is affected by the children's behaviors in class. Further research is needed to fully understand this relationship.

TABLE 1 | Descriptive statistics and inter-correlations (N = 32).

	M	*SD*	1	2	3	4	5	6	7	8	9
(1) Child's CA (months)	54.47	10.00									
(2) Child's MA (months)	63.79	16.03	0.932**								
(3) Teacher's age (years)	40.97	10.43	0.188	0.090							
(4) Teacher's spontaneity	24.90	6.23	−0.152	−0.092	−0.145						
(5) Teacher's expressiveness	26.53	4.15	−0.089	0.007	−0.245	0.680**					
(6) Teacher's fun	26.84	3.94	0.149	0.218	−0.175	0.334	0.534**				
(7) Teacher's creativity	17.15	2.79	0.078	0.110	0.236	0.164	0.381*	0.798**			
(8) Teacher's silliness	14.03	3.80	−0.413*	−0.315	−0.521**	0.481**	0.394*	−0.124	−0.304		
(9) Teacher's total playfulness	109.46	14.84	−0.140	−0.039	−0.277	0.853**	0.880**	0.674**	0.498**	0.478**	
(10) Child's playfulness	−2.82	0.89	0.163	0.143	−0.301	0.379*	0.109	−0.003	−0.157	0.335*	0.243

*CA, chronological age; MA, mental age; *p < 0.05; **p < 0.01.*

The other teacher behavior found in the present study to be significantly related to child playfulness, silliness – defined as childlike behavior (Glynn and Webster, 1992) – and has not been widely investigated. Jung (2011) found that silly facial expressions play a part in teachers' playful behaviors that help young children cope with stressful transitions. Cohen (2008) suggested that caregivers use silliness to solve stressful conflicts. Similarly, Kuhaneck et al. (2010) suggested that a therapist who feels free to act silly can more easily engage children in an activity, and that silliness, like other playful behaviors can be acquired. Silliness may therefore help teachers engage children in play and other activities. Another possible to explanation for the relationship between teacher silliness and child playfulness may be that since silliness is a very salient behavior it may be easy for children to copy and incorporate in their own playful behavior.

In the current study, the mediation model showed that teachers' silliness was higher when children were younger, so that younger children would be more playful given higher levels of teachers' silliness. No direct correction was found between children's age and playfulness level. However, it seems that age did have a role in teacher playfulness, so that it was indirectly related to children's playfulness. This may be explained by the fact that in recent years, play is sidelined by early learning standards and assessment of academic attainments (Roskos and Christie, 2007; Dickey et al., 2016); thus, the older the children are, the less time they have to play in school. However, a Vygotskian approach to play indicates that play and academic development are not mutually exclusive – in fact, scaffolding play can promote not only the development of play itself, but also the acquisition of academic skills (Bodrova, 2008). Preselection can also explain this finding: teachers who perceived themselves as "sillier" chose to work with younger kids with whom they felt more comfortable to expresses this trait. Johnson et al. (2013) argued that "No longer is it enough for an ECE teacher to simply respect playfulness in young children; they must also be playful themselves and master play facilitation techniques" (p. 271). The findings of the current study provide tentative support for this claim.

Knowing the benefits of play and playfulness in ECE (Singer, 2013) and their importance for child development, it is meaningful to address teachers' understanding of play and playfulness, and to promote their own playfulness in order to enable them to help children express and develop their playfulness. According to Broadhead et al. (2010) it is essential that teachers observe children, especially when they play, so that they can gain the ability to understand and support playful behavior and learning.

Although playfulness is usually considered a personal trait, playful behaviors such as silliness may also be seen as skills that can be acquired and honed. Using silliness, spontaneity and other play behaviors for pedagogical purposes is easier when one has mastered play-based assessment and communication techniques (Jones and Reynolds, 2015). Unfortunately, however, play and playfulness are often neglected in teacher education (Johnson et al., 2013).

Carter (1993) suggests three stages in training teachers to be more playful, including aspects highlighted in the current research. First, the teachers need to identify their experience and attitude toward play. Next, they need to be taught to pay attention and understand children's play. When these two skills are attained, they will be able to move to the third stage of training, which includes various playful activities allowing teachers to practice playful behavior. Similarly, Jones and Reynolds (2015) suggest that practice and exercises in remembering one's own past play can help teachers stay in touch with the child inside them, which will help them act more playfully. Trawick-Smith and Dziurgot (2010) showed that teachers who had better education were more likely to perform good-fit play interactions. Teacher training programs should reconsider how to expand teachers' knowledge about and understanding of play and playfulness and how they might develop their own playfulness – information that is unfortunately lacking in current early-years teacher training (Jung and Jin, 2015).

The current study is a pilot study, and despite its interesting findings, its limitations must be taken into consideration. First, its sample is small and not fully representative. For example, all the teachers in the sample were female. Although ECE teachers in Israel are females, it is important to include male teachers in research and to investigate gender differences. Future research should continue this investigation using larger and more representative samples. Second, longitudinal research should be conducted in order to determine causality. Third, teachers'

playfulness was measured using a self-report questionnaire, since, to the best of our knowledge, no observational measurement for adult playfulness exists. However, in order to gain deeper understanding of how teachers use playful behaviors in play interaction in ECE, observational measurement should be considered. Furthermore, all teachers have complete the APS *after* the play-interaction had been videotaped. The play interaction could thus have affected the way they completed the survey. Finally, despite its advantages, the APS, was criticized for poor theoretical background, psychometrics and validity. Therefore, future research should consider using alternative measurements of adult playfulness (such as provided by Barnett, 2007; Shen et al., 2014; Proyer, 2017). In addition, the current research uses new translation of the APS scale that has been developed for this study. The validity of the Hebrew translation of the APS needs further testing.

In addition, further research should investigate additional child and teacher characteristics that can be related to those findings, for example, by looking into the variability among teachers in their play understanding and practices. Promoting teachers' playfulness and playful ECE environment can eventually lead to better classrooms and play experiences for children. As part of a broader research project that investigated play interactions between and children adults at home and educational settings, it will be interesting to look at differentness in playfulness between parent and teachers and its relations to child's playfulness. Finally, in is important to remember that one-on-one teacher–child play interactions are not common in today's ECE settings. Future studies can examine group play interactions with teachers, as well as whether improved teacher–child play interactions also lead to improved child–child play and other interactions.

AUTHOR CONTRIBUTIONS

The author was responsible for research design, data collection, data analysis, and writing the paper.

ACKNOWLEDGMENTS

The author is grateful to Professor Cory Shulman, Jerusalem, for fruitful discussions. The author will also want to thank the teachers and children who generously gave their time during the course of the study.

REFERENCES

Andrich, D. (1988). *Rasch Models for Measurement (No. 68)*. New York, NY: Sage. doi: 10.4135/9781412985598

Aras, S. (2016). Free play in early childhood education: a phenomenological study. *Early Child Dev. Care* 186, 1173–1184. doi: 10.1080/03004430.2015.1083558

Ashiabi, G. S. (2007). Play in the preschool classroom: its socioemotional significance and the teacher's role in play. *Early Child. Educ. J.* 35, 199–207. doi: 10.1007/s10643-007-0165-8

Barnett, L. A. (1991). Characterizing playfulness: correlates with individual attributes and personality traits. *Play Cult.* 4, 371–393.

Barnett, L. A. (2007). The nature of playfulness in young adults. *Pers. Individ. Diff.* 43, 949–958. doi: 10.1016/j.paid.2007.02.018

Bateson, P., Bateson, P. P. G., and Martin, P. (2013). *Play, Playfulness, Creativity and Innovation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139057691

Ben-Sasson, A., Cermak, S. A., Orsmond, G. I., and Tager-Flusberg, H. (2007). Extreme sensory modulation behaviors in toddlers with autism spectrum disorders. *Am. J. Occup. Ther.* 61, 584–592. doi: 10.5014/ajot.61.5.584

Bodrova, E. (2008). Make-believe play versus academic skills: a Vygotskian approach to today's dilemma of early childhood education. *Eur. Early Child. Educ. Res. J.* 16, 357–369. doi: 10.1080/13502930802291777

Bozionelos, N., and Bozionelos, G. (1999). Playfulness: its relationship with instrumental and expressive traits. *Pers. Individ. Diff.* 26, 749–760. doi: 10.1016/S0191-8869(98)00207-4

Broadhead, P., Howard, J., and Wood, E. (Eds.). (2010). *Play and Learning in the Early Years: From Research to Practice*. Newcastle upon Tyne: Sage.

Bundy, A. C. (1997). "Play and playfulness: What to look for," in *Play in Occupational Therapy for Children*, eds L. D. Parham, and L. S. Fazio (St. Louis, MO: Mosby), 52–66.

Bundy, A. C., Luckett, T., Naughton, G. A., Tranter, P. J., Wyver, S. R., Ragen, J., et al. (2008). Playful interaction: occupational therapy for all children on the school playground. *Am. J. Occup. Ther.* 62, 522–527. doi: 10.5014/ajot.62.5.522

Bundy, A. C., Nelson, L., Metzger, M., and Bingaman, K. (2001). Validity and reliability of a test of playfulness. *Occup. Ther. J. Res.* 21, 276–292. doi: 10.1177/153944920102100405

Carter, M. (1993). Catching the spirit: training teachers to be playful. *Child Care Inf. Exch.* 89, 37–39.

Case-Smith, J. (2013). "I foundational knowledge for occupational therapy for children," in *Occupational Therapy for Children-E-Book*, eds J. Case-Smith, and J. C. O'Brien (Maryland Heights, MO: Mosby).

Childress, D. C. (2010). Play behaviors of parents and their young children with disabilities. *Top. Early Child. Spec. Educ.* 31, 112–120. doi: 10.1177/0271121410390526

Christian, K. M. (2012). *The Construct of Playfulness: Relationships with Adaptive Behaviors, Humor, and Early Play Ability*. Ph.D. dissertation, Case Western Reserve University, Cleveland.

Cohen, L. J. (2008). *Playful Parenting: An Exciting New Approach to Raising Children that will Help you Nurture Close Connections, Solve Behavior Problems, and Encourage Confidence*. New York, NY: Ballantine Books.

Cornelli Sanderson, R. (2010). *Towards a New Measure of Playfulness: The Capacity to Fully and Freely Engage in Play* Ph.D. dissertation, Loyola University, Chicago.

Dickey, K., Castle, K., and Pryor, K. (2016). Reclaiming play in schools. *Child. Educ.* 92, 111–117. doi: 10.1080/00094056.2016.1150742

Duncan, P. W., Bode, R. K., Min Lai, S., Perera, S., and Glycine Antagonist in Neuroprotection Americans Investigators (2003). Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Arch. Phys. Med. Rehabil.* 84, 950–963. doi: 10.1016/S0003-9993(03)00035-2

Elliott, F., and Jarneman, M. (2017). *"But It's Not Always So Easy to Join the Play Because One Should Be Here and One Should Be There": Teacher Participation in Children's Pretend Play: A Case Study of One Preschool From a Cultural Historical Activity Theory Perspective*. Jönköping: Jönköping University.

Fabrizi, S. E. (2014). *Measuring the Playfulness of Children with Special Needs in Occupational Therapist Led, Caregiver-Included Community Playgroups*. Fort Lauderdale, FL: Nova Southeastern University.

Fisher, E. P. (1992). The impact of play on development: a meta-analysis. *Play Cult.* 5, 159–181.

Frost, J. L., Wortham, S. C., and Reifel, R. S. (2001). *Play and Child Development*. Upper Saddle River, NJ: Merrill-Prentice Hall.

Glynn, M. A., and Webster, J. (1992). The adult playfulness scale: an initial assessment. *Psychol. Rep.* 71, 83–103. doi: 10.2466/pr0.1992.71.1.83

Glynn, M. A., and Webster, J. (1993). Refining the nomological net of the adult playfulness scale: personality, motivational, and attitudinal correlates for highly intelligent adults. *Psychol. Rep.* 72, 1023–1026. doi: 10.2466/pr0.1993.72.3.1023

Graham, B. C., Sawyers, J. K., and DeBord, K. B. (1989). Teachers' creativity, playfulness, and style of interaction with children. *Creat. Res. J.* 2, 41–50. doi: 10.1080/10400418909534299

Hayes, A. F. (2009). Beyond Baron and Kenny: statistical mediation analysis in the new millennium. *Commun. Monogr.* 76, 408–420. doi: 10.1080/03637750903310360

Johnson, J., Celik, S., and Al-Mansour, M. (2013). "Play in early childhood education," in *Handbook of Research on the Education of Young Children*, 3rd Edn, eds B. Spodek and O. N. Saracho (New York, NY: Routledge), 265–274.

Jones, E., and Reynolds, G. (2015). *The Play's the Thing: Teachers' Roles in Children's Play*, 2nd Edn. New York, NY: Teachers College Press.

Jung, E., and Jin, B. (2015). College coursework on children's play and future early childhood educators' intended practices: the mediating influence of perceptions of play. *Early Child. Educ. J.* 43, 299–306. doi: 10.1007/s10643-014-0658-1

Jung, J. (2011). Caregivers' playfulness and infants' emotional stress during transitional time. *Early Child Dev. Care* 181, 1397–1407. doi: 10.1080/03004430.2010.532873

Kangas, M. (2010). Creative and playful learning: learning through game co-creation and games in a playful learning environment. *Think. Skills Creat.* 5, 1–15. doi: 10.1016/j.tsc.2009.11.001

Kangas, M., and Ruokamo, H. (2012). "Playful learning environments: effects on children's learning," in *Encyclopedia of the Sciences of Learning*, ed. N. M. Seel (New York, NY: Springer), 2653–2655.

Kangas, M., Siklander, P., Randolph, J., and Ruokamo, H. (2017). Teachers' engagement and students' satisfaction with a playful learning environment. *Teach. Teach. Educ.* 63, 274–284. doi: 10.1016/j.tate.2016.12.018

Koch, G. G. (1982). "Intraclass correlation coefficient," in *Encyclopedia of Statistical*, Vol. 4, eds S. Kotz and N. L. Johnson (New York, NY: John Wiley & Sons), 212–217. doi: 10.1002/0471667196.ess1275.pub2

Kuhaneck, H. M., Spitzer, S. L., and Miller, E. (2010). *Activity Analysis, Creativity and Playfulness in Pediatric Occupational Therapy: Making Play Just Right*. Sudbury, MA: Jones & Bartlett.

Lieberman, J. N. (1974). *Explorations in Teacher Characteristics: Playfulness in the Classroom Teacher. ERIC Document Reproduction Service No. ED 050368*. Washington, DC: ERIC, 1–10.

Lieberman, J. N. (1977). *Playfulness: Its Relationship to Imagination and Creativity*. New York, NY: Academic Press.

Lobman, C. L. (2003). What should we create today? Improvisational teaching in play-based classrooms. *Early Years* 23, 131–142. doi: 10.1080/09575140303104

Lobman, C. L. (2006). Improvisation: an analytic tool for examining teacher–child interactions in the early childhood classroom. *Early Child. Res. Q.* 21, 455–470. doi: 10.1016/j.ecresq.2006.09.004

McAfee, O., and Leong, D. (2010). *Assessing and Guiding Young Children's Development and Learning*, 5th Edn. Boston, MA: Pearson.

McMillan, D. (2017). "Towards the playful professional," in *Playful Teaching and Learning*, eds G. Walsh, D. McMillan, and C. McGuinness (New York, NY: Sage), 198–212.

Menshe-Grinberg, A., and Atzaba-Poria, N. (2017). Mother-child and father-child play interaction: the importance of parental playfulness as a moderator of the links between parental behavior and child negativity. *Infant Ment. Health J.* 38, 772–784. doi: 10.1002/imhj.21678

Mullen, E. M. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: AGS, 58–64.

Okimoto, A. M., Bundy, A., and Hanzlik, J. (2000). Playfulness in children with and without disability: measurement and intervention. *Am. J. Occup. Ther.* 54, 73–82. doi: 10.5014/ajot.54.1.73

Pilowsky, T., Yirmiya, N., Shulman, C., and Dover, R. (1998). The autism diagnostic interview-revised and the childhood autism rating scale: differences between diagnostic systems and comparison between genders. *J. Autism Dev. Disord.* 28, 143–151. doi: 10.1023/A:1026092632466

Pinchover, S., Shulman, C., and Bundy, A. (2016). A comparison of playfulness of young children with and without autism spectrum disorder in interactions with their mothers and teachers. *Early Child Dev. Care* 186, 1893–1906. doi: 10.1080/03004430.2015.1136622

Pramling Samuelsson, I., and Johansson, E. (2009). Why do children involve teachers in their play and learning? *Eur. Early Child. Educ. Res. J.* 17, 77–94. doi: 10.1080/13502930802689053

Preacher, K. J., and Hayes, A. F. (2008). "Contemporary approaches to assessing mediation in communication research," in *The Sage Sourcebook of Advanced Data Analysis Methods for Communication Research*, eds A. F. Hayes, M. D. Slater, and L. B. Snyder (Thousand Oaks, CA: Sage), 13–54.

Proyer, R. T. (2011). Being playful and smart? The relations of adult playfulness with psychometric and self-estimated intelligence and academic performance. *Learn. Individ. Diff.* 21, 463–467. doi: 10.1016/j.lindif.2011.02.003

Proyer, R. T. (2012). Examining playfulness in adults: testing its correlates with personality, positive psychological functioning, goal aspirations, and multi-methodically assessed ingenuity. *Psychol. Test Assess. Model.* 54, 103–127.

Proyer, R. T. (2013). The well-being of playful adults: adult playfulness, subjective well-being, physical well-being, and the pursuit of enjoyable activities. *Eur. J. Hum. Res.* 1, 84–98 doi: 10.7592/EJHR2013.1.1.proyer

Proyer, R. T. (2014). To love and play: testing the association of adult playfulness with the relationship personality and relationship satisfaction. *Curr. Psychol.* 33, 501–514. doi: 10.1007/s12144-014-9225-6

Proyer, R. T. (2017). A new structural model for the study of adult playfulness: assessment and exploration of an understudied individual differences variable. *Pers. Individ. Diff.* 108, 113–122 doi: 10.1016/j.paid.2016.12.011

Pyle, A., DeLuca, C., and Danniels, E. (2017). A scoping review of research on play-based pedagogies in kindergarten education. *Rev. Educ.* 5, 311–351. doi: 10.1002/rev3.3097

Roskos, K., and Christie, J. (2007). "Play in the context of the new preschool basics," in *Play and Literacy in Early Childhood: Research from Multiple Perspectives*, 2nd Edn, eds K. Roskos and J. Christie (Mahwah, NJ: Lawrence Erlbaum), 83–100.

Rubin, K. H., Fein, G. G., and Vandenberg, B. (1983). "Play," in *Handbook of Child Psychology: Socialization, Personality, and Social Development*, Vol. 4, eds E. Hetherington and P. H. Mussen (New York, NY: Wiley), 698–774.

Shen, X., Chick, G., and Pitas, N. A. (2017). From playful parents to adaptable children: a structural equation model of the relationships between playfulness and adaptability among young adults and their parents. *Int. J. Play* 6, 244–254. doi: 10.1080/21594937.2017.1382983

Shen, X. S., Chick, G., and Zinn, H. (2014). Playfulness in adulthood as a personality trait: a reconceptualization and a new measurement. *J. Leis. Res.* 46, 58–83.

Singer, E. (2013). Play and playfulness, basic features of early childhood education. *Eur. Early Child. Educ. Res. J.* 21, 172–184. doi: 10.1080/1350293X.2013.789198

Singer, E., Nederend, M., Penninx, L., Tajik, M., and Boom, J. (2014). The teacher's role in supporting young children's level of play engagement. *Early Child Dev. Care*, 184, 1233–1249. doi: 10.1080/03004430.2013.862530

Synodi, E., Linardakis, M., Sadauskienė, R., Kochanskienė, A., and Rimavičienė, D. (2015). Teachers' approach to playfulness in the process of education/learning in Lithuania and Greece. *Signum Temporis* 8, 15–19.

Tegano, D. W., Groves, M. M., and Catron, C. E. (1999). Early childhood teachers' playfulness and ambiguity tolerance: essential elements of encouraging creative potential of children. *J. Early Child. Teach. Educ.* 20, 291–300. doi: 10.1080/0163638990200307

Trawick-Smith, J., and Dziurgot, T. (2010). Untangling teacher-child play interactions: do teacher education and experience influence "good-fit" responses to children's play? *J. Early Child. Teach. Educ.* 31, 106–128. doi: 10.1080/10901021003781148

Vygotsky, L. (1978). Interaction between learning and development. *Read. Dev. Child.* 23, 34–41.

Walsh, G., Sproule, L., McGuinness, C. and Trew, K. (2011). Playful structure: a novel image of early years pedagogy for primary school classrooms. *Early Years* 31, 107–119. doi: 10.1080/09575146.2011.579070

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence-Revised*. San Antonio TX: Psychological Corporation.

Wisneski, D. B., and Reifel, S. (2012). "The place of play in early childhood curriculum," in *Curriculum in Early Childhood Education: Re-Examined, Rediscovered, Renewed*, eds N. File, J. J. Mueller, and D. B. Wisneski (New York, NY: Routledge), 175–187.

Wong, S., and Logan, H. (2016). "Play in early childhood education: an historical perspective," in *Play: A Theory of Learning and Change,* ed. T. Brabazon (New York, NY: Springer), 7–26. doi: 10.1007/978-3-319-25549-1_2

Wood, E., and Attfield, J. (2005). *Play, Learning and the Early Childhood Curriculum*. New York, NY: Sage. doi: 10.4135/9781446280393

Wood, E. (2008). "Developing a pedagogy of play," in *Early Childhood Education: Society and Culture,* eds A. Anning, J. Cullen, and M. Fleer (New York, NY: Sage), 19–30.

Youell, B. (2008). The importance of play and playfulness. *Eur. J. Psychother. Counsel.* 10, 121–129. doi: 10.1080/13642530802076193

Yu, P., Wu, J. J., Chen, I. H., and Lin, Y. T. (2007). Is playfulness a benefit to work? Empirical evidence of professionals in Taiwan. *Int. J. Technol. Manag.* 39, 412–429. doi: 10.1504/IJTM.2007.013503

Yue, X. D., Leung, C. L., and Hiranandani, N. A. (2016). Adult playfulness, humor styles, and subjective happiness. *Psychol. Rep.* 119, 630–640. doi: 10.1177/0033294116662842

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Pinchover. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 14 August 2018
doi: 10.3389/fpsyg.2018.01440

The Positive Relationships of Playfulness With Indicators of Health, Activity, and Physical Fitness

René T. Proyer[1,2]*, *Fabian Gander*[2], *Emma J. Bertenshaw*[3] and *Kay Brauer*[1]

[1] Personality and Assessment, Department of Psychology, Martin-Luther University of Halle-Wittenberg, Halle, Germany, [2] Personality and Assessment, Department of Psychology, University of Zürich, Zurich, Switzerland, [3] Unilever R&D, Colworth Science Park, Bedford, United Kingdom

Adult playfulness is a personality trait that enables people to frame or reframe everyday situations in such a way that they experience them as entertaining, intellectually stimulating, or personally interesting. Earlier research supports the notion that playfulness is associated with the pursuit of an active way of life. While playful children are typically described as being active, only limited knowledge exists on whether playfulness in adults is also associated with physical activity. Additionally, existing literature has not considered different facets of playfulness, but only global playfulness. Therefore, we employed a multifaceted model that allows distinguishing among *Other-directed*, *Lighthearted*, *Intellectual*, and *Whimsical* playfulness. For narrowing this gap in the literature, we conducted two studies addressing the associations of playfulness with health, activity, and fitness. The main aim of Study 1 was a comparison of self-ratings ($N = 529$) and ratings from knowledgeable others ($N = 141$). We tested the association of self- and peer-reported playfulness with self- and peer-reported physical activity, fitness, and health behaviors. There was a good convergence of playfulness among self- and peer-ratings (between $r = 0.46$ and 0.55, all $p < 0.001$). Data show that both self- and peer-ratings are differentially associated with physical activity, fitness, and health behaviors. For example, self-rated playfulness shared 3% of the variance with self-rated physical fitness and 14% with the pursuit of an active way of life. Study 2 provides data on the association between self-rated playfulness and objective measures of physical fitness (i.e., hand and forearm strength, lower body muscular strength and endurance, cardio-respiratory fitness, back and leg flexibility, and hand and finger dexterity) using a sample of $N = 67$ adults. Self-rated playfulness was associated with lower baseline and activity (climbing stairs) heart rate and faster recovery heart rate (correlation coefficients were between −0.19 and −0.24 for global playfulness). Overall, Study 2 supported the findings of Study 1 by showing positive associations of playfulness with objective indicators of physical fitness (primarily cardio-respiratory fitness). The findings represent a starting point for future studies on the relationships between playfulness, and health, activity, and physical fitness.

Keywords: adult playfulness, playfulness, health, activity, physical fitness, health behavior, OLIW, peer-ratings

OPEN ACCESS

Edited by:
Mark Hallahan,
College of the Holy Cross,
United States

Reviewed by:
Konrad Schnabel,
International Psychoanalytic University
Berlin, Germany
Doug Maynard,
SUNY New Paltz, United States

***Correspondence:**
René T. Proyer
rene.proyer@psych.uni-halle.de

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *28 October 2017*
Accepted: *23 July 2018*
Published: *14 August 2018*

Citation:
Proyer RT, Gander F, Bertenshaw EJ and Brauer K (2018) The Positive Relationships of Playfulness With Indicators of Health, Activity, and Physical Fitness. Front. Psychol. 9:1440. doi: 10.3389/fpsyg.2018.01440

INTRODUCTION

There is an agreement in the literature that play behaviors serve important functions in numerous developmental processes in infancy and childhood (see e.g., Bruner et al., 1976; Burghardt, 2005). For example, for animals it has been argued that play at a young age may help, for example, in muscle development and facilitate physical balance (e.g., Fagen, 1981). Playfulness as a *personality trait* in adults has hitherto been widely neglected in research and practice across disciplines. In particular, its correlates with variables of physical functioning in adult life are widely unknown. Nevertheless, a growing number of studies exist that support the notion that playfulness (the personality trait associated with play as the actual behavior) may serve an important role in several life domains of adults too. Amongst others, it has been shown that playfulness relates to positive outcome variables such as coping (e.g., Staempfli, 2007; Magnuson and Barnett, 2013), work performance and innovative behavior at work (Glynn and Webster, 1992; Yu et al., 2007), creativity and intrinsic motivation (Amabile et al., 1994; Proyer, 2012b), virtuousness (Proyer and Ruch, 2011), sexual selection (Chick et al., 2012; Proyer and Wagner, 2015), academic success (Proyer, 2011), low expressions in the Impostor phenomenon (Brauer and Proyer, 2017), or subjective well-being (Proyer, 2013, 2014a,b; Proyer et al., 2018a). The present study aims at extending these findings to health, activity, and physical fitness.

While there is no agreement in the literature about a definition of playfulness in adults as a personality trait, the recent years have seen an increase in the study of the variable. It has been argued that research has partially suffered from the usage of conceptualizations and assessment instruments that have failed to clearly differentiate between the core of playfulness and its consequences (e.g., when using statements such as "I laugh a lot" for the study of individual differences in *playfulness*; Proyer, 2012a; Proyer and Jehle, 2013), a lack of distinctiveness from related traits (e.g., humor, creativity, or curiosity; e.g., Proyer, 2018), and unwanted overlap with basic personality traits (mostly extraversion and emotional stability; Proyer and Jehle, 2013). Based on a mixed-methodology (e.g., psycho-linguistic, psychometric, and qualitative approaches; for an overview see Proyer, 2017) a new definition that aims at focusing on the core characteristics of playfulness has been proposed; namely,

> "Playfulness is an individual differences variable that allows people to frame or reframe everyday situations in a way such that they experience them as entertaining, and/or intellectually stimulating, and/or personally interesting. Those on the high end of this dimension seek and establish situations in which they can interact playfully with others (e.g., playful teasing, shared play activities) and they are capable of using their playfulness even under difficult situations to resolve tension (e.g., in social interactions, or in work type settings). Playfulness is also associated with a preference for complexity rather than simplicity and a preference for—and liking of—unusual activities, objects and topics, or individuals" (Proyer, 2017 p. 114).

Hence, it is argued that playfulness in adults also contributes to other life domains (e.g., intellectual achievements, or personal involvement)—rather than entertainment alone, which has been highlighted as its main function in earlier definitions (e.g., Murray, 1938; Glynn and Webster, 1992; Barnett, 2007). A playful attitude can also help to gain a new perspective on serious topics and assist in coping with adverse circumstances (cf. the notion of serious-cheerfulness as one pursuit of the *homo ludens* in Rahner, 1948/2008; see also Proyer and Rodden, 2013; Proyer, 2014a). On a more descriptive level, playful people are typically seen as funny, humorous, spontaneous, unpredictable, active, energetic, adventurous, convivial, and cheerful and tend to display playful behavior by telling jokes, playing pranks, and horsing around (Barnett, 2007, 2011). In line with these descriptions, playfulness at a younger age has also been linked to greater physical activity or physical spontaneity (e.g., coordinated movements; Lieberman, 1977; see also Barnett, 1991; Singer et al., 1980).

Proyer (2017) proposes a new structural model of adult playfulness that differentiates among four facets; namely, (a) *O*ther-directed, (b) *L*ighthearted, (c) *I*ntellectual, and (d) *W*himsical playfulness (OLIW-model; see **Table 1** for a description). Additionally, previous research has shown that playfulness could also be conceptualized and measured on a global level (playfulness in general), in terms of an easy onset and high intensity of playful experiences along with the frequent display of playful activities (Proyer, 2012a). We will use both a global assessment of playfulness and a measure for the four facets for a more fine-grained differentiation of the variable.

A particularly understudied topic in the research on adult playfulness is the study of physical activity and fitness. It has been

TABLE 1 | Description of the different playfulness facets.

Trait	Description
Global Playfulness	Global Playfulness is an individual differences variable that allows people to frame or reframe everyday situations in a way such that they experience them as entertaining, and/or intellectually stimulating, and/or personally interesting.
Other-directed	The facet of other-directed playfulness is characterized by the use of playful behaviors in social situations. High scorers use playfulness to ease tense situations, and cheer other people up, they enjoy horsing around with friends and engage, generally, in a playful interaction style with other people
Lighthearted	The facet of lighthearted playfulness is characterized by a spontaneous, carefree view of life. High scorers do not think much about possible consequences of their behavior but prefer and enjoy improvising in comparison with elaborate preparation.
Intellectual-Creative	The facet of intellectual playfulness is characterized by the enjoyment of playing with ideas. High scorers like to puzzle over problems and to come up with new, creative solutions for problems.
Whimsical	The facet of whimsical playfulness is characterized by a preference for breaking ranks. High scorers are amused by oddities and have a preference for extraordinary things and people. Others often regard them as extravagant.

suggested that play in children might serve as a practice for future skills (although this idea has also been disputed; see Chick, 2001; Burghardt, 2005). A first hint on a positive association between playfulness and physical health, greater activity, and fitness comes from Lieberman's (1977) early work on playfulness in children. For adults, she argues that "[. . .] through its component parts of sense of humor, manifest joy, and spontaneity, it has major implications for childrearing practices, educational planning, career choices, and leisure pursuits" (Lieberman, 1977 p. xi) and later she notes "[. . .] playfulness as a quality of play would developmentally transform itself into a personality trait of the player in adolescence and adulthood" (Lieberman, 1977, p. 23).

Associations Between Playfulness, Health, Activity, and Fitness

There are few published studies exploring the relationship between playfulness and indicators of health, activity, and fitness. Therefore, the present studies aim to extend and build upon limited knowledge of the relationships. Existing literature and theoretical reasoning allow us to derive an initial framework that supports the notion of an association. There are several ways to think about the potential link between adult playfulness and markers of physical functioning: Firstly, the association may be developed *indirectly via positive affect*: Playful personalities give rise to feeling playful more frequently which is associated itself with experiencing a pleasant, positive feeling. Positive affect is associated with successful health outcomes and behaviors in some studies, although the evidence is mixed in general (e.g., Lyubomirsky et al., 2005; Cameron et al., 2017). Accordingly, Fredrickson (1998, 2001) argues that positive emotions may contribute to physical resources such as coordination, strength, or cardiovascular health by strengthening personal resources (e.g., via joint social activities that elicit positive emotions and that may consequently also result in greater levels of activities)[1]. The experienced positive affect may also buffer against (emotional) distress that may be a hindering factor for engaging in physical activity or limiting a persons' desire to actively engage with their environment. Secondly, a positive association between playfulness and physical functioning can also manifest as *action orientated tendencies* toward curiosity, exploration, or physical activities; playful people may be interested in trying and doing a greater range of activities. This notion receives support from studies showing that those high in playfulness also seem to have an active interest in the pursuit of leisure time activities (e.g., Mannell, 1984), experience low boredom in their leisure time (e.g., Barnett, 2011), have good skills to cope with adversities and possess a mastery orientation (e.g., Staempfli, 2007; Magnuson and Barnett, 2013; Proyer, 2014a) and have an interest in the pursuit of enjoyable activities (e.g., doing fun things with others or communing with nature; Proyer, 2013).

While a full analysis of motivational factors associated with playfulness is missing, there are data linking greater expressions of playfulness with intrinsic motivation (Amabile et al., 1994; Proyer, 2012b). Given that many models which describe the structure of adult playfulness, contain other-directed facets (e.g., Lieberman, 1977; Barnett, 2007; Proyer, 2017) one might assume that affiliation may also be related to playfulness. This may help playful adults not only to engage in team sports, but also in other joint activities with others (e.g., outdoor activities) and, in general, be associated with greater engagement with one's environment. Playfulness may facilitate achievement in active sports as it has been linked with intrinsic motivation and motivation toward achievement (see e.g., Proyer, 2011, 2012b), innovation and creativity (e.g., Yu et al., 2007; Bateson and Martin, 2013; for an overview see Proyer et al., 2018b), and competitiveness (see Csikszentmihalyi, 1975) to name but a few.

Thirdly, those high in playfulness may be in general more interested in *promoting their own health*. For example, Proyer (2013) found positive associations between playfulness and health-behaviors such as pursuing a more active way of life. Such activities may then contribute positively to an individual's health. However, playfulness also showed some negative relationships with specific health behaviors, such as security orientation (e.g., wearing a safety belt in the car or avoiding violence), substance use (e.g., drinking coffee or alcohol), and hygiene (e.g., regularly brushing teeth, or using floss). This may be associated with a certain lighthearted attitude (Proyer, 2017) in dealing with the daily life and speaks for the need to differentiate between facets of playfulness and different indicators of physical functioning. Fourthly, playfulness may either directly or indirectly have a sustained effect on the development of health via skills needed for physical functioning, based on mechanisms not identified above. It is possible that genetic predispositions interact with trait playfulness to increase physical activity. Another example could be the choice of vocational activity that may be influenced by the fit between a person's trait playfulness and his/her preferred activities. Proyer (2013) showed that global playfulness goes along with better (physical) coordination skills (Proyer, 2013; $N = 255$, self-reported data in a correlational study). In this sense, playfulness may facilitate greater physical activity as those higher in playfulness have greater skills in certain areas (e.g., coordination) and use them accordingly (e.g., in sports, or other activities that require greater skill-levels), or in the sense that playfulness supports the development and acquisition of certain health-related skills.

Taken together these findings speak for greater levels of activity among highly playful people. Thus, playfulness might affect health and physical skills through increased activity. On the other hand, highly playful people may be more likely to be physically active because they aim for greater physical and or mental health. The relation between playfulness, health, activity, and fitness is likely to be complex, so in these studies we focus on testing particular associations between personality and physical

[1]Fredrickson (2003) notes: "For example, joy and playfulness build a variety of resources. Consider children at play in the schoolyard or adults enjoying a game of basketball in the gym. Although their immediate motivations may be simply hedonistic—to enjoy the moment—they are at the same time building physical, intellectual, psychological and social resources. The physical activity leads to long-term improvements in health, the game-playing strategies develop problem-solving skills, and the camaraderie strengthens social bonds that may provide crucial support at some time in the future [. . .]. Similar links between playfulness and later gains in physical, social and intellectual resources are also evident in nonhuman animals, such as monkeys, rats and squirrels" (p. 333).

activity. Since playfulness is assumed to be a multi-dimensional construct, the playful facets [e.g., playfulness in its *Other-directed* (e.g., team sport, group activities) or *Lighthearted* (e.g., being open for new activities) facet] may be more strongly correlated with different indicators of health, activity, and fitness than a single global score that averages playfulness across different domains. Also, the findings so far relied on self-reports for both playfulness and indicators of health, activity, and fitness and it is unclear to what extent these results are due to a shared method influence. The problem in this case is that "[. . .] one has no way of distinguishing trait variance from unwanted method variance" (Campbell and Fiske, 1959, p. 102). One possible solution is the consideration of additional methods and testing whether the findings are comparable.

We aim to expand the validity of the findings by additionally collecting peer-ratings of a well-acquainted informant for each participant, on both the playfulness and indicators of health variables. This approach has two main merits: Firstly, there is a broad range of research that has shown that although self-ratings provide a valid source of information, self-perceptions are prone to biases (cf. Connolly and Ones, 2010). For example, Vazire and Mehl (2008) have shown that peer-ratings are not only as accurate as self-ratings, but that they *independently* predict behavioral outcomes and provide a unique insight. Hence, we aim to implement to collect peer-ratings for each tested participant. Moreover, we follow Hofstee's (1994) approach to aggregate the self- and peer-rating to provide an approximation of the participants "true" personality. We will use this for studying associations of playfulness with health, activity, and fitness. Secondly, this approach allows reducing common method variance (Campbell and Fiske, 1959) based on the same assessment techniques (i.e., questionnaires).

The present set of studies aimed at replicating and extending previous findings (e.g., Proyer, 2013) in several ways. Study 1 uses an online survey methodology and extends previous findings by considering a broad array of indicators of health, activity, fitness; studying different facets of playfulness, in addition to global playfulness. Further, we collected informant ratings of well-acquainted peers for playfulness and health indicators. This approach allows us to test whether the correlational patterns are comparable for the same variable in self- and peer-ratings with other variables. Study 2 uses objective indicators of health, activity, and fitness, as well as playfulness (as measured in a laboratory setting) for examining their associations in a methodologically more rigorous approach, in order to overcome the shortcomings of having to rely on self-ratings only.

STUDY 1

Study 1 examines the relationships of self-and peer-rated global playfulness and its facets with mental and physical health (self-ratings) as well as physical activity, physical fitness, and health-behaviors (self- and peer-ratings). Based on earlier findings (Proyer, 2013), we expected positive relationships of playfulness with self-rated mental health and selected health behavior such as leading an active way of life, but also substance use. In line with Proyer (2013), we expected negative relationships with compliance (e.g., taking care and complying to regulations in road traffic), but positive associations with levels of activity and physical health. On the level of specific facets, we expected the largest relationships for *Other-directed* playfulness, since this aspect has the strongest conceptual overlap with being active and engaging in (play) activities with others in comparison to other, more cognitive aspects of playfulness. We assume that there are indirect effects of having more contact with others and either inviting them to join an own activity or participating in theirs. This type of playful exchange with others should be particularly helpful in engaging in activities. Of course, not all of these will be positively associated with health and physical fitness but leading a more active way of life should also include physical activities. The other facets are expected to show comparatively smaller or no relationships with health, fitness, and well-being.

Finally, we expected these relationships to be present in all data sources (i.e., when looking at relationships between self-rated playfulness and self-rated health/activity/fitness, but also self- and peer-ratings, peer- and self-ratings, and peer- and peer-ratings). Previous studies found that self- and peer-rated playfulness converges in a range of 0.44–0.57 across the four OLIW-facets in a mixed sample of highly acquainted people (good friends, romantic partners, siblings, etc., Proyer, 2017), between 0.33 and 0.58 in heterosexual couples in a romantic relationship (Proyer et al., 2018a), and between 0.21 and 0.37 in a zero-acquaintance setting (ratings of short self-descriptions of up to five sentences; Proyer and Brauer, 2018). Thus, the convergence was within the range which has been reported for other personality traits. For example, Funder et al. (1995) reported for facets of the big five an average correlation of $r = 0.37$ between self- and parents' ratings, and relationships of $r = 0.36$ and $r = 0.30$ when testing college and hometown acquaintances, respectively. We expect overlap in the range of these coefficients in the present study. We also expect robust convergence among self-and peer-ratings for our measures of activity, fitness, and health behaviors. However, it must be noted that some of these may be more difficult to observe than others. It has been argued that personality traits differ in their observability/evaluativeness (see Vazire, 2010) and similarly, depending on the level of acquaintance, some health behaviors or pursued activities may be low in observability. Moreover, we will test the overlap between the observed relationships of self- and peer-rated playfulness with activity, fitness, and health behaviors by computing the vector correlations (see Borkenau and Liebler, 1992) for each playfulness facet. This approach allows to estimate the overall overlap of the observed correlations when comparing the self- vs. peer ratings.

Methods

Sample

A total of $N = 529$ participants (81.1% women) aged 18–78 years ($M = 37.41$, $SD = 15.90$) took part in an online survey. A breakdown into age-categories shows that a large portion of participants were between 18 and 29 years old (43.5%), 10.8% were between 30 and 39, 20.6% were between 40 and 49, 15.3% were between 50 and 59, and 9.8% were 60 years or older. A large part of the sample (41.6%) had a degree from a university or a university of applied sciences (BSc or higher), 36.1% held a degree allowing them to attend a university or a university

of applied sciences, 16.3% completed vocational training, 5.8% have completed mandatory school, and only one participant (0.2%) had not completed mandatory school. Most participants (50.3%) were Swiss, or German (43.2%). More than half of the participants were currently in a relationship (60.3%), 33.3% were single, 5.1% were divorced/separated, and 1.3% were widowed. Most participants were currently employed (67.5%). Overall, the sample is diverse regarding age and other demographic characteristics.

About a fourth of the sample (141 participants; 26.7%) also provided peer-reports of the playfulness measures. The peer-raters were mostly female (53.9%) and aged 17 to 83 ($M = 39.78$, $SD = 14.94$). The largest portion of peer-raters was the romantic partner of the person who provided the self-report (48.9%), a family member (27.7%), or a (close) friend (23.4%).

Instruments

Playfulness measures

The *Short Measure of Adult Playfulness* (SMAP; Proyer, 2012a) assesses global playfulness with five items on a 7-point scale (1 = "does not apply at all" to 7 = "applies completely"). A sample item is "I am a playful person." Internal consistency was high ($\alpha = 0.90$).

The *OLIW-playfulness questionnaire* (OLIW; Proyer, 2017) assesses the four facets of playfulness (*Other-directed, Lighthearted, Intellectual,* and *Whimsical* playfulness) with seven items each on a 7-point scale (1 = "does not apply at all" to 7 = "applies completely"). A sample item is "I can use my playfulness to bring joy to other people or cheer them up" (*Other-directed* playfulness). Internal consistencies ranged from $\alpha = 0.66$ (*Intellectual*) to $\alpha = 0.77$ (*Whimsical*).

Health, activity, and fitness measures

Single items for the assessment of *Subjective Physical and Mental Health*; assessed with one item each, ranging from 0 (= "very bad") to 10 (= "very good"). Single-item ratings of the subjective health status have been reported to be stable (Miilunpalo et al., 1997) and substantially related to external criteria such as number of physician visits (Miilunpalo et al., 1997), or mortality (for an overview see Idler and Benyamini, 1997).

The *General Level of Activity* (GLA; Proyer et al., 2014) was assessed with four single items for the comparison of (a) the own activity level in general ("I would consider myself a not very active person" vs. "a very active person"); (b) the own activity level in comparison with the person's peers ("Compared to other persons of my age and gender I would consider myself a not very active person" vs. "a very active person"); (c) comparison with people that are generally very active ("Some people are very active. They try to be active whenever possible and are looking for ways to complete tasks in a way that involves movement and physical activity. To what extent does this describe you?"); and (d) comparison with people that are generally not very active ("Some people are not very active. Although they are not lazy, they are never as active as they could be. To what extent does this describe you?" [recoded item]). All items are answered on a 7-point scale (1 = "not at all" to 7 = "to a very large extent"). Internal consistency was high ($\alpha = 0.89$).

The *International Physical Activity Questionnaire* long-form (IPAQ; Craig et al., 2003) asks about the time spent (i.e., minutes per week) with physical activity in the past 7 days in four domains (work, active transportation, domestic and garden, and leisure-time). The questionnaire distinguishes among activities of low, moderate, and vigorous intensity. The IPAQ allows calculating specific scores for the four domains, and a total score. Additionally, scores for the time spent sitting, and the time spent in transportation ("passive transportation") can be computed. Craig et al. (2003) report acceptable measurement properties, while moderate validity of the German long form in comparison with objective data was reported (Wanner et al., 2016).

The *Physical Fitness Questionnaire* (FFB-MOT; Bös et al., 2002) asks for the ease of performing twelve physical exercises. The FFB-MOT assesses general physical fitness in four basic motor abilities and a total score ("global fitness"): Cardiorespiratory fitness (e.g., "running one kilometer without a break"), strength (e.g., "carrying a heavy basket [8kg] over several floors"), flexibility (e.g., "tying shoelaces while standing upright"), and coordination (e.g., "doing a somersault"). Items are rated on a five-point scale (1 = "I cannot perform this exercise" to 5 = "I can perform this exercise without any problems"). Internal consistency was $\alpha = 0.82$ for the total score and ranged from $\alpha = 0.66$ (coordination) to $\alpha = 0.85$ (cardiorespiratory fitness) for the basic abilities.

The Multiple Health Behavior Questionnaire (MHB-39; Wiesmann et al., 2003) assesses the frequency of performing 39 health-related behaviors on a five-point scale (1 = "*never*" to 5 = "*always*"). The internal consistency of the total scale was high ($\alpha = 0.83$). Additionally, Wiesmann et al. (2003) and Proyer et al. (2013) extracted six orthogonal factors in a principal component analysis. These factors were labeled: Active way of life (e.g., being frequently physically active), compliance (e.g., visiting a physician when becoming aware of physical symptoms), substance use (e.g., drinking coffee or alcohol), security orientation (e.g., wearing a seatbelt in the car), diet (e.g., eating sweet dishes), and hygiene (e.g., using dental floss). For comparison purposes across studies and self- and peer-ratings, we used the regression weights from Proyer et al. (2013) to replicate their factor solution in this study.

Procedure

Participants were recruited through online-advertisements (forums, mailing lists), and by contacting participants from earlier (unrelated) studies. All participants gave informed consent and completed all instruments online. After the end of the survey, participants were asked to forward the link of the online study to a person who knows them well (no further restrictions/inclusion criteria were implemented). These people were asked to complete peer-ratings of the SMAP, the OLIW, and the GLA, the FFB-MOT, and the MHB-39.

Participants were not financially compensated but received an automated feedback on their individual scores upon completion of the study and had the opportunity to enter a prize draft for one of 10 online shopping vouchers worth 25 Swiss Francs. Peer-raters did not receive any incentive for participation.

Results

In a first step, we examined the convergence of self- with peer-rated measures (see **Supplementary Table A**). Overall, the self-peer convergence was high for playfulness and its facets (all *r*s >0.46), while all other relationships outside the main axis were smaller in size (all *r*s ≤ 0.34). We also examined the agreement among self- and peer-ratings for activity, fitness, and health behaviors. Again, the relationships were high and ranged between $r = 0.47$ and $r = 0.74$ (mean = 0.68), while the correlations between the same constructs (i.e., the correlations in the main axis) were numerically higher than those between different constructs (see **Supplementary Table B** for all coefficients). Thus, self- and peer-ratings converged very well and were in the expected range for all measures.

Further correlational analyses (not reported in full detail) revealed small relationships with gender (men showed higher scores in global [$r = -0.10$] and *Lighthearted* playfulness [$r = -0.11$]) and small to medium-sized relationships with age (global [$r = -0.10$] and *Other-directed* playfulness [$r = -0.28$] showed negative relationships with age, while *Lighthearted* [$r = 0.16$], *Intellectual* [$r = 0.20$], and *Whimsical* [$r = 0.14$] playfulness were positively related to age). Therefore, we controlled for the influence of gender and age in all main analyses.

In a next step, we examined the relationships of playfulness with health, activity, and fitness. First, we present the associations of self- and peer-rated playfulness with *self-rated* indicators of health, activity, and fitness (**Table 2**), and afterwards the associations of self- and peer-rated playfulness with *peer-rated* indicators (**Table 3**), and finally the relationships between averaged self- and peer-ratings of playfulness and the indicators of health, activity, and fitness (**Supplementary Table C**).

Analysis of the Self-Ratings of Playfulness

Global playfulness was widely unrelated to health, activity, and fitness but showed associations with health behaviors such as leading an active way of life or substance consumption (both positive), and compliance (negative) in self-ratings. *Other-directed* playfulness was positively associated with mental health, activity and all fitness measures (with the exception of strength) and showed positive relationships to health behaviors overall (especially, active way of life and substance consumption). *Lighthearted* playfulness positively related to mental and physical health and activity but was unrelated to physical fitness. It showed positive relationships to leading an active way of life and substance consumption, and negative relationships to compliance. *Intellectual* playfulness was positively associated with mental health, the global activity level, and all fitness measures (with the exception of strength), and was positively related to health behaviors overall, safety, leading an active way of life, and diet. Finally, *Whimsical* playfulness was unrelated to health, and fitness (with the exception of flexibility), but went along with higher levels of global activity. Further, it was related to health behaviors overall, safety, leading an active way of life, and diet (positive), and compliance (negative).

Analysis of the Peer-Ratings of Playfulness

When analyzing the peer-ratings (see **Table 2**), we found generally small effect sizes (mean $r = 0.02$; range = [−0.20;0.18]) and in most cases correlation coefficients failed to reach statistical significance. For a better understanding of the relationships between self- and peer-rated playfulness and the tested measures, we computed vector correlations[2] to estimate their overlap. While the self- and peer-correlations showed only a comparatively small overlap for global ($r = 0.38$) and *Intellectual* ($r = 0.43$) playfulness, there was a more substantial overlap for *Other-directed* and *Lighthearted* ($r = 0.64$) as well as *Whimsical* playfulness ($r = 0.56$). Hence, the inspection of single coefficients showed that differences (e.g., different signs for associations) in the correlational patterns existed particularly for global and *Intellectual* playfulness.

We analyzed the relationships between *peer-ratings* of health behaviors, activity, and fitness, and self- and peer-rated playfulness (see **Table 3**) and, again, computed the agreement of the relationships via vector correlations. When using the peer-rated perspective toward health behaviors their relationships with self- and peer-ratings of playfulness showed greater convergence, as it was substantial for global ($r = 0.87$), *Other-directed* ($r = 0.76$), *Lighthearted* ($r = 0.85$), and *Whimsical* ($r = 0.73$) while the lowest overlap existed for *Intellectual* playfulness ($r = 0.51$). Hence, the relationships between self- and peer-rated playfulness seem to widely converge in the sense that they share overlap in their relationships toward physical activities and health. However, the findings do not indicate a perfect overlap and should be interpreted cautiously. In any case they support the notion that the reported associations cannot be explained by a common method bias in the self-ratings of playfulness and the other measures.

Peer-rated global playfulness positively related to some health behaviors (mostly leading an active way of life) and showed some negative relationships to fitness. Most playfulness facets (*Other-directed*, *Intellectual*, and *Whimsical*) were positively related to activity, and went along with leading an active way of life.

Finally, we analyzed the relationship between self-rated health behavior and the aggregated self- and peer-ratings for playfulness to provide a more accurate estimate of the traits. As expected, the use of aggregated self- and peer-ratings contributed to the explanation of the relationships between playfulness and indicators of health (e.g., R^2 increase ≤ 18%; see **Supplementary Table C**). The analyses of aggregated ratings widely confirmed the findings. Again, *Other-directed*, *Intellectual*, and *Whimsical* playfulness showed positive relationships to activity and leading an active way of life. However, while the consideration of peer-ratings might reduce the influence of biases due to the shared method, peer-ratings did not provide substantial additional information: They did not explain additional variance in health, activity, or fitness variables above the influence of self-ratings.

[2]Vector correlations were computed on basis of Fisher's *r*-to-*z* correlations (see Borkenau and Liebler, 1992).

TABLE 2 | Relationships of self- and peer-rated playfulness with different self-rated indicators of health, activity, and fitness, controlled for gender and age.

Self-ratings	Self						Peer					
	SMAP	OTD	LTH	INT	WHI	R^2 (OLIW)	SMAP	OTD	LTH	INT	WHI	R^2 (OLIW)
Mental health	0.04	0.14**	0.26***	0.14**	−0.02	0.09***	0.10	0.14	0.13	0.13	0.01	0.03
Physical health	0.02	0.04	0.17***	0.06	0.01	0.03**	0.07	0.08	0.06	0.10	−0.04	0.02
Global activity level (GLA)	0.06	0.20***	0.12**	0.18***	0.15**	0.05***	0.03	0.13	0.06	0.16	0.12	0.03
Physical activity (IPAQ)	0.02	0.10*	0.09*	0.07	0.06	0.01	0.07	−0.01	0.07	0.06	0.03	0.01
Physical fitness (FFB-MOT)	0.00	0.12**	0.08	0.15***	0.01	0.03**	−0.12	0.06	−0.02	0.07	0.00	0.01
Strength	0.00	0.02	0.05	0.07	0.01	0.01	−0.08	0.02	−0.02	0.05	−0.03	0.01
CR-Fitness	0.01	0.10*	0.07	0.12**	0.00	0.02*	−0.17*	0.02	−0.09	0.04	−0.03	0.01
Flexibility	0.01	0.11**	0.07	0.15***	0.09*	0.02*	0.01	0.04	0.08	0.04	0.18*	0.03
Coordination	0.00	0.09*	0.03	0.10*	−0.05	0.02*	−0.09	0.07	−0.02	0.08	−0.13	0.03
Health behaviors (MHB-39)	−0.01	0.12**	0.03	0.17***	0.10*	0.03**	−0.01	−0.02	−0.16	−0.02	−0.03	0.03
Safety	0.05	0.00	0.02	0.08*	0.11**	0.02*	0.00	0.00	−0.08	−0.02	−0.05	0.00
AWOL	0.12**	0.33***	0.25***	0.24***	0.24***	0.14***	0.10	0.16*	0.04	0.09	0.17*	0.05
Compliance	−0.09*	−0.03	−0.12**	−0.03	−0.13**	0.02*	−0.10	−0.13	−0.13	−0.20*	−0.13	0.05
Diet	−0.01	0.06	0.04	0.11**	0.05	0.01	−0.02	0.04	0.06	0.06	0.06	0.01
Substance consumption	0.09*	0.09*	0.11*	0.05	0.04	0.02	−0.10	0.05	0.07	−0.13	0.08	0.06
Hygiene	−0.07	−0.02	−0.07	0.05	−0.06	0.02	−0.01	−0.09	−0.15	−0.01	0.03	0.03

*N = 529 for self-ratings, N = 128-141 for peer-ratings. SMAP, short measure of adult playfulness; OTD, Other-directed; LTH, Lighthearted; INT, Intellectual; WHI, Whimsical playfulness. CR-Fitness, Cardio-Respiratory Fitness. AWOL = Leading an active way of life. R^2 (OLIW) = Explained variance by all playfulness facets combined, over the influence of gender and age. *p < 0.05. **p < 0.01. ***p < 0.001. Two-tailed.*

TABLE 3 | Partial correlations of self- and peer-rated playfulness with different peer-rated indicators of health, activity, and fitness, controlled for gender and age.

Peer-Ratings	Self						PEER					
	SMAP	OTD	LTH	INT	WHI	R^2 (OLIW)	SMAP	OTD	LTH	INT	WHI	R^2 (OLIW)
Global activity level (GLA)	−0.04	0.16	0.11	0.23**	0.10	0.05	0.12	0.26**	0.03	0.18*	0.23**	0.11**
Physical fitness (FFB-MOT)	−0.17*	0.05	0.05	0.06	−0.06	0.01	−0.03	0.11	−0.02	0.15	0.07	0.03
Strength	−0.21*	−0.05	−0.06	0.02	−0.11	0.02	−0.13	0.05	−0.05	0.14	−0.02	0.03
CR-fitness	−0.13	0.07	0.11	0.06	−0.04	0.02	−0.05	0.11	−0.02	0.15	0.06	0.03
Flexibility	−0.09	0.11	0.04	0.11	0.04	0.01	0.03	0.13	0.08	0.15	0.17*	0.04
Coordination	−0.15	−0.02	0.01	0.04	−0.12	0.02	0.00	0.02	−0.11	0.06	−0.02	0.02
Health behaviors (MHB-39)	0.05	0.11	−0.02	0.08	0.07	0.02	0.19*	0.28**	−0.11	0.31***	0.05	0.15***
Safety	0.11	−0.04	−0.01	0.08	0.05	0.02	0.18*	0.02	−0.03	0.22*	−0.02	0.06*
AWOL	0.15	0.37***	0.22**	0.23**	0.19*	0.14***	0.27**	0.44***	0.13	0.32***	0.26**	0.23***
Compliance	−0.06	−0.13	−0.14	−0.15*	−0.13	0.04	0.00	0.15	−0.13	0.08	−0.15	0.09*
Diet	−0.13	0.12	0.08	0.09	0.11	0.02	−0.12	0.03	0.02	−0.01	0.00	0.00
Substance consumption	−0.02	0.03	0.15	0.04	0.02	0.02	−0.06	0.13	0.15	−0.03	0.05	0.05
Hygiene	−0.10	−0.05	−0.09	−0.01	−0.06	0.01	−0.12	−0.16	−0.14	−0.11	0.06	0.04

*N = 128-141 for peer-ratings. SMAP, short measure of adult playfulness; OTD, Other-directed; LTH, Lighthearted; INT, Intellectual; WHI, Whimsical playfulness. CR-Fitness, Cardio-Respiratory Fitness. AWOL = Leading an active way of life. R^2 (OLIW) = Explained variance by all playfulness facets combined, over the influence of gender and age. *p <0.05. **p < 0.01. ***p < 0.001.Two-tailed.*

Discussion

This study provides support for the notion of a contribution of playfulness to physical functioning. The findings were widely in line with expectations and show differential effects for the single facets of playfulness. As expected, self- and peer-ratings for playfulness and the indicators for health, activity, and fitness converged very well. As in previous studies (e.g., Proyer, 2017; Proyer and Brauer, 2018), playfulness and its facets were accurately perceived by their acquaintances. Thus, it can be concluded that peer-raters can observe playfulness and physical functioning well, and that their perception does not differ strongly from the self-perceptions by the individuals. These findings support the notion that the results obtained for self-reports are not an artifact due to the usage of the same method of assessment. This is also corroborated by the fact that most relationships of playfulness and health, activity,

and fitness were somewhat parallel in self- and peer-reports (despite the expected lower coefficients in the analyses of the peer-ratings). However, vector correlation analyses indicated that relationships between self- and peer-ratings of *Intellectual* playfulness differed from each other independently of whether the outcome was assessed by self- or peer-ratings. In line with previous studies (e.g., Proyer and Brauer, 2018; Proyer et al., 2018a), we argue that the comparatively low observability of *Intellectual* playfulness contributes to differences in self- and peer-views and the associations toward external variables.

When combining the findings of the different data sources and focusing on those that were found in multiple combinations of self- and peer-ratings, *Other-directed* playfulness showed positive relationships with mental health, while no relationships with physical health were observed. The global level of activity (i.e., GLA) showed the most robust associations with *Other-directed* and *Intellectual* playfulness. For physical fitness, no robust relationships across multiple data sources were observed, except for a positive correlation between flexibility and *Whimsical* playfulness and small relationships between *Intellectual* playfulness and cardiorespiratory fitness and flexibility. Health behaviors overall were mostly positively related to *Other-directed* and *Intellectual* playfulness. All facets of playfulness were positively related to pursuing an active way of life; as expected *Other-directed* playfulness was the numerically strongest correlate in self- and peer-ratings. Finally, *Intellectual* playfulness was positively related to safety and negatively to compliance, whereas *Lighthearted* playfulness positively related to substance consumption.

Thus, based on the findings of Study 1, we conclude that there are positive relationships of playfulness with activity and mental health (for *Other-directed* and *Intellectual* playfulness). In line with the literature (e.g., Yang et al., 2016), one might argue that engaging in social acts contributes to engagement in health behavior. However, it should be tested whether being high in *Other-directed* playfulness is, indeed, correlated with engagement in group-sports (e.g., being in a soccer team). The relationships between *Intellectual* playfulness and activity/health might reflect previous findings of the correlation between cognitive ability and engaging in health behaviors (e.g., Gottfredson and Deary, 2004). Thus, it is not surprising that those preferring complexity over simplicity also show higher inclinations to lead an active life and report higher fitness. Overall, the associations of playfulness with physical fitness were positive and there was a robust relationship of all playfulness facets with leading an active way of life. However, there are also some aspects of playfulness that might also have negative consequences and go along with negative health behaviors (such as substance consumption or lacking compliance). Overall, the findings showed that playfulness is mostly positively associated with indicators of mental and physical health. Thus, one might argue, that those high in playfulness are at an advantage for health-related outcomes. However, causality cannot be determined, thus, it is unclear whether playfulness facilitates being active and striving for health or vice versa.

Study 1 has several limitations. The male:female ratio was imbalanced, as women were over-represented in the sample. While we control for the impact of age and gender in our analyses, it must be acknowledged that the study should be replicated using a more balanced sample. While we would argue that the inclusion of peer-ratings is a strength of the study this approach also has certain problems. For example, we did not control for the type of acquaintance (e.g., romantic partner, friend, family member, work colleague etc.; see Funder et al., 1995), which may have an effect on the findings. There is evidence that stronger acquaintanceship contributes to the accuracy of perceiving others' personality (e.g., Watson et al., 2000) and one might expect that accuracy for judging health behaviors would also be highest in highly acquainted peers. Taking into account that all peer-raters were at least friends with whom they rated and that there was substantial self-peer convergence for all measures, one might argue that holding acquaintanceship constantly high (e.g., by exclusively employing romantic partners as peer-raters) would be the best option to gather an accurate estimate of health behaviors. On a more global level, the inclusion of the peer-ratings show that the contribution of playfulness to the understanding of the different activity-measures do not seem to be based on a joint method-bias (i.e., all self-ratings), but have substance above and beyond similarities in the way the data were collected. Another limitation concerns the assessment of physical and mental health, as these were assessed with 1-item measures. Although these correlated with measures of fitness, activity, and health behaviors single-item reliability cannot be estimated and thus warrants cautious interpretation.

Although in our sample healthy and active people were slightly overrepresented (slightly negative skewed distributions for health-related variables), the sample covered a broad range from very active and healthy to not very active and unhealthy people. The same was true for playfulness: The whole range of the theoretical scale in the playfulness measures was represented and means were comparable to what has been reported before for samples from the general population (Proyer, 2012a, 2014b).

STUDY 2

In Study 2, we address some limitations of Study 1 by testing the association of playfulness with physical activity and fitness with different, more rigorous methodological approaches, including objective measures. We used an interview approach for assessing physical activity. The strength of this approach is that we had the possibility of directly inquiring about activity and discussing questions with each of the participants, rather than have to rely on the answers in questionnaires. Of course, it must be acknowledged that these data are also self-reports. Additionally, we administered a broad array of field tests, covering cardiorespiratory fitness (climbing stairs; Boreham et al., 2000), flexibility (stretching; Wells and Dillon, 1952), strength (hand-grip strength; ACSM, 2001), endurance (repeatedly standing up from a chair; Bohannon, 1995), and dexterity (placing pins with both hands simultaneously in small holes on a metal plate; Schoppe, 1974) for objective assessments of physical fitness. These tasks were selected to represent a broad range of indicators of fitness, which allows differentiating among

these components. Hence, Study 2's main contribution is the inclusion of objectively measured indicators of fitness.

As in Study 1, playfulness was assessed with established self-report measures. However, we also included indicators of playful behaviors directly. Two types of behaviors were considered. Firstly, we observed participants during a waiting period and assessed how many playful items they interacted with deliberately during this period. The items were pre-defined (see procedure for details). The expectation was that greater levels of playfulness would be associated with more playful activities in a standardized time period. Secondly, participants completed the task for assessing their dexterity twice; namely, once under the standard condition and once in an impaired condition using goggles that simulate different levels of alcohol intoxication. These are typically used within driver education programs to demonstrate varying degrees of visual impairments due to intoxication. Participants could freely choose the level of impairment for doing this task and were allowed to try them out as long as they wanted. We expected that playful people select stronger levels of impairment as this would allow for greater expressions of their playfulness and since playfulness is associated with a mastery orientation (Proyer, 2014a) and a liking of competitiveness (Csikszentmihalyi, 1975). Finally, we assessed body height and weight to control for influences of participants' body mass index in addition to controlling for gender and age as in Study 1.

Based on the results of Study 1, we expected positive associations of *Other-directed* and *Intellectual* playfulness with all tested aspects of activity and fitness. For dexterity, we expected even stronger relationships in the impairment condition than in the standard condition, since playful people are hypothesized to adapt quicker to new circumstances and are willing to work under less structured external conditions (e.g., Proyer, 2012a, 2014a). Finally, we expected these associations with fitness also to be present in the more objective indicators of playfulness.

Methods

Sample

A total of $N = 67$ participants (73.1% women) aged from 19 to 75 ($M = 39.21$, $SD = 18.54$) took part in Study 2. A large part of the sample (43.3%) has a degree from a university or a university of applied sciences, 40.3% have a degree allowing them to attend a university or a university of applied sciences, 14.9% completed vocational training, and 1.5% have completed mandatory school. Most participants (89.6%) are Swiss. About half of the participants were currently in a relationship (70.1%), 26.9% single, and 3.0% were divorced or separated. Most participants were employed (61.2%). Thus, based on the demographic composition, the sample is highly comparable to the sample of Study 1. Further, as planned, we have successfully over-sampled low-scorers (scores of 3 and below; 20.9%) and high-scorers (scores of 5 and above; 41.3%) for being able to differentiate among participants with more extreme expressions, rather than having a larger number of participants in the middle range. Finally, the sample was also diverse with regard to their body mass index, which ranged from 16.5 to 31.0 ($M = 22.40$, $SD = 3.23$).

A power analysis showed that the study's sample size allowed detection of correlation effects of $\rho = 0.28$ with a power of 0.80 ($\alpha = 0.05$). Thus, medium-to-large effects can be detected through conventional null-hypothesis significance testing. Therefore, we will interpret correlations in terms of its effect size. Our effect size of interest is set at $r = 0.21$ following Gignac and Szodorai's (2016) recommendations for studies on individual differences.

Instruments

Questionnaire measures

As in Study 1, the *Short Measure of Adult Playfulness* (SMAP; Proyer, 2012a) and the OLIW-playfulness questionnaire (OLIW; Proyer, 2017) were used. Internal consistencies were acceptable (SMAP: $\alpha = 0.90$; OTD: $\alpha = 0.74$, LIG: $\alpha = 0.68$, INT: $\alpha = 0.68$, WHI: $\alpha = 0.83$).

The *International Physical Activity Interview* short-form (IPAQ; Craig et al., 2003) asks about the time spent with physical activity in the past 7 days. The time spent with physical activity of low, moderate, and vigorous levels of intensity is assessed, and, additionally, the time spent sitting. Craig et al. (2003) report acceptable measurement properties.

Objective measures

We assessed participants' body height and weight using standard instruments in order to control for the influence of body mass index in subsequent analyses.

The *Hand-Grip Strength Test* (ACSM, 2001) is an indicator of isometric strength of the hand and forearm muscles. This is measured using a hand dynamometer that has to be squeezed with the dominant hand as hard as possible. The standard instruction allows the participants to try three times in a row while only the best trial is recorded. Bohannon (1998) reports high convergent validity with other measures of arm strength. In the present study, a steel spring dynamometer (Collin's) was used.

The *1-min Sit-to-Stand Test* (Bohannon, 1995) is a measure for lower body muscular strength and endurance (representative normative data for Switzerland have been published by Strassmann et al., 2013). Participants are instructed to stand up and sit back down on a chair as many times as possible during 1 min. Ritchie et al. (2005) report good reliability and convergent validity with other measurements.

The *Stair-Climbing Exercise* (STE; Boreham et al., 2000) is an indicator for overall cardio-respiratory fitness. Participants walked 100 steps (i.e., 10 flights of 10 stairs) at a fixed, metronome-paced speed (i.e., 90 steps per minute). The variable of interest was the change in heart rate. Heart rate (beats per minute; bpm) was monitored continuously during the exercise, and during short periods before (5 min) and after the exercise (1 min), for having estimates for baseline heart rate, and the recovery of the heart rate after exercise. Heart rate was measured using a commercial device (Polar H7) using a chest strap that has been shown to yield highly reliable results that are comparable to electrocardiogram measurement (Wang et al., 2017).

The *Sit and Reach test* (Wells and Dillon, 1952) is a test of back and leg flexibility. Participants sit on the floor and bend their arms forward, as far as possible. The standard instruction allows

the participants to try two times in a row while only the best trial is recorded. Wells and Dillon (1952) report good reliability and validity of this task, and others report good criterion validity for hamstring extensibility (Mayorga-Vega et al., 2014). In the present study, the "zero point," where the participant's fingertips reach as far as their feet, was set at 26 cm.

The *Test of Fine Motor Functions* (TOFMF; Schoppe, 1974) is a test for hand- and finger dexterity. The participant has to take short pins out of a box and put them as quickly as possible in the corresponding hole, with both hands simultaneously. We measured the time between the first pin is set and the last pin is set. Hamster (1980) reports good convergent validity of this task with other motor tasks. This test was conducted twice; once under standard conditions, and once under an impaired vision condition, using "Drunk buster goggles" that simulate reduced alertness, slowed reaction time, confusion, visual distortion, alteration of depth and distance perception, reduction of peripheral vision, poor judgment and decision making, and lack of muscular coordination.

Procedure

As in Study 1, we recruited participants through online-advertisements (forums, mailing lists), and by contacting participants from earlier (unrelated) studies. The participants completed online versions of the playfulness instruments. After completion, they were invited to the lab study that took place at the University of Zurich. Depending on the availability of the participants, the lab study took place a couple of days to several weeks after the completion of the online study. Three instructors (two psychology students and one graduate student) were trained by the principal investigators to conduct the lab studies. This study had two parts, an activity part and a playfulness part. At the beginning of the lab study, participants were informed again on the study and gave informed consent. Afterwards, the participants' height and weight were measured and participants put on the chest strap for measuring the heart rate. Then, the interview on the physical activity during the last 7 days (IPAQ) was conducted. The administration of the IPAQ took about 5 min on average. We measured the participants' heart rate during 5 min while sitting. Afterwards, participants conducted the Stair-Climbing Exercise. During the stair climbing, the heart rate was continuously measured. Immediately after the stair climbing, participants sat down for 1 min during which the recovery of the heart rate was measured. Afterwards, participants completed the Hand-Grip Strengths Test, the Sit-To-Stand Test, and the Sit-And-Reach Test.

Subsequently, participants were told that the instructor has to process the so far collected data and they have to wait for about 5 min. However, they were invited to look at some materials that will be used later on in the experiment (six in total). Participants were introduced to three "drunk buster goggles" that simulate different degrees of alcohol intoxication (i.e., 0.04–0.06%, 0.06–0.08%, and 0.08–0.15% blood alcohol content [BAC]). The participants were invited to try out and play with some items (e.g., a yo-yo, a Rubik's cube, a marble labyrinth). During these 5 min, the instructor was paying attention to which goggles the participants try on and what items they played with, for assessing objective indicators of playful behaviors. Afterwards, participants conducted the Test of Fine Motor Functions twice: Once in a normal, unimpaired condition and a second time using a drunk buster goggle of their choice. Together with the participants' behaviors during the waiting time, this choice should serve as a more objective indicator for playfulness. Finally, the participants were debriefed and received a reimbursement of 25 Swiss Francs for their participation. The full procedure of Study 2 is given in **Table 4**.

TABLE 4 | Schematic of study 2.

Online	Self-report measures on playfulness (SMAP, OLIW)
On Site	Introduction; informed consent
	Assessment of body height and weight
	Interview on physical activity (IPAQ)
	Activity Part
	• Cardiorespiratory measure (Stair-Climbing Exercise; including Baseline and Recovery assessment of heart rate)
	• Hand-Grip Strength Test
	• Sit-To-Stand Test
	• Sit-And-Reach Test
	Playfulness Part
	• Waiting period (5 min)
	• Test of Fine Motor functions (unimpaired)
	• Test of Fine Motor functions (impaired)
	Debriefing

Results

First, we analyzed the relationships of playfulness and its facets with the data from the interview on physical activity in the last 7 days, in order to replicate the findings of Study 1 using a different methodology (**Table 5**) and while also controlling for gender and age. Additionally, we have controlled for body mass index due to its relationship to both playfulness (mostly *Lighthearted* playfulness [$r = 0.27$, $p = 0.028$] but also nonsignificant trends for *Intellectual* [$r = 0.14$], *Whimsical* [$r = 0.13$], and *Other-directed* [$r = 0.08$] playfulness) and fitness measures (mostly recovery heart rate [$r = 0.35$] and lower body strength and endurance, $r = -0.40$).

Table 5 shows that *Other-directed* and *Lighthearted* playfulness were positively related to the total amount of time spent physically active in the last 7 days as derived from the interview, and negatively related to the amount of time spent sitting (as did global playfulness). *Other-directed* playfulness went along with more time spent with moderate or vigorous activity, whereas *Lighthearted* playfulness was positively related to the time spent walking. The other playfulness facets and global playfulness showed fewer and smaller relationships to the

TABLE 5 | Partial correlations of playfulness (SMAP) and its facets (OLIW) with self-reported activity in the last 7 days.

	SMAP	OTD	LTH	INT	WHI	Items	Goggle
Activity total	0.13	0.29*	0.26*	0.18	0.14	0.21*	0.05
Time spent walking	−0.01	0.08	0.25*	0.17	−0.15	−0.12	0.17
Time spent with moderate activity	0.14	0.24*	0.06	0.19	0.11	0.18	0.17
Time spent with vigorous activity	0.13	0.22*	0.12	0.02	0.26*	0.30**	−0.15
Time spent sitting	−0.33**	−0.25*	−0.21*	−0.15	−0.06	−0.01	−0.05

Note N = 67. All correlations are controlled for gender, age, and body mass index. SMAP, short measure of adult playfulness; OTD, Other-directed; LTH, Lighthearted; INT, Intellectual; WHI, Whimsical playfulness. Items = Number of items played with (range = 0–6); Goggle = Strength of chosen goggle (range = 0.04–0.06% blood alcohol content (BAC) = 1;0.06-0.08% BAC = 2;0.08-0.15% BAC = 3).
**p <0.05. **p <0.01. One-tailed.*

different types of activities. When looking at the more objective indicators of playfulness, we found positive associations between the number of items participants played with and total activity and the time spent with vigorous activity. Further, the level of impairment of the chosen goggle showed a tendency to go along with more time spent walking and with moderate activity.

Second, we examine the objective measures of fitness and strengths and their relationships with playfulness and its facets (**Table 6**).

The table shows that playfulness demonstrated mostly the expected pattern with objective of fitness, strength, and dexterity: Lower heart rates (i.e., average heart rate at baseline, during the stair climbing, and during recovery) went along with higher scores in global playfulness, *Other-directed* playfulness, while there were also some effects for *Intellectual* and *Whimsical* playfulness. All playfulness facets—but not global playfulness—were positively related to hand and forearm strength. *Lighthearted* and *Whimsical* playfulness went along with better back- and leg flexibility. *Other-directed* playfulness tended to go along with better lower body strength and endurance. Also, there were trends toward positive relations to hand and finger dexterity in the unimpaired condition, while *Lighthearted* and *Whimsical* playfulness went along with better performance and *Intellectual* playfulness with worse performance in the impairment condition.

Finally, we analyzed the relationships of the objective indicators of playfulness; these are the number of items the participants played with during the waiting time (ranging from 1 to 6; $M = 4.69$, $SD = 1.36$) and the strength of the chosen impairment in the goggle that is, the level of simulated alcohol intoxication. Most participants (59.7%) chose the weakest goggle simulating a blood alcohol content (BAC) of 0.04–0.06%, 16.4% chose the goggle of medium intensity (0.06–0.08% BAC), and 23.9% chose the strongest goggle (0.08–0.15% BAC). Preliminary analyses showed, that these indicators were indeed related to playfulness: The strength of the chosen goggle positively related to global ($r[62] = 0.32$, $p = 0.005$) and *Lighthearted* playfulness ($r[62] = 0.21$, $p = 0.048$), whereas the relationships with the other facets were mostly in the intended direction, but did not reach significance (*Other-directed*: $r[62] = 0.17$, $p = 0.089$; *Intellectual*: $r[62] = 0.19$, $p = 0.067$; *Whimsical*: $r[62] = 0.01$, $p = 0.480$). The number of items played with was unrelated with self-rated playfulness (all $rs < 0.09$), and the strength of the chosen goggle ($r = 0.09$, $p = 0.511$). Since these indicators do not allow for fine-grained distinctions and are rather rough assessments of playful behavior, we considered these associations with the trait-measures of playfulness to be adequate for considering these indicators as objective measures of playfulness.

Further, we found that the number of items participants played with during the waiting time was associated with greater strength and endurance in the lower body. Also, those who chose goggles simulating higher levels of intoxication tended to show a lower baseline heart rate. No other robust effects were found ($|r| < 0.16$, $ps > 0.10$) but with few exceptions, all effects were in the expected direction and were highly similar to those from the self-reports on playfulness.

Discussion and General Discussion

Study 2's main contribution is corroborating findings from Study 1 with multiple methods and adds objective measures of physical activity to research in adult playfulness. The findings widely confirmed the results of Study 1 regarding activity with the strongest relationship found for *Other-directed*, and tendencies for *Intellectual* playfulness. In this study we also found evidence in favor of *Lighthearted* playfulness. Further, objective indicators of playfulness also yielded some positive associations with activity. In contrast to Study 1, we also detected relationships between playfulness and fitness: global, *Other-directed*, and *Intellectual* playfulness positively related to cardiorespiratory fitness, while all playfulness facets positively related to measures of strength. For dexterity, only small effects in the expected direction were obtained. Also, the objective indicators of playfulness only yielded few robust associations, but were generally in the expected direction.

The more exploratory analyses on the self-selected visual impairment show that greater playfulness was associated with the selection of greater impairments, which may increase the play-experience, but also could be interpreted as a sign of competitiveness. The latter has already been discussed in previous research in its association with play and playfulness (e.g., Rogers et al., 1987; see also Csikszentmihalyi, 1975). It needs mentioning that the selection of the degree of visual impairment is only one potential behavioral indicator of playfulness. This is a comparatively new line of research (linking trait playfulness in adults with *miniature situations* representative of the actual behavior associated with the trait) and warrants further verification. As noted for so-called objective personality tests in the Cattellian tradition (allowing for the assessment of *T*-data; for an overview see Ortner and Proyer, 2018) single tests assessing a specific behavior do not correlate systematically with self-reports (as in our study). Hence, one aim of future research will be the development of further objective tests for the assessment of playfulness and aggregate them for validation studies (e.g., when relating them to self-reports) and their validation against other data sources (e.g., *L*-data). Nevertheless, a limitation for the usage of the selection

TABLE 6 | Partial correlations of playfulness and its facets with objective measures of fitness, strength, and dexterity.

				Correlations						
	M	*SD*	Range	SMAP	OTD	LTH	INT	WHI	Items	Goggle
Baseline heart rate (bpm)	780.60	100.54	56–109	−0.19	−0.26*	−0.10	−0.18	−0.07	−0.01	−0.18
Activity Heart Rate (bpm)	1190.14	140.25	88–151	−0.24*	−0.32**	−0.15	−0.25*	0.01	0.02	−0.07
Recovery Heart Rate (bpm)	980.21	200.58	59–146	−0.23*	−0.27*	−0.10	−0.14	−0.22*	−0.11	−0.14
Hand and Forearm Strength	650.92	210.69	27–127	0.13	0.30*	0.23*	0.61***	0.22*	−0.02	0.15
Back and Leg Flexibility (cm)	310.57	90.48	5–46	0.10	0.02	0.27*	0.12	0.24*	0.13	0.04
Lower Body Strength and Endurance (number of repetitions)	350.03	90.18	18–61	0.06	0.20	0.07	−0.11	−0.06	0.24*	0.09
Hand and Finger dexterity, standard condition (s)	770.15	200.89	50–173	−0.15	−0.20	−0.01	0.01	−0.04	0.04	−0.11
Hand and Finger dexterity, impairment condition (s)	1330.01	640.98	70–520	−0.02	0.04	−0.17	0.18	−0.20	–	–

N = 61-67. All correlations are controlled for gender, age, and body mass index Correlations with fine motor skills in the impairment condition are controlled for the selected goggles SMAP, Short Measure of Adult Playfulness; OTD, Other-directed; LTH, Lighthearted; INT, Intellectual; WHI, Whimsical playfulness Baseline Heart Rate = Average heart rate (bpm) while sitting for 5 min; Activity Heart Rate = Average heart rate (bpm) during stair climbing exercise; Recovery Heart Rate = Average heart rate (bpm) during 1 min after the exercise Hand and Forearm Strength = Arbitrary scale; Back and Leg Flexibility = Distance participants were able to bend their arms forward ("zero point" = 26 cm); Lower Body Strength and Endurance = Number of repetitions in the 1-min Sit-to-Stand Exercise; Hand and Finger dexterity = Seconds to complete the Test of Fine Motor Functions in standard condition or impairment condition (using a goggle simulating alcohol intoxication) Items = Number of items played with (range = 0 to 6); Goggle = Strength of chosen goggle (range = 0.04-0.06% blood alcohol content (BAC) = 1;0.06-0.08% BAC = 2;0.08-0.15% BAC = 3).

**p<0.05. **p<0.01. ***p<0.001. One-tailed tests.*

of the degree of visual impairment as an approximation for playful behavior is that it will require further validation in future studies.

Again, this study highlights the importance of differentiating among different facets of playfulness as they have different predictive value. The findings also show that a global assessment can only give a general sense of the direction of the associations, but cannot provide a more fine-grained differentiation. It should be noted that the size of the correlation coefficients between *Hand and Forearm Strength* and *Intellectual* playfulness (about 37% shared variance) is an anomaly in comparison with the other coefficients. This finding requires replication (as is warranted for the other findings) and most likely seems to be attributable to specifics of this particular sample. While we caution against over-interpretation, future research is warranted to test how robust this association is.

Several limitations in these studies must be addressed: Again, the sample is imbalanced with respect to the male:female ratio. Further, the question of the generalizability of the findings needs to be discussed. We have tested a rather diverse sample in Study 1 (also in Study 2, but of smaller size), but, of course, the samples are not fully representative, nor gender balanced. However, since most relationships with demographic characteristics were rather small, and we corrected for influences of age and gender, we do not have much reason to doubt the validity of the findings. However, it is possible that there was a sampling effect (i.e., studies on activity mainly attract active individuals). Although our sample was diverse in this regard, we did not find participants that were not physically active at all. Thus, some effects might be underestimated in the present study since no physically non-active participants were tested. While we were able to replicate some findings of Study 1 with Study 2, some reported effects (those on the more objective measures) warrant replication in future studies using independently collected and more representative samples. Although coefficients were in the expected range, the sample size allowed us to detect mainly medium-to-large effects through testing statistical significance. Hence, we interpreted smaller effects cautiously and upon effect size (cf. Gignac and Szodorai, 2016) instead of relying solely on statistical significance. However, future studies require larger sample sizes to detect potential small effects. Whereas the findings show that playfulness is related to health, activity, and fitness, it also became apparent, that the relationships are of small-to-medium size. Replication and extension is warranted.

Overall, our studies show that playfulness is positively correlated with greater levels of activity, specific aspects of fitness (mostly cardiorespiratory fitness), and specific health behaviors (mostly leading an active way of life). Whereas the different facets of playfulness vary in their relationships with indicators of physical activity, *Other-directed* playfulness seems to be the most relevant aspect to levels of fitness, activity, and health, as it yielded the strongest and most robust associations across studies and methods, followed by *Intellectual* and *Lighthearted* playfulness. A future research direction could be disentangling the effects of social activities and a social inclination in general and testing the specific contribution of *Other-directed* playfulness (e.g., being able to interact playfully with others to have more fun during the exercises, or using joint activities as an additional resource to promote fitness). *Whimsical* playfulness, on the other hand, seems to be widely unrelated to physical activity (with exceptions, though—e.g., it was associated with more reported time spent with vigorous activity in Study 2). Hence, future studies should examine the distinct effects of single facets in the prediction of health-related outcomes.

The relationships reported in these studies are generally small in terms of size but robust across different methods and cannot be explained by a method bias (shared method variance). The relationships can also be found when using more

objective measures of physical activity, rather than the self- and peer-ratings. As expected, correlations are smaller due to the use of different methods. However, most relationships between playfulness and physical activity were in the expected direction.

This research was conducted cross-sectionally therefore, does not allow for the interpretation of causality (or the direction of the relationships between playfulness, and health, activity, and fitness). However, an initial working model could be proposed to conduct research that explores the direction of the relationship and test underlying mechanisms between playfulness and physical activity, and mental and physical health. Such a model could be framed in the context of health behavior models. Personality variables can affect health through influencing a person's compliance with health-oriented behaviors (e.g., Wiebe and Smith, 1997; Vollrath et al., 1999; Kubzansky et al., 2009) and between personality and exercise behavior (e.g., Rhodes and Smith, 2006; Allen et al., 2017). However, none of these have looked at the personality trait of playfulness. One of the most influential models to be applied to physical activity is the *Theory of Planned Behavior* (TPB), a social-cognitive model that proposes attitudes and beliefs (e.g., perceived ability to be active, and perceptions about what other people think about it), influence intentions to be active, which in turn determines actual behavior (e.g., Ajzen, 1985). Individual factors such as personality are thought to differentially influence the role of these predictors in the model (see Ajzen, 2011) Indeed, research suggests that some personality factors may play a role in moderating determinants within the TPB upon physical activity (e.g., Courneya et al., 1999; Rhodes et al., 2005; Vo and Bogg, 2015), although to the best of our knowledge, playfulness has not yet been investigated. It is plausible that playfulness could exert effects upon motivational and volitional aspects of physical activity goal pursuit. For example, playful people that enjoy group exercise and socializing, or reframe activity to make it more entertaining, may be more likely to enjoy physical activity, and have routines that promote regular exercise adherence. Currently, we also expect that there are many, bi-directional links between playfulness and physical activity and overall well-being. Furthermore, these links likely operate directly and/or be mediated by current well-being/health.

As discussed earlier, there is evidence for a positive relationship between being playful and the experience of positive emotions (e.g., joy or contentment). Positive affect may have a wide ranging influence on well-being—very much in the sense of Fredrickson's (2001) notion of a positive upward spiral associated with the experience of positive emotions (see also Panksepp, 1993). This relationship may also be helpful for a better understanding of why playfulness relates to physical activity or leading an active way of life in general. More specifically, positive affect could influence physical activity goal pursuit (e.g., Cameron et al., 2015, 2017) for example via the feelings-as-information route, whereby people interpret how they feel in general to be a favorable judgment about a target health behavior. Future tests can clarify the impact of positive affective states upon physical activity levels.

Another potential pathway that warrants further investigation is the idea that playfulness improves flexible thinking (i.e., the ability to shift perspectives, seeing new solutions, and adapting to new situations). *Intellectual* types of playfulness (Proyer, 2017) may be particularly useful for the pursuit of, and engagement in physical activities (e.g., in the sense of generating interest, or for maintaining high motivation). More broadly, psychological flexibility may be a fundamental aspect of psychological health (Kashdan and Rottenberg, 2010) and may have the potential of also contributing to physical well-being and activity.

Future studies should examine these hypotheses in both acute studies and within longitudinal designs to assess potential health-benefits and outcomes (e.g., longevity) in more detail. If playing and being playful facilitates the emergence of positive emotions (e.g., Fredrickson, 2001) and, amongst others, contribute to better coping with stressors (e.g., Staempfli, 2007; Magnuson and Barnett, 2013; Proyer, 2014a), there may also be long-term effects observable (see Gordon, 2014). Finally, it should be acknowledged that in biology there is the idea that animals primarily play when they feel safe and have enough energy to do so (e.g., when being healthy; for an overview see Burghardt, 2005). Hence, there may be a different working mechanism to consider: Only those that are healthy and not exposed to severe psychological or environmental stressors and active can "afford" to play, while others must be more protective of their available resources.

The next logical step from our perspective would be devising intervention studies. For example, measuring playfulness (or specific facets) and testing whether this influences physical activity determinants and levels over time. Experience sampling methods, where participants indicate levels of playfulness, positive affect and activity across the day, could elucidate the interplay between these components, and help identify relevant situations and behaviors for intervention studies. Another area to research would be the fit between type of physical activities pursued and preference in specific domains of playfulness. In this respect, observation studies of extreme groups in natural environments would be helpful (e.g., observing sport teams, or other people pursuing different types of physical activities) to relate behavior to playfulness. Appealing to an adult's sense of playfulness seems advisable when developing interventions to increase engagement in physical activity. One might think of the development of a program that facilitates physical activity in a way that demands certain levels of playfulness (e.g., by embedding competitions, facilitating playful interactions with others, playing for a "reward," or having playful reinforcers such as those involving humorous content). This trend has already started and various behavior change programs have been developed with the goal of "gamifiying" more traditional interventions (e.g., Howells et al., 2016), and has also been suggested for interventions aiming at addressing health (Cugelman, 2013).

A caveat of our research is that we have not covered potential negative effects of play and playfulness. For example, Burghardt (2005) lists examples of cruel aspects of play (e.g., when cats play with their prey, killing it slowly) as well as its risky, dangerous,

or addictive components (e.g., when playing risky types of sports and similar behavior)—hence, play and playfulness may not always be fun nor positively contributing to health and well-being (e.g., when experimenting with substances). Further, we did not assess variables such as consumption of legal and illegal drugs, medical conditions, or physical restrictions—these variables might also affect the relationships between playfulness, activity, and health, or limit the relationships that can be found in such a study (a playful individual could be more active if s/he would not suffer from a medical condition). Therefore, it may be necessary to have an even more fine-grained list of actual play-behaviors (e.g., Proyer, 2017) and test which of them predicts health-behaviors and activity. The latter may also be helpful to address the question of whether playful people engage (exceedingly) in computer/online-based games and games played on mobile devices and whether this may also have detrimental effects on physical activity levels.

Finally, it would be interesting to extend the research on playfulness to other physical and mental skills; it might be the case that highly skilled individuals in a broad array of fields (e.g., arts, sports, or intellectual domains) might be higher in playfulness than the general population since learning and practicing skills might be facilitated by playfulness. In fact, Study 2 provided some first insights on how playful individuals might acquire skills: Those high in playfulness chose to perform the fine-motor task in a more playful way in the self-selected vision impairment condition–even if there was no necessity to do so and this leads to a decreased performance in a current task. However, this attitude of seeking and playing with challenge might explain how playfulness adds to the acquisition and mastering of new skills.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Swiss Psychological Association with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. According to the local ethics committee (Kantonale Ethikkommission Zürich), the present study did not require a formal approval.

AUTHOR CONTRIBUTIONS

RP, FG, and EB: conception and design of the work; RP and FG: data collection; RP, FG, EB, and KB: data analysis and interpretation; RP and FG: drafting the article; RP, FG, EB, and KB: critical revision of the article; RP, FG, EB, and KB: final approval of the published version.

FUNDING

This research was funded by Unilever R&D, United Kingdom.

ACKNOWLEDGMENTS

The authors are grateful to Corina Passini and Melanie Sirocic for their help with the data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01440/full#supplementary-material

REFERENCES

ACSM (2001). *ACSM's Resource Manual for Guidelines for Exercise Testing and Prescription*. Philadelphia, PA: Lippincott Williams & Wilkins.

Ajzen, I. (1985). "From intentions to actions: a theory of planned behavior," in *Action-control: From Cognition to Behavior*, eds J. Kuhl and J. Beckman (Heidelberg: Springer), 11–39.

Ajzen, I. (2011). The theory of planned behaviour: reactions and reflections. *Psychol. Health* 26, 1113–1127. doi: 10.1080/08870446.2011.613995

Allen, M. S., Magee, C. A., Vella, S. A., and Laborde, S. (2017). Bidirectional associations between personality and physical activity in adulthood. *Health Psychol.* 36, 332–336. doi: 10.1037/hea0000371

Amabile, T. M., Hill, K. G., Hennessey, B. A., and Tighe, E. M. (1994). The work preference inventory: assessing intrinsic and extrinsic motivational orientations. *J. Pers. Soc. Psychol.* 66, 950–967. doi: 10.1037/0022-3514.66.5.950

Barnett, L. A. (1991). Characterizing playfulness: correlates with individual attributes and personality traits. *Play Cult.* 4, 371–393.

Barnett, L. A. (2007). The nature of playfulness in young adults. *Pers. Indiv. Differ.* 43, 949–958. doi: 10.1016/j.paid.2007.02.018

Barnett, L. A. (2011). How do playful people play? gendered and racial leisure perspectives, motives and preferences of college students. *Leisure Sci.* 33, 382–401. doi: 10.1080/01490400.2011.606777

Bateson, P., and Martin, P. (2013). Play, playfulness, creativity, and innovation. Cambridge: Cambridge University Press.

Bohannon, R. W. (1995). Sit-to-stand test for measuring performance of lower extremity muscles. *Percept. Motor Skill* 80, 163–166. doi: 10.2466/pms.1995.80.1.163

Bohannon, R. W. (1998). Hand-grip dynamometry provides a valid indication of upper extremity strength impairment in home care patients. *J. Hand Ther.* 11, 258–260. doi: 10.1016/S0894-1130(98)80021-5

Boreham, C. A. G., Wallace, W. F. M., and Nevill, A. (2000). Training effects of accumulated daily stair-climbing exercise in previously sedentary young women. *Prev. Med.* 30, 277–281. doi: 10.1006/pmed.2000.0634

Borkenau, P., and Liebler, A. (1992). Trait inferences: sources of validity at zero-acquaintance. *J. Pers. Soc. Psychol.* 62, 645–657. doi: 10.1037/0022-3514.62.4.645

Bös, K., Abel, T., Woll, A., Niemann, S., Tittlbach, S., and Schott, N. (2002). Der fragebogen zur erfassung des motorischen funktionsstatus (FFB-Mot) [The Physical Fitness Questionnaire (FFB-Mot)]. *Diagnostica* 48, 101–111. doi: 10.1026/0012-1924.48.2.101

Brauer, K., and Proyer, R. T. (2017). Are Impostors playful? testing the association of adult playfulness with the impostor phenomenon. *Pers. Indiv. Differ.* 116, 57–62. doi: 10.1016/j.paid.2017.04.029

Bruner, J., Jolly, A., and Sylva, K. (1976). *Play: Its Role in Development and Evolution*. New York, NY: Basic Books.

Burghardt, G. M. (2005). *The Genesis of Animal Play: Testing the Limits*. Cambridge, MA: MIT Press.

Cameron, D. S., Bertenshaw, E. J., and Sheeran, P. (2015). The impact of positive affect on health cognitions and behaviours: a meta-analysis of the experimental evidence. *Health Psychol. Rev.* 9, 345–365. doi: 10.1080/17437199.2014.923164

Cameron, D. S., Bertenshaw, E. J., and Sheeran, P. (2017). Positive affect and physical activity: Testing effects on goal setting, activation, prioritisation, and attainment. *Psychol. Health* 33, 1–17. doi: 10.1080/08870446.2017.1314477

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56, 81–105. doi: 10.1037/h0046016

Chick, G. (2001). What is play for? sexual selection and the evolution of play. *Play Cult. Stud.* 3, 3–25.

Chick, G., Yarnal, C., and Purrington, A. (2012). Play and mate preference: testing the signal theory of adult playfulness. *Am. J. Play* 4, 407–440.

Connolly, B. S., and Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychol. Bull.* 136, 1092–1122. doi: 10.1037/a0021212

Courneya, K. S., Bobick, T. M., and Schinke, R. J. (1999). Does the theory of planned behavior mediate the relation between personality and exercise behavior?. *Basic Appl. Soc. Psychol.* 21, 317–324.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., et al. (2003). International physical activity questionnaire: 12-country reliability and validity. *Med. Sci. Sports Exerc.* 35, 1381–1395. doi: 10.1249/01.MSS.0000078924.61453.FB

Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *J. Hum. Psychol.* 15, 41–63. doi: 10.1177/002216787501500306

Cugelman, B. (2013). Gamification: what it is and why it matters to digital health behavior change developers. *JMIR Serious Games* 1:e3. doi: 10.2196/games.3139

Fagen, R. (1981). *Animal Play Behavior*. New York, NY: OUP.

Fredrickson, B. L. (1998). What good are positive emotions? *Rev. Gen. Psychol.* 2, 300–319.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: the broaden-and-build theory of positive emotions. *Am. Psychol.* 56, 218–226. doi: 10.1037/0003-066X.56.3.218

Fredrickson, B. L. (2003). The value of positive emotions: the emerging science of positive psychology is coming to understand why it's good to feel good. *Am. Sci.* 91, 330–335. doi: 10.1511/2003.4.330

Funder, D. C., Kolar, D. C., and Blackman, M. C. (1995). Agreement among judges of personality: interpersonal relations, similarity, and acquaintanceship. *J. Pers. Soc. Psychol.* 69, 656–672. doi: 10.1037/0022-3514.69.4.656

Gignac, G. E., and Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Pers. Indiv. Diff.* 102, 74–78. doi: 10.1016/j.paid.2016.06.069

Glynn, M. A., and Webster, J. (1992). The adult playfulness scale: an initial assessment. *Psychol. Rep.* 71, 83–103. doi: 10.2466/pr0.1992.71.1.83

Gordon, G. (2014). Well played: the origins and future of playfulness. *Am. J. Play* 6, 234–266.

Gottfredson, L. S., and Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Curr. Dir. Psychol. Sci.* 13, 1–4. doi: 10.1111/j.0963-7214.2004.01301001.x

Hamster, W. (1980). *Die Motorische Leistungsserie – MLS (Handanweisung/Handbook)*. Mödling: Schuhfried.

Hofstee,W. K. B. (1994). Who should own the definition of personality? *Eur. J. Pers.* 8, 149–162.

Howells, A., Ivtzan, I., and Eiroa-Orosa, F. J. (2016). Putting the 'app' in happiness: a randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. *J. Happiness Stud.* 17, 163–185. doi: 10.1007/s10902-014-9589-1

Idler, E. L., and Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *J. Health Soc. Behav.* 38, 21–37. doi: 10.2307/2955359

Kashdan, T. B., and Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clin. Psychol. Rev.* 30, 865–878. doi: 10.1016/j.cpr.2010.03.001

Kubzansky, L. D., Martin, L. T., and Buka, S. L. (2009). Early manifestations of personality and adult health: a life course perspective. *Health Psychol.* 28, 364–372. doi: 10.1037/a0014428

Lieberman, J. N. (1977). *Playfulness: Its Relationship to Imagination and Creativity*. New York, NY: Academic Press.

Lyubomirsky, S., King, L., and Diener, E. (2005). The benefits of frequent positive affect: does happiness lead to success?. *Psychol. Bull.* 131, 803–855. doi: 10.1037/0033-2909.131.6.803

Magnuson, C. D., and Barnett, L. A. (2013). The playful advantage: how playfulness enhances coping with stress. *Leis. Sci.* 35, 129–144. doi: 10.1080/01490400.2013.761905

Mannell, R. C. (1984). A psychology for leisure research. *Loisir Soc. Soc. Leis.* 7, 11–21.

Mayorga-Vega, D., Merino-Marban, R., and Viciana, J. (2014). Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: a meta-analysis. *J. Sports Sci. Med.* 13, 1–14.

Miilunpalo, S., Vuori, I., Oja, P., Pasanen, M., and Urponen, H. (1997). Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *J. Clin. Epidemiol.* 50, 517–528. doi: 10.1016/S0895-4356(97)00045-0

Murray, H. A. (1938). *Explorations in Personality*. Oxford: Oxford University Press.

Ortner, T. M., and Proyer, R. T. (2018). "Behavioral and performance measures of personality," in *Encyclopedia of Personality and Individual Differences*, eds V. Zeigler-Hill and T. K. Shackelford (Cham: Springer). doi: 10.1007/978-3-319-28099-8_1281-1

Panksepp, J. (1993). "Rough and tumble play: a fundamental brain process," in *SUNY Series, Children's Play in Society. Parent–child Play: Descriptions and Implications*, ed K. MacDonald (Albany, NY: State University of New York Press), 147–184.

Proyer, R. T. (2011). Being playful and smart? the relations of adult playfulness with psychometric and self-estimated intelligence and academic performance. *Learn Indiv. Differ.* 21, 463–467. doi: 10.1016/j.lindif.2011.02.003

Proyer, R. T. (2012a). Development and initial assessment of a short measure for adult playfulness: the SMAP. *Pers. Indiv. Differ.* 53, 989–994. doi: 10.1016/j.paid.2012.07.018

Proyer, R. T. (2012b). Examining playfulness in adults: testing its correlates with personality, positive psychological functioning, goal aspirations, and multi-methodically assessed ingenuity. *Psychol. Test. Assess. Model.* 54, 103–127.

Proyer, R. T. (2013). The well-being of playful adults: adult playfulness, subjective well-being, physical well-being, and the pursuit of enjoyable activities. *Eur. J. Humor Res.* 1, 84–98. doi: 10.7592/EJHR2013.1.1.proyer

Proyer, R. T. (2014a). Perceived functions of playfulness in adults: does it mobilize you at work, rest, and when being with others? *Eur. Rev. Appl. Psychol.* 64, 241–250. doi: 10.1016/j.erap.2014.06.001

Proyer, R. T. (2014b). Playfulness over the lifespan and its relation to happiness. *Z. Gerontol. Geriat.* 47, 508–512. doi: 10.1007/s00391-013-0539-z

Proyer, R. T. (2017). A new structural model for the study of adult playfulness: assessment and exploration of an understudied individual differences variable. *Pers. Indiv. Differ.* 108, 113–122. doi: 10.1016/j.paid.2016.12.011

Proyer, R. T. (2018). Playfulness and humor in psychology: an overview and update. *Humor* 31, 259–271. doi: 10.1515/humor-2016-0080

Proyer, R. T., and Brauer, K. (2018). Exploring adult playfulness: examining the accuracy of personality judgments at zero-acquaintance and an LIWC analysis of textual information. *J. Res. Pers.* 73, 12–20. doi: 10.1016/j.jrp.2017.10.002

Proyer, R. T., Brauer, K., Wolf, A., and Chick, G. (2018a). Adult playfulness and relationship satisfaction: an APIM analysis in romantic couples. Manuscript submitted for publication.

Proyer, R. T., Gander, F., and Wellenzohn, S. (2014). *The General Level of Activity Questionnaire*. Unpublished Instrument. Zurich: University of Zurich.

Proyer, R. T., Gander, F., Wellenzohn, S., and Ruch, W. (2013). What good are character strengths beyond subjective well-being? the contribution of the good character on self-reported health-oriented behavior, physical fitness, and the subjective health status. *J. Posit. Psychol.* 8, 222–232. doi: 10.1080/17439760.2013.777767

Proyer, R. T., and Jehle, N. (2013). The basic components of adult playfulness and their relation with personality: the hierarchical factor structure of seventeen instruments. *Pers. Indiv. Differ.* 55, 811–816. doi: 10.1016/j.paid.2013.07.010

Proyer, R. T., and Rodden, F. A. (2013). Is the Homo Ludens cheerful and serious at the same time? an empirical study of hugo rahner's notion of ernstheiterkeit. *Arch. Psychol. Relig.* 35, 213–231. doi: 10.1163/15736121-12341262

Proyer, R. T., and Ruch, W. (2011). The virtuousness of adult playfulness: the relation of playfulness with strengths of character. *Psychol. Well Being* 1, 1–12. doi: 10.1186/2211-1522-1-4

Proyer, R. T., Tandler, N., and Brauer, K. (2018b). "Playfulness and creativity: a selective review," in *Creativity and Humor*, eds S. R. Luria, J. Baer, and J. C. Kaufman (San Diego, CA: Academic Press).

Proyer, R. T., and Wagner, L. (2015). Playfulness in adults revisited: the signal theory in German speakers. *Am. J. Play* 7, 201–227.

Rahner, H. (1948/2008). *Der spielende Mensch* [Man at play]. Einsiedeln: Johannes (orig. published 1948).

Rhodes, R. E., Courneya, K. S., and Jones, L. W. (2005). The theory of planned behavior and lower-order personality traits: interaction effects in the exercise domain. *Pers. Indiv. Diff.* 38, 251–265. doi: 10.1016/j.paid.2004.04.005

Rhodes, R. E., and Smith, N. E. I. (2006). Personality correlates of physical activity: a review and meta-analysis. *Br. J. Sports Med.* 40, 958–965. doi: 10.1136/bjsm.2006.028860

Ritchie, C., Trost, S., Brown, W., and Armit, C. (2005). Reliability and validity of physical fitness field tests for adults aged 55 to 70 years. *J. Sci. Med. Sport* 8, 61–70. doi: 10.1016/S1440-2440(05)80025-8

Rogers, C. S., Meeks, A. M., Impara, J. C., and Frary, R. (1987). "Measuring playfulness: development of the child behaviors inventory of playfulness," in *Paper presented at the Southwest Conference on Human Development* (New Orleans, LA).

Schoppe, K. J. (1974). Das MLS-Gerät: ein neuer testapparat zur messung feinmotorischer leistungen [The MLS-device: a new device for the assessment of fine motor skills]. *Diagnostica* 20, 43–47.

Singer, J. L., Singer, D. G., and Sherrod, L. R. (1980). A factory analytic study of preschoolers' play behavior. *Acad. Psychol. B.* 2, 143–156.

Staempfli, M. B. (2007). Adolescent playfulness, stress perception, coping and well-being. *J. Leis. Res.* 39, 393–412. doi: 10.1080/00222216.2007.11950114

Strassmann, A., Steurer-Stey, C., Lana, K. D., Zoller, M., Turk, A. J., Suter, P., et al. (2013). Population-based reference values for the 1-min sit-to-stand test. *Int. J. Public Health* 58, 949–953. doi: 10.1007/s00038-013-0504-z

Vazire, S. (2010). Who knows what about a person? the self–other knowledge asymmetry (SOKA) model. *J. Pers. Soc. Psychol.* 98, 281–300. doi: 10.1037/a0017908

Vazire, S., and Mehl, M. R. (2008). Knowing me, knowing you: the accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *J. Pers. Soc. Psychol.* 95, 1202–1216. doi: 10.1037/a0013314

Vo, P. T., and Bogg, T. (2015). Testing theory of planned behavior and neo-socioanalytic theory models of trait activity, industriousness, exercise social cognitions, exercise intentions, and physical activity in a representative U.S. sample. *Front. Psychol.* 6:1114. doi: 10.3389/fpsyg.2015.01114

Vollrath, M., Knoch, D., and Cassano, L. (1999). Personality, risky health behaviour, and perceived susceptibility to health risks. *Eur. J. Pers.* 13, 39–50.

Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., et al. (2017). Accuracy of wrist-worn heart rate monitors. *JAMA Cardiol.* 2, 104–106. doi: 10.1001/jamacardio.2016.3340

Wanner, M., Probst-Hensch, N., Kriemler, S., Meier, F., Autenrieth, C., and Martin, B. W. (2016). Validation of the long international physical activity questionnaire: influence of age and language region. *Prev. Med. Rep.* 3, 250–256. doi: 10.1016/j.pmedr.2016.03.003

Watson, D., Hubbard, B., and Wiese, D. (2000). Self-other agreement in personality and affectivity: the role of acquaintanceship, trait visibility, and assumed similarity. *J. Pers. Soc. Psychol.* 78, 546–558. doi: 10.1037/0022-3514.78.3.546

Wells, K. F., and Dillon, E. K. (1952). the sit and reach—a test of back and leg flexibility. *R. Q. Exerc. Sport* 23, 115–118. doi: 10.1080/10671188.1952.10761965

Wiebe, D. J., and Smith, T. W. (1997). "Personality and health: progress and problems in psychosomatics," in *Handbook of personality psychology*, eds R. Hogan, J. A. Johnson, and S. R. Briggs (San Diego, CA: Academic Press), 891–918.

Wiesmann, U., Timm, A., and Hannich, H.-J. (2003). Multiples Gesundheitsverhalten und Vulnerabilität im Geschlechtervergleich [A gender comparison of multiple health behavior and vulnerability]. *Z. Gesundh.* 11, 153–162. doi: 10.1026/0943-8149.11.4.153

Yang, C. Y., Boen, C., Gerken, K., Li, T., Schorpp, K., and Harris, K. M. (2016). Social relationships and physiological determinants of longevity across the human life span. *Proc. Natl. Acad. Sci. U.S.A.* 113, 578–583. doi: 10.1073/pnas.1511085112

Yu, P., Wu, J.-J., Chen, I.-H., and Lin, Y.-T. (2007). Is playfulness a benefit to work? empirical evidence of professionals in Taiwan. *Int. J. Technol. Manag.* 39, 412–429. doi: 10.1504/IJTM.2007.013503

Conflict of Interest Statement: The authors declare that this study received funding from Unilever. EB was a staff member of Unilever R&D and contributed to the objectives, design of the study and the final manuscript.

Copyright © 2018 Proyer, Gander, Bertenshaw and Brauer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 September 2018
doi: 10.3389/fpsyg.2018.01704

How Playfulness Motivates – Putative Looping Effects of Autonomy and Surprise Revealed by Micro-Phenomenological Investigations

Katrin S. Heimann and Andreas Roepstorff*

Interacting Minds Centre, Aarhus University, Aarhus, Denmark

Play and playfulness have repeatedly been suggested to promote learning and performance, also in environments traditionally not connotated with play. However, finding empirical evidence for these claims has been aggravated by the lack of a definition of play and playfulness fitting to this description. This paper proposes to consider playfulness as an attitude, mode or mental stance, that can be modulated independent of the activity pursued and of the general character of the person. It furthermore introduces the micro-phenomenological method to assess the process and outcome of such modulation. To explore this, we devised a simple building task in a controlled within-subject design, interviewing each participant on how they accomplished the task when asked to perform it so that it either felt playful or not playful. The outcomes of this data driven approach supported this notion of playfulness as a stance, and allowed for specific hypotheses about the temporal course and mechanisms of becoming playful. They suggest that an experience of autonomy and self-expression may be key to the success of the modulation. They furthermore indicate that the resulting playful state may allow for an exploratory engagement with materials that can lead to surprising results. Such unexpected results seem to enhance participants' feeling of competence which, in turn, may increase the motivation for the task. We discuss these results within the framework of Deci and Ryan's motivational theory and in relation to current research on gamification and learning.

Keywords: playfulness, micro-phenomenology, motivation, gamification, autonomy, competence, creativity

OPEN ACCESS

Edited by:
René T. Proyer,
Martin Luther University of Halle-Wittenberg, Germany

Reviewed by:
Shulamit Pinchover,
The New School, United States
Doug Maynard,
SUNY New Paltz, United States

****Correspondence:***
Katrin S. Heimann
katrinheimann@cas.au.dk

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *04 April 2018*
Accepted: *23 August 2018*
Published: *11 September 2018*

Citation:
Heimann KS and Roepstorff A (2018) How Playfulness Motivates – Putative Looping Effects of Autonomy and Surprise Revealed by Micro-Phenomenological Investigations. Front. Psychol. 9:1704. doi: 10.3389/fpsyg.2018.01704

INTRODUCTION

> *"P: I think . . . the most memorable thing was that I started smiling when you said it [*to be playful*]. And I felt like "Oh-yes!" And I felt like I could think about it and take my time instead of just rushing into it . . . So, I was... I was excited... but still. . . mh. . . calm and... or not calm... but, but like... more settled in a way. . . Before* [when advised not to be playful*]. . . I was very driven by. . . pressure, but maybe a little bit stressed and now I was just... driven by how I..., how I wanted to have fun and build these things and... and just play with it"* (Participant 3, talking about her experience to accomplish a building task in a playful stance).

Throughout decades, playing and being playful have been described as favorable or even conditional for humans' well-being, performance, development, and even cultural evolution (see for example Murray, 1938; Huizinga, 1955; Bateson and Martin, 2013; Gray, 2013). Extending on these claims, the leading hypothesis of the quickly growing field of gamification is the assumption that play elicits high levels of motivation and creative behavior and that this can be utilized in education and work contexts to boost learning and performance. Studies exploring related hypotheses have risen to millions per year (see Scholar PLOTr[1]; key words play/playfulness/gamification and motivation/creativity, etc.), though with very mixed results (see for example Hamari et al., 2014; Hanus and Fox, 2015; Nicholson, 2015; Sailer et al., 2017). It has repeatedly been suggested that this ambiguity might not be due to low correlations, plainly. Rather the problem is likely to lie in a lack of agreement on how to actually define or identify the phenomena. Already Sutton Smith, reviewing play theories of the 100 years before 1997, summarized the state of the art as follows: *"We all play occasionally, and we all know what playing feels like. But when it comes to making theoretical statements about what play is, we fall into silliness. There is little agreement among us, and much ambiguity"* (Sutton-Smith, 2009, p. 1). Interestingly, this holds still true in present time, with fun researcher De Kowen stating: *"I'm beginning to think that I'll never be able to define playfulness comprehensively enough to embrace it in its fullness. It's just too diverse, too idiosyncratic, personal, profound to allow itself to be confined into anything satisfyingly definition-like. I've come to the conclusion that the best we can do is describe experiences, instances, moments in our lives that appear, in retrospect, at least, to have proven themselves unquestionably, undeniably, overwhelmingly playful"* (De Kowen, 2017).

This paper embraces De Kowen's emphasis on the experience of being playful, but it does not concur with his claim that this experience does not allow for generalization and can only be captured anecdotally. Instead it presents an *empirical study* of the *experiential nature* of becoming playful. Our findings resonate with and extend on gamification research and allow to formulate specific hypotheses regarding the function of play and playfulness, in particular the connection between playfulness, motivation, and creativity.

The Conceptual Challenge

There have been at least three different approaches guiding the many attempts to capture play and playfulness: by focusing on features of play/playful activities, by focusing on playfulness as a character trait and by focusing on playfulness as a frame of mind.

Firstly, most of the earlier definitions have attempted to identify specific features of activities that legitimize them to be called play or playful (for example play needs to be a spontaneous activity, not rulebound, non-literal, based on active engagement etc., see Blurton Jones, 1972; Reynolds, 1976; Rubin et al., 1983). However, empirical efforts to support the respective criteria have repeatedly failed (cf. Sutton-Smith and Kelly-Byrne, 1984).

[1]https://www.csullender.com/scholar/

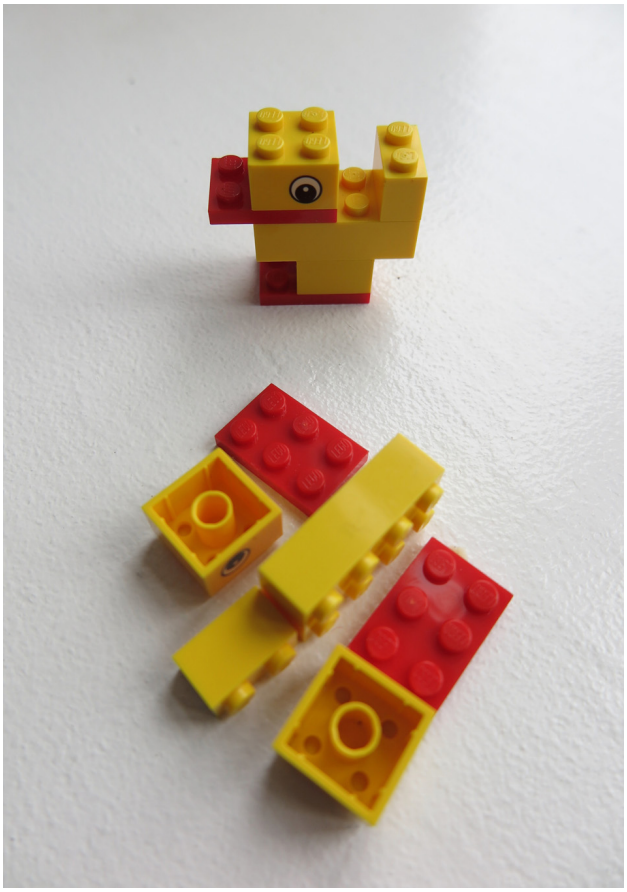


FIGURE 1 | Prototype duck and LEGO bricks set used for the task.

Secondly, Lieberman (1965, 1966) was one of the first to shift attention from the *activity* to the *participant*, thus attempting to define playfulness as a character trait, with sub-traits such as cognitive spontaneity, physical spontaneity, social spontaneity, manifest joy and sense of humor (for recent work on this see Glynn and Webster, 1992; Proyer, 2012; Shen et al., 2014; see also Barnett, 2017 for a culture comparative approach).

However, much research in the fields of psychology, education, and in particular management and business administration seems to assume that one can have a more or less playful way to fulfill a task, independent of the activity and despite the fact that persons can be playful to a different degree. Building on the theories of Murray (1938); Caillois (1961) and Apter (1991), it has thus thirdly been proposed that playfulness should be conceptualized *"as the attitude of a person when he or she is engaged mentally and physically in the state of play. [. . .] Any object can become a tool for play and any situation can be approached in a playful manner when the person is in such frame of mind"* (Arrasvuori et al., 2010, p. 2; see also Boberg et al., 2015); and elaborating on this: *"Playfulness is the expression of a universal capacity that can either be nurtured and encouraged or constrained and limited by both internal and environmental variables"* (Sanderson, 2010).

In fact, already the psychologist Susanna Millar has pleaded for such analytical shift, claiming that "*perhaps play is best used as an adverb; not as a name of a class or activities, nor as distinguished by the accompanying mood, but to describe how and under what conditions an action is performed [. . .]*" (Millar, 1968).

The Methodological Challenge

However, as argued by Sutton Smith, there is very little empirical work to further define this specific experiential state – and its relation to motivation and creativity:

> *"What do the players reckon to be the character of and the reasons for their own participation? Obviously, there is not much research to be referred to here, although there is a considerable amount of anecdotal opinion to be cited*" (Sutton-Smith, 2009, p. 16).

The "obviously" in this quote most likely refers to the circumstance that introspection, the only way to assess subjective experiences, is not a method traditionally relied on within psychology. Two main reasons seem to be responsible for this circumstance. Firstly, subjects are often untrained to attend to and to linguistically express the micro-gestures in their minds. More importantly, they are even considered untrustworthy and highly susceptible sources of confabulation about their own mental actions (see for example Nisbett and Wilson, 1977; Johansson et al., 2005; but also Petitmengin et al., 2013 and, more generally, Jack and Roepstorff, 2003). Secondly, it poses a challenge for experimental psychology to derive generalizable information (numbers) from highly individual reports of experiences from complex, real life situations as the most commonly used quantitative methods, such as averaging, do not easily lend itself to this purpose. In consequence, the use of introspection has long been excluded in cognitive psychology as a means to study the human mind. However, as often pointed out, this is a curious circumstance given the fact that the harshest critics themselves not only commonly do trust their own judgments about their minds but also use these judgments – or "anecdotes" – to derive ideas and hypotheses for their experiments. As Jack and Roepstorff stress, the purpose cannot be to condemn introspection, but to validate and expand its use as a scientific tool (Jack and Roepstorff, 2003).

Micro-phenomenology (MP) is an interview and analysis approach explicitly developed for this purpose. It aims to facilitate the access to subjective experience and to analyze and represent it in a manner fitting to the scientific aim of generalization of results (for descriptions and validity tests of the method see Petitmengin-Peugeot, 1999; Petitmengin, 2006; Petitmengin and Bitbol, 2009; Bitbol and Petitmengin, 2013; Petitmengin et al., 2013, in preparation). It is also distinctly different from other assessments of experiences in so far as it looks at the detailed unfolding of an experience over time, rather than asking for overall characteristics (as assessed by questionnaires, etc., see for example Arrasvuori et al., 2010). This allows to explore (micro)-mechanisms and causalities otherwise unattended to, and it makes the approach a potentially valuable tool when trying to understand how a certain process might be facilitated or hindered – an obvious goal when thinking of the applicability of findings about playfulness for psychology, education or business. So far, micro-phenomenology has been applied to a wide range of topics, and the granted insights may have direct applicability in clinic, education, or organization. For example, it has been used to train epilepsy patients to detect (and treat) early signs of an arriving seizure, to help academics to understand and find first steps out of a writer's block (Bojner Horwitz et al., 2013) and to assess and tackle some of the critical factors of overworking in executives (Créno and Cahour, 2015). It is furthermore used to reach a more fine-grained understanding of phenomena traditionally considered to be out of reach for empirical assessment such as emotions (as experiences, see for example Depraz et al., 2017).

Aims of This Paper

In the following we will further present this method and our precise research design developed to explore the following questions:

- *Is there empirical support for the hypothesis that playfulness can be considered an experiential stance that can be internally or externally manipulated?*
- *What are the experiential characteristics of the process of becoming playful in such way?*
- *Can experiential reports help to inform current discussions about the relation of play/playfulness and learning/performance?*

MATERIALS AND METHODS

Participants

Participants of this study were 22 young adults, 8 male, 14 female, 23.4 mean age (SD = 4.16), all but one with completed high school degree, 5 with further university degree. All were students of Vestjylland's folk high school (Vestjyllands Hojskole) located in Velling, Denmark, a boarding school offering adult education for national and international students in several creative topics such as writing, dancing, theater, music as well as sustainable thinking and entrepreneurship[2]. The students were recruited on campus as volunteers after a 30-min interactive lecture introducing the method of micro-phenomenology as a tool of cognitive science, including a live interview about a simple spelling task with one of the teachers as interviewee. This setup allowed us to get participants motivated for the task and comfortable with the setup and the reflective requirements of the method.

All data were recorded within 1 week in May 2017, in a classroom of the Hojskole. Interviews were conducted in English by the same interviewer (Katrin Heimann). All students understood and spoke English on a level allowing them to study for a university degree. All participants received 150 DKK as reimbursement. They all gave written informed consent to procedure and data use and were debriefed after the experiment. The study was carried out in accordance with

[2]http://www.danishfolkhighschools.com/about-folk-high-schools/history/

the recommendations of the guidelines of the Human Subjects Committee of the Cognition and Behaviour lab at Aarhus University and The Central Denmark Region Committees on Health Research Ethics. The protocol was approved by the Human Subjects Committee and exempted from need of approval by The Central Denmark Region Committees on Health Research Ethics.

Data Recording Procedure

A micro-phenomenological interview begins with the elicitation of a particular singular experience. This is considered necessary to avoid a mere reproduction of information or knowledge about the phenomenon in focus (as triggered by simply asking what it means to get and be playful), and instead to foster access to an actual experience. For this purpose, MP always involves a clear reference event that the interviewee repeatedly is reminded to refer to. As the aim of our study was to explore the experiential process of becoming playful, we would ideally have referred to an instance when the participant had such an experience. However, asking for a personal memory fulfilling this condition would most likely have triggered experiences of very different nature across participants, which, in turn would have compromised the analysis of the data. We also wanted to avoid creating a "playful" experience by a prechosen context modulation (such as a "gamified" design) as for this we would have had to use (and therefore prime with) our anecdotal intuitions of how to get and be playful. We therefore decided for a controlled within-subject design that gave the participants the task to decide what it takes to become and be playful or not playful:

Briefly, for each participant we prepared six equal sets of six LEGO bricks, each set allowing to build a small duck.

From one set we built a prototype duck and placed it on the table. The other five sets were arranged in separate heaps in front of the participant. Each participant was then given one out of two tasks (order counterbalanced across participants):

Task (a): "I would now like you to build five LEGO ducks out of these sets. You can rebuild the prototype you see on the table or just build any duck or duck-like creature you like – that is up to you. The only thing that is really important for us and this experiment is that you do this as playfully as you can. Please find a way of doing it, so that it feels playful and nothing but playful."

Task (b): "I would now like you to build five LEGO ducks out of these sets. You can rebuild the prototype you see on the table or just build any duck or duck-like creature you like – that is up to you. The only thing that is really important for us and this experiment is that you do it in a non-playful manner. Please find a way of doing it, so that it feels not playful at all."

Notably, this design had only a minimal difference in instructions: the suggestion to build so that it feels as playful as possible or not playful at all. This contrast allowed us to explore whether adopting a stance of playfulness would allow for similar experiential qualities across participants, independent of the particular setting, activity or character of the participant (all constant across conditions).

If participants asked for further explication for the non-playful condition, we answered that it should rather feel like work. This occurred for the majority of participants, and we will discuss this as a possible priming issue in the last section of the paper.

After each building session, we ran a micro-phenomenological interview with the participants involving the following procedure: we started out by asking the participant the following question:

> "Now, I would like you to go back to the moment in which I asked you *I would now like you to build five LEGO ducks out of these sets. You can rebuild the prototype you see on the table or just build any duck or duck-like creature you like – that is up to you. The only thing that is really important for me and this experiment is that you feel as playful as you can when doing it/you don't feel playful at all when doing it.* Please take your time to go back to that moment and then tell me what you experienced. How did you accomplish this task?"

We used the principles of the micro-phenomenological interview technique to guide the interviewee to and through her experience avoiding to prime for certain answers or foster confabulation a posteriori. The main tool for this is that the interviewer, after the initial open question, only unfolds the answers of the participants in their own words. Thus, he mainly suggests to dive deeper into certain experiential episodes by repeating the participants' phrasing and asking for further explication of the actual experience in all its dimensions. Most importantly, he avoids to prime the participant, e.g., by reformulating the experience according to his own experiences or prior knowledge or by asking for dimensions not mentioned by the participants himself. For more information see Petitmengin (2006).

In each interview, when the context allowed it, we furthermore asked "Did you manage to become playful/non-playful?" and – depending on the answer – "How did you experience it as playful/non-playful?" (and asking deeper into this as explained above). This allowed to explore the participants' success in shifting his/her own inner stance and to explore how the result of the effort felt like with reference to the specific task.

Participants also filled out a questionnaire that explored basic demographics and asked about the overall experience in terms of *playfulness*, *enjoyment*, *fun,* and *duration perception*. We furthermore recorded videos, photographs of the ducks built, measured heart rate and galvanic skin response. Finally, we administered a short version of the Torrence creativity task post-experiment. The current paper only analyzes and discusses the interview data (based on video recordings) and the demographics. Complete interview data can be provided by the authors upon request.

Data Analysis

Micro-phenomenology is based on the "elicitation interview" technique developed by Vermersch (1994) to help practitioners reflect their own praxis in the field. Thus, the original technique was designed to reveal individual thought and action processes to enhance each participants' specific professional activities. More recently, micro-phenomenology has been adapted and refined for use in cognitive science, thus focusing more on how to generalize across participants. Obviously, this is dependent on a

well-designed protocol. In this study we used the following four-step procedure based on Vermersch (1994, regarding the pre-analysis), Petitmengin (2001, 2006, 2009, regarding time analysis and analysis proper), and especially Depraz et al. (2017, regarding the generative aspects of the analysis proper and the constructive analysis).

(1) **Pre-analysis**: This step comprises

   (a) *transcription* of the data, providing the raw verbal text, "the Verbatim";
   (b) direct and ongoing *commentaries* along the transcription;
   (c) *decoupling*, that is differentiating in the Verbatim between utterances describing actual lived experience and "satellite information" such as commentaries, theoretical generalizations, context information etc.

The end result of this analysis step consists of transcripts of the participants' utterances (with interviewer's speech removed) that describes only her lived experience (after satellite information has been removed).

(2) **Time-analysis**: This step consists in reorganizing the Verbatim of the lived experience so that it follows the timeline of the actual experience rather than the timeline of the interview.

(3) **Analysis proper**: This step involves analyzing the decoupled Verbatim with regard to structural categories, identified *a priori* on the basis of knowledge about dimensions of experience in general and *a posteriori* based on the experiential dimensions revealed in the actual data. The end result of this analysis step is a representation of the experience of each participant that explicates the type of experience each temporal phase contains: such as a memory, an imagination, a linguistic thought, a feeling etc. The following categories and subcategories guided our analysis in this phase (they were generally derived from Depraz et al., 2017, but see further specifications below):

   (a) *body* – referring to participants' mentioning of (1) kinesthesis, that is bodily alterations/movements and, in our case, also distinct actions recalled; (2) perceptions, in our case including also more abstract perceptions, such as a pressure from outside etc., to the degree these are given a bodily basis.
   (b) *cognition* – comprising references to (1) memories, (2) imaginations, (3) attentional moves (focused/open), or (4) (meta) thought processes.
   (c) *feeling/emotion* – comprising references to affective reactions to the protocol that could be categorized as emotions such as happiness or as more diffuse states such as feeling ill at ease/stressed/bored etc. In the results we have grouped category (c) with category (a) as there was a direct link of perception and feeling.

(4) **Constructive analysis**: This step involves building a reduced model of the experience in general, keeping in mind the initial research questions. Thus, for each participant, we reviewed the results from the **Time Analysis** (2) as well as the **Analysis Prope**r (3) both with respect to the two conditions and with respect to participants' reports of whether they succeeded in entering the specific states. This allowed to highlight structural differences between the conditions (playful/non-playful). We also used this approach to explore if there was an order effect (when participants were first asked to build in a playful or in a non-playful way), but we did not find any systematic differences.

It is important to mention that the interviews and the analysis attempted to map the experiential process of becoming playful/non-playful, without explicitly frontloading specific theoretically motivated elements into the analysis. To the extent possible, the outcome of the analysis can therefore be considered data driven rather than hypothesis driven.

The primary aim was to analyze interviews to get insight into the experiential structure of becoming playful. However, we also used the videos to document the kind of products (duck-figures) produced in each condition and we evaluated the interviews in order to see if participants succeeded in manipulating the playfulness of their stance. In the following, we provide numbers (indicating how many participants out of 22 responded in a certain way) when generalizations across participants were possible, and we use direct quotes to either illustrate the generalizable phenomenology or to hypothesize about phenomena not touched upon in enough interviews to make an overall claim. The quotes are marked with PF (playful) and NPF (non-playful).

RESULTS

Success of Modulation of Playfulness

The task to build while modulating one's own feeling of playfulness seemed feasible for most participants. Only 3 out of 22 participants reported a difficulty to achieve a playful stance. To build in a non-playful stance seemed to be slightly harder with 12 participants reporting difficulties (though not failures). As the experiential analysis revealed, this might be related to the use of the LEGO material facilitating a playful stance rather than a non-playful stance. We will refer to these instances with more detail below.

Experiential Structure and Products

In total, we identified four different phases of the experience in general:

(1) Modulation of the playful stance
(2) Imaginative building preparations
(3) Building
(4) Product evaluation.

While there might be other phases involved in other tasks, we assume that the ones detected here cover more general process characteristics of any such task, such as a preparation phases (1 and 2), a conduct phase (3) and an evaluation phase (4). In the following we will describe each of the phases with respect to the experiential categories as defined above. To give a full picture, we also include third-person reports of building products or behavior in these otherwise first-person descriptions (differences clearly marked in the text).

Modulation of Playful Stance

The majority of participants reported about thoughts about how to modulate their own playfulness that occurred to them right after they received the task. Similar micro-experiences also happened in later phases of the task, seemingly as the result of participants evaluating the state achieved so far and possibly trying to correct it to better fit the goal (that is to be playful/non-playful). In the following, we report about the precise content and phenomenology of this phase using the categories described above.

Cognition, meta-thoughts/inner speech

Almost all participants reported the occurrence of conscious linguistic thoughts about the meaning of the task given. Strikingly, the vast majority of these reflections addressed a modulation of the feeling of autonomy as well as certain directions for how and what to build. In the playful condition, participants mentioned that the demand to be playful essentially meant them to be set free to do whatever they wanted to (14 participants) or to create something clearly inspired by their own ideas or intuitions rather than any pre-given options (7 participants). In six cases, these thoughts were reported to have come to participants as direct inner speech, giving the claims a very distinct and confident character, see for example:

> *"P: Ehm. . . I remember smiling eh. . . first when you said it... that I was about to play... I immediately started smiling... and eh... then I thought "Okay now I can do whatever I want and I can take my time" (P3, PF).*

Fifteen participants in total, still in the playful condition, furthermore described that they experienced an urge to be creative or to build different ducks – at least as long that this would not lead to a stressful experience:

> *"P: I think my first thought was to like make five different, because I found that kind of playful before . . . but then I. . . also. . .thought back to that experience and it was also. . . actually kind of. . . work-like I think, because it was also bit stressful, you know, to have to make five different... that look like a duck... so, mh. . . I think I just. . .to make it really playful it had to be. . . like I could do it different, but it didn't have to be like five different ducks. . ." (P7, PF).*

In contrast, in the non-playful condition, 16 participants reported thoughts indicating that experiencing constraints and stress was taken as essential for fulfilling the condition. This included a constraint of building as fast as possible (time pressure) and of fulfilling certain expectations on how the product should look like (evaluation pressure). See for example:

> *"P: I thought a lot more about speed... And I was. . . much more worried with what the two of you would think. . . Ehm. . . yeah, but speed was the first thing that popped into my mind, in order to making it feel more like work, and then at some point correctness" (P6, NPF, focus on time pressure).*

or also:

> *"P: I was thinking about that I should do this with more than five ducks and just keep on going. And that it needed to be the same as this one [points at prototype duck] . . . it's like, I couldn't use my imagination and I just needed to produce... in the right way" (P4, NPF, focus on evaluation pressure).*

The last quote is also an example of one of 13 participants who explicitly mentioned having thought that the task to be non-playful directly implied copying the prototype duck. Elaborations from four of these participants suggested that this might be related to the expectation that copying is a meaningless and boring activity.

Notably, still in the non-playful condition, two participants reported these thoughts occurring to them in the form of inner speech. However, the experience seemed very different from that of the playful mode, in which participants reported hearing their own voice contently noting the freedom given by the instructions (see above). In contrast, in the non-playful condition participants reported a constant reminder from "a" (thus not necessarily their own) voice asking them to get going. See for example:

> *"P: It's just like... with like working I... I kind of heard the sound like in the military. . . like "Working! Go! [tuftuf] [. . .] I just. . . remember there was this like voice "come on... make ducks" like a voice saying like "come on" or something like. . .*
>
> *I: Okay. There was a voice that told you to get... get on with it?*
>
> *P: Yeah, not. . .not like pep-talkish, but like "do it. . .do it"" (P19, NPF).*

Body, perceptions; feeling/emotion

Strikingly, many participants reported that their thoughts and inner voices were accompanied by immediate bodily reactions fitting the assumed affordances of the conditions. Of the participants indicating that they reached a playful stance, nine mentioned the immediate *feeling* of being relieved of any or certain obligations, as well as being encouraged to explore and enjoy themselves. See for example:

> *"P: Ehm... well, I think the. . . the most memorable thing was that I started smiling when you said it. And I felt like oh – yes! And like, I didn't. . . and also yeah, I like I. . . I felt like I could think about it and take my time instead of just rushing into it. So, I was. . . I was excited... but... but still mh. . . calm and... or not calm... but, but like. . . more. . . more settled in a way. . ."(P3, PF).*

In contrast, of the participants indicating that they reached a non-playful stance, 16 reported the *feeling* of an obligation to fulfill the task in one or the other way, causing them stress and partly also boredom:

> *"P: It was maybe a bit more stressful, because I felt I should do it in the right way. . . So I needed to [gestures repetition]. . . get over, do it again, if it. . . if it wasn't right" (P4, NPF).*

This happened even when participants explicitly noted that they themselves were the ones actually being in charge about the actions to be performed. See for example:

"P: You said that it was my decision. You said that. . . [I: mh] I could [I: yeah] make them however I [I: yeah] wanted, but. . . when you said that it was work and not play, I automatically assumed that I was creating a product that was. . . supposed to look like the original model" (P3, NPF).

Notably, the three participants that admitted difficulties in reaching a playful stance connected this to the feeling of heteronomy (that is being determined by other circumstances than own will). While one participant explained that she also tried to accomplish the task "correctly," a demand which to her compromised her feeling of playfulness, the other two explicitly stated that being given a task or being in an experimental situation in general had hindered them reaching a real playful stance.

Cognition, memories

12 participants reported the occurrence of memories directly related to the modulation of playfulness.

In the playful condition, these were without exceptions memories of play during own childhood or of playing (as an adult) with children; while in the non-playful condition, participants recalled mostly memories of prior or current work places. Underlining this difference again, the five participants, for whom the building activity in the non-playful condition evoked memories of play in childhood, reported some degree of difficulty achieving the non-playful state. See for example:

"I: Did feel like working actually? P: Mmh. . . kind of yeah. . . but. . . I still have a lot of memories with LEGO... So, in that way it was still... I mean I remember ehm. . . these brochures with different things you could build. . . like big castles and stuff. . . And I remember my father and I playing with them and... and building them according to these brochures. So, in that way it... it had a hint of that and it... it like... yeah but...but mostly it felt like a factory job to a degree" (P3, NPF).

Conditions further differed regarding the occurrence of those memories. In the task to build playfully, participants reported memories happening like a flashback, without conscious intent. In contrast, for at least four cases in the non-playful condition, the memory seemed to have been intently evoked to help the elicitation of a certain inner stance. The following quote shows how such effort facilitated getting into a work mood in the task given:

"P: Okay... I. . . was thinking about eh. . . my university. . . because, I am studying architecture and we usually have to build something... models, yeah. . . and I tried to be in that position. . . [. . .]

I imagined, if I were. . . eh... in the work place. . . with my. . . with my... partners and. . . [laughs] I don't know. . . I... I tried to imagine it. . .the surroundings. And. . . (pauses)

I: How was it? How. . . How was that situation for you?

P: It was so many people in my room and. . . all of them. . . did their own work and. . . and their own eh. . . projects

I: Mh. . . And how did it feel to be there?

P:...Like I needed to do it

I: You needed to do what?

P: To... to build ducks, but not exactly the same ducks.

I: Mh. So, but how did you... So, this was a transfer from the university context to here?

P: Yes" (P 3, NPF).

Cognition, attentional move

In both conditions, participants reported a heightened attention for the task. However, while in the playful condition such state seemed to come naturally with the building, in the non-playful condition, 4 participants reported this move as an effortful activity, demanded by the task:

"I: Do you remember anything else of your body feeling? Or of your thoughts when you were building the ducks? P: Ehm. . . I... I. . .I think it... this... this whole being drawn in. . . I am building something, which sucks me in [. . .] it takes my focus" (P2, PF, no effort).

"Well I had to do it like working...so I tried to be more focused. . . like saying, at least when doing the first copy. . . like. . . "okay." I... I would normally just put it together, but here I am like... "okay. . . take it step by step... placing the first brick. . . then the next one... "and building up from there. . . bit more metho... dolo... gically. . ." (P1, NPF, effort).

To sum up, participants' descriptions indicated that from the very start of the experience, the different tasks triggered micro-experiences clearly differing between conditions regarding content and experience. In the playful condition participants indicated pleasant conscious thoughts about the association of such mood with the experience of autonomy and creative production. Noteworthy, such thoughts seemed to be accompanied by immediate feelings fitting these requirements, such as relief, freedom, inspiration, and enjoyment. Such feelings might have been facilitated by spontaneous memories to pleasant childhood play experiences and an effortless raise of attention for the task. In the non-playful condition on the other hand, participants indicated demanding conscious thoughts about the association of such mood with the experience of outer and inner constraints, pressure and meaningless repetitive actions. Such thoughts seemed to be accompanied by immediate feelings fitting these requirements, such as heteronomy, stress, and boredom. These feelings might have been facilitated by intently evoked memories of former working places fulfilling these conditions and the perceived strain to constantly focus attention on the task.

Imaginative Building Preparations

This phase was derived from a number of participants who referred in their experiential reports to mental imaginations that preceded the duck building. Further instances were found within the building phase in which participants used such as an inspirational as well as corrective tool for their activity.

Cognition, imaginations

Five participants reported visual imaginations of ducks as a perceived mean to facilitate, inform, or inspire the building task faced. See for example:

"P: I visualized, I think, eh. . . duck... ducklings actually. Eh... a few weeks ago Justine had some ducklings and that's were. . . those were

the first that came into the mind – my mind. And then some images of the rattle Donald Duck. And ehh. . . yeah. . . then I just tried to put together the bricks. . .” (P10, PF).

While some of such visualizations were of very concrete character, others seem to be more schematic. Thus, participant 10 reported shortly imagining the ducklings of a neighbor that he had seen the week before, and participants 5 and 10 both recalled in a moment thinking about the comic figure Donald Duck, hissing. On the other hand, participants 1 and 8 indicated that they were rather imagining different “duck positions,” without being able to precisely describe the visualizations connected, but rather calling them “schematic concepts.”

In general, reports of imaginative experiences seemed more common in the playful condition. The only case (out of five in total) that was reported for the non-playful condition was one of “schematic character” – possibly however, this might be at least connected to if not caused by the bias for copying the prototype in the non-playful condition, which naturally affords less creative planning.

Feeling/emotions

Interestingly, three out of the five participants reporting visual imaginations expressed some kind of emotional dissatisfaction quickly developing along with this experience. The reason for this seemed to lie in the circumstance that the translation of the mental images into a LEGO brick construction posed a strong, often unsolvable challenge. That is, participants mentioned that they were not able to live up to the pictures they drew in their head when dealing with the actual bricks.

The report of one of the participants furthermore indicated that this feeling might be more prominent in the non-playful condition, while in the playful condition participants might be able to somehow “let go” of the self-opposed matching task, allowing anything to happen:

“P: I tried to use these bricks [points with both index finger at third duck] in order to build something that would look like a duck flying. And. . . well, I encountered the same problem again (like in the non-playful condition, added by authors). I couldn’t get the image to fit. . . and sort of doing the working thing, I’ve really tried to, like... “okay this image needs. . . to fit. . .” and like, make a model. . . Then here it was more like experimental. . . Just going, could I do this, that or something else” (P1, PF).

To sum up, participants’ descriptions indicated that the different tasks triggered imaginative efforts clearly differing between conditions regarding content and experience. In the playful condition several participants indicated concrete associations to ducks seen in different contexts (of real life or illustrations seen). The translation of such imaginations into the building material was experienced as not very feasible, however this failure did not lead to frustration but rather allowed to open for the interaction with the material. In contrast, in the non-playful condition only one participant indicated an imagination preceding the building phase. However, this association to be of a schematic character, and the difficulty to translate such image into the building material was experienced as frustrating, leading to repetitive trials, rather than opening for creative production.

Building

With this phase we refer to any micro-experiences that accompanied the actual building of ducks or “duck-like” creatures.

We begin by a description of the building outcomes, before elaborating on the experiential reports of the participants:

In the playful condition, only three participants built the same duck again and again, while the other 19 participants built five ducks that did not look alike (though in four participants one of these was an exemplar of the prototype).

In contrast, in the non-playful condition, 10 out of the 22 participants built prototype ducks only, one participant always built the same duck different from the prototype, nine built one to three prototypes, while the rest of the ducks differed to a smaller or bigger degree, and only two participants built five different ducks. Three of the participants, who in the non-playful condition did not restrict themselves to copying, explicitly stated that this was due to a strategy change happening while building. Interestingly, they indicated that the experience was getting too easy to be still considered work or a job:

“P: I think that when I had built that one and it was so easy, then I was just like. . . mh. . . it’s. . . doesn’t really feel like I have. . . done like. . . a job, if I just do five of those. . . Cause it’s... [shrugs with shoulders]” (P7).

It is possible that these cases are caused by the identification of a non-playful stance with a work attitude. We will get back to this in the discussion.

It is also worth noting that in the non-playful condition, all participants built *ducks* – as instructed. By contrast, in the playful condition, four of the participants explicitly reported having constructed *something else* than ducks:

“P: I just started trying to. . . look what came out of it. I didn’t really think about it. . . but still. . . ehm. . . it’s a bit difficult to explain, I feel, but mh... I... because I did it mostly like: I didn’t think so much about it, but at the same time I was trying to be maybe a little, or maybe a bit creative.

I: How did you do that to be creative with that?

P: Mh. to. . . build something that looks like an animal. I was going for... making animals.

I: You were going for making animals. Not ducks actually. Just animals?

P: Yeah [. . .]

I: Did you think about a specific animal here or how did you. . .? [points at duck3]

P: A little bit. I don’t remember the... it’s not an “animal-animal,” but there is... so, there is this game. . . [. . .]where they eat ehm... [gestures eating] It’s “pacman” I think. Yeah. I thought a bit about that

I: Mh. Okay. Did you... was that eh... eh... a thought that you had before building it, or? During or?

P: Eh... after

I: Afterward?

P: Yeah

I: You saw it and then you thought about pacman?

P: [nods]” (P21, PF).

The quote indicates that participants might not initially have intended to break the rule (of building ducks). It rather indicates

that participants, when being playful, did not have a precise idea of what (else than ducks) to build in the first place and that the new products were a result of the playful stance which made them stay open to the process:

"P: Mh... yeah, it feels playful that there isn't like any rules. That can I, like, create whatever. . . I want to. . . and that... was also playful that you didn't, like [picks up duck1], necessarily, had to, like, see what it is [laughs]. What animal. . . So that I could like. . . ehm. . . change my mind about what it is... this was a man [picks up duck4] or. . . some other animal..." (P23, PF).

Such explorative mood seemed to be enhanced by the LEGO material that fostered a trial and error approach:

"*P: Well... It's LEGO. . . so it makes it bit hard to like really feel working with it. . . [. . .] it's... hm – it's maybe not a clear thought I had, but it pretty quickly once I got my hands on, it. . . was... well. . . somewhat natural just to fiddle with it. . .I played a lot with LEGO as a child. . .so it's sort of like. . . fiddle hand. . . and. . . yeah...[. . .] also maybe it's because it's such an easy task. . . I just like want to do it. . .rather then put too much thinking into it" (P1, NPF).*

The same participant also indicated that the LEGO material provided enforced this approach by making it difficult to fulfill a precisely set goal, such as an interior image of a duck. This might have helped to create an open space for the unexpected to happen:

"I think. . . I wanted to do a flying duck. . . but that one [clearing throat and points at duck 3] ended up being, like, a duck standing with its wings out. That one [points at duck 4] ended up being like a duck in the water. . .. with wings out. And that [points at duck 5] might be a flying duck – or something else. . . I am not sure. [P softly laughing]

I: And you thought about these things, while you were building this?

P: Yeah, so. . . oh. . . no, not while building – at some point there was just like. . . I needed one brick more or something and I couldn't really find it and then was just like. . . okay. . . And... then the model changed. . . Like, I sort of thought, ok, maybe it can become what I want it to be, but now it looks like this. – And. . .I am not sure if it was like. . . once the last brick was sitting or when I was done. . . but it sort of was like, okay that's what it is" (P 1, PF).

The following experiential descriptions extend on this finding.

Cognition, metathoughts/inner speech

We consistently found conscious linguistic thoughts guiding the building experience. In particular in the playful condition, participants seemed to use this tool to manage a delicate balance of creative pleasure and stress:

"P: I think I... ehm... I just didn't want it to. . . to make it a. . . a. . . more difficult task. I think it would re...quire... too much ehh. . . [I: mh] creativity from my site and I thought this was like the safe way and the... it wouldn't feel like... I think I easily would feel like... that I did it wrong, if I did anything else. [. . .]

"I kept telling myself that... "But it's not a competition!" to make myself feel more... calm about it.

[. . .] Maybe. . . playful and doing things a 100% correct... is not super... ehh..." (P6, PF).

Body, perception; Feelings/emotion

Further qualifying the experience of the building phase in the playful condition, 14 out of 22 participants reported feeling relaxed, free and having fun. One participant even replaced the word playful by "joyful" in the conversation:

"P: So. . . now when it has to be more joyful, so I don't have to think about it, but I already get the basics. Or I tried it in fact. . . So now it's. . . it's like I know this thing, but I can do whatever I want

[. . .] I: So, you said you could just do whatever you want. And when. . . you always say eh. . . "joyful" so is that an equivalent for you to playful?

P: Yeah, yeah!" (P11, PF).

On the other hand, 10 out of 22 participants reported negative feelings arising from the non-playful building, with boredom and stress being most prominent.

"P: This was just boring and... I... felt this pressure, you know, . . .I have to do this, because this is a working task and I find no ehh... happiness in this one. . . (I felt) a challenge also, because it's work so I should know how to build this... probably build the same for 8 h... 1000s of times every day. And eh... I lost most of the interest in [reaches for duck 3 and picks it up] how to... build this, because I. . . I knew it" (P2, NPF).

Furthermore, in the playful condition, two participants pointed toward an aesthetic quality of LEGO: its particularly pleasant tactual experience.

"I like in LEGO that they. . . [picks up prototype] the machine. . . the... this fabrication machine. . .it's so perfect you know. They always... fit [presses head down... into another and they... I enjoy eh... this feeling when it... you know? . . . sticks. . . it's so smooth and it has a little tension I... I like when they [claps hands together] go. . .together."(P2, PF).

"And so it's lots of positive associations with the touch and the feel of it and. . . the feeling when. . . [mimics pressing bricks on top of each other] the sound and the feeling of it connect.

It's like... 'cause there is a feeling, there is a sound. It's like, if you. . . if you close a book [gestures closing a book]. It can be a... a very subtle sound. . . It can be like "ahh". . . It's a really good feeling" (P22, PF).

In contrast, this pleasure was not found in the non-playful condition:

"I lost most of the interest in [reaches for duck 3 and picks it up] how to. . . build this, because I... I knew it. I didn't find it eeh... pleasant anymore to put together the bricks... I just pressed [demonstrates] on the top... I didn't feel, you know... anymore this. . . pleasure. . . of clicking them together" (P2, NPF).

The interview with participant four suggests that it is the experienced evaluation pressure evoked by the non-playful condition that hindered such an experience:

"P: I was maybe a bit more stressful, because I felt I should do it in the right way. So, I needed to [gestures repetition]... get over if it...if it wasn't right.

I: And how does it feel to be more stressful?

P: Mh. . . not nice [laughs] . . . yeah...it makes it more difficult to build, actually, when you more stressful, because you are not relaxing. . .so it was a bit di... more difficult to do it. Yeah.

I: How was that?

P: Mhm. . . Just really, if. . . you put the brick in the wrong place you had to move it and I think I could feel it in my fingers. They were less relaxed. . .So it was more difficult actually to put a brick in the right place.

I: Okay. . . in the way that you grasped them?

P: Yeah, or put it" (P4, NPF).

To sum up, participants' products and descriptions indicated that, also within the building action, the different tasks triggered micro-experiences clearly differing between conditions regarding content and experience.

The majority of participants in the playful condition built five different constructions, some of which did not even represent ducks. Rather than the outcome of a conscious strategy, this appeared at least partly to be the result of a distinct openness to the process induced by the autonomous stance taken. Participants' reports furthermore suggested that such openness might have been facilitated by a conscious care to keep up a good mood and a low stress level. This mood management appears in turn to have allowed for a higher sensibility toward the building material, by facilitating a perception of its aesthetic qualities, which again allowed for further exploration and openness toward the process.

In the non-playful condition on the other hand, the majority of participants produced several copies of one and the same construction. Furthermore 10 out of 22 participants reported negative feelings, such as stress and boredom, arising from such building. Interviews also indicated that such feelings might have reduced the sensibility for the material, by this further narrowing the action space available.

Product Evaluation

We identified this phase based on a number of participants reporting detailed inner reactions to their own finished products. While most participants built the ducks one at a time, the evaluation most often referred to was the one at the very end of the building face, looking at all their products together.

Body, kinesthesis/perception

After having finished the last duck, participants often took a moment to look at their products (we did see at least 12 doing so clearly in the video), partly even rearranging them (six participants), before telling the experimenter that they had finished the task. This behavior was particularly obvious in the playful condition (comprising all of the 12 clear cases), possibly influenced by participants building more diverse ducks (see above).

Feeling/emotion

Participants' reports of these moments were marked by descriptions of feeling/emotion, with joy and surprise being the most prominent. In fact, four participants stressed that they had not known that they would be able to produce such products:

"I: And when you saw it ready. . . how did you feel?

P: I felt ehm. . . satisfied [laughs]. . . yeah... I was actually a bit surprised, that I. . . I did sort of . . .got something like that" (P8, PF).

In the non-playful condition, on the other hand, one participant pointed out that not being surprised was essential for this condition:

"P: I was thinking about that I should do this with more than five ducks and just keep on going [And that it needed to be the same as this one [points at prototype duck] . . . it's like, I couldn't use my imagination and I just needed to produce... in the right way.

[. . .] I think because than I think it was... would be boring. . . just like, for a long time p... period...

[. . .] There wouldn't be any surprises along the way. It was just be the same again and again"(P4, NPF).

To sum up, interviews indicated that also participants' experiences with their finished products clearly differed between conditions regarding content and experience. In the playful condition, the majority of participants spent obvious time on looking at their products and reported positive emotional reactions to such (satisfaction and surprise), while in the playful condition, such reports were much more scarce and rather marked by negative expressions (boredom etc.).

In the following we will review these findings regarding the whole experience and reflect them in the light of current discussions in the field of gamification.

DISCUSSION

Our study aimed to assess three questions: (a) Can we find empirical support for the proposition to look at play and playfulness as a stance or state of mind that can be modulated by internal or external variables? (b) What are the experiential characteristics of the process of becoming playful in such way? (c) Can such investigations inform current discussions about the relation of play/playfulness and learning/performance?

To assess this, we used a controlled within-subject design, interviewing each participant on how they accomplished a building task when asked to perform it so that it either felt playful or not playful. In the following we discuss these questions in turn.

Playfulness as a State of Mind

The interviews indicated that participants in general could relate to the instructions and were able to enter a playful stance, with 19 out of 22 participants reporting that they managed to do so. It seemed slightly harder for them to achieve a non-playful state, with 12 participants reporting difficulties to do so. Close evaluation of the interviews suggested that difficulty to get in a non-playful state might be an effect of the LEGO material provided that carried too many associations to and memories of play. Some participants, albeit less pronounced, expressed difficulty in becoming playful; this appeared bound to the experimental situation, which restricted participants' feeling of autonomy. Taken together these findings support the claim that participants could assume a particular playful stance, generated by a voluntary internal modulation. They further show that external variables (e.g., experimental context and setup) must support this internal

effort for it to be realized completely. In particular, they must support a degree of autonomy which playfulness seems to require.

Experiential Qualities

This conclusion is supported by the most striking finding in the micro-phenomenological assessment: a feeling of autonomy seems to be constitutional for the ability to modulate playfulness. 14 participants reported immediately thinking that this task essentially meant being free to build what they wanted to – an indication of the importance of freedom of choice. Seven furthermore reported that they felt encouraged to create something "with their own minds" – an indication that the personal meaning of the task was essential to them feeling playful.

This contrasts starkly with the description of 16 participants in the non-playful condition who reported thinking about or immediately feeling constraints regarding their building actions and/or a deprivation of meaning regarding the task. One participant even explicitly referred to the quasi-sensation of a voice telling her to "do this, do this" which vividly exemplifies the lack of autonomy experienced.

Interviews furthermore indicated a range of strategies that participants considered as key in achieving and maintaining one or the other of the two stances explored: in particular, several participants recalled the need to actively use attention and memories to get into a non-playful stance, while the transition to be playful seemed to happen almost effortless, facilitated by spontaneous flashbacks into childhood or joyful situations of playing with children. There were indications though that this difference might be linked to the material used in the building task as we will discuss later.

Our data further revealed that the two conditions differed in what participants considered to be appropriate building products: 13 participants, when advised to be playful, recalled explicit thoughts about building five different ducks in the playful condition. This seemed to happen with the intention to enhance self-expression and creativity and was pursued as long as this task did not imply too much stress – a feeling apparently connected with work. In contrast to this, when requested to act non-playfully, 15 participants reported an inner advice to only copy ducks. The purpose of this seemed to be to create an atmosphere of restriction, meaninglessness, and boredom – and it was pursued as long as this task did not get too easy – a circumstance seemingly excluding an activity to be considered as work.

Participants' reports correlated strongly with participants' final products in the two conditions: the majority of participants in the playful condition built five different constructions, while in the non-playful condition, the majority produced several copies of one and the same construction. Notably, four participants in the playful condition in fact reported not building "ducks" at all, that is, their drive for freedom and creativity made them even ignore a critical part of the (very minimal) task instruction.

The interviews suggest that this higher expression of creativity in the playful condition may be the outcome of a dynamic process set in motion by taking an autonomous stance: freed from specific constraints and goals, participants seem to enter a curiosity driven interaction with the material, which allows for an unknown outcome to occur. This process might have been enhanced by an aesthetic way of perceiving the building material, enforcing an exploratory approach due to the sensory and reflective pleasure involved. Interestingly, this process may result in unexpected products, and the realization of this appears to enhance participants' feeling of competence.

These findings suggest that participants entered the playful condition by a contextual reinterpretation of the situation. This involved allowing oneself to feel autonomous and be exploratory without these self-imposed directives becoming constraints and stressful factors. In stark contrast, entering a non-playful condition was achieved by establishing a context of self-imposed constraints (e.g., time pressure or evaluation) and by reducing exploration, surprise and enjoyment to a minimum. This also appeared to reduce their experienced feelings of competence and motivational drive. Tellingly, if these constraints were not met, participants reported that their experience did not meet the requirements of the non-playful task.

The importance of autonomy as well as enjoyment for play and playfulness have been noted before, for instance in attempts to identify play with reference to specific features of the activity or the players (see section "Introduction"). However, these approaches do usually not not make any claims about the status and role of such features in the temporal course of modulating and being in a playful stance. To take one example, Bateson and Martin (2013) list that play behavior must be (a) spontaneous, intrinsically motivated and fun and (b) the players free from illness or stress (see also Fagen, 1981; Burghardt, 2005). They furthermore claim that "playful play is accompanied by a particular positive mood state in which the individual is more inclined to behave (and in the case of humans, think) in a spontaneous and flexible way." The psychologist Erikson offers a temporally more detailed model, however from a developmental perspective. He argues for a "play stage" in human development that is entered right after the "early childhood stage," which is aimed at achieving autonomy. Given the development of autonomy in the play stage, the child can learn to take initiative and engage in a world shared with others which is accompanied by a sense of mastery (see Erikson, 1959; as well as Proyer, 2018). We believe that our data offers a refined picture of how these mechanisms unfold in the distinct time course of each single playful experience. They show how a momentary modulation of autonomy influences consequent behavior and experience, how mood management and manipulation are involved in this process and how this affect the feeling of mastery/competence. These findings can be closely connected with current research about the relation between play and motivation and learning.

Playfulness and Learning

Our findings bear striking similarity to those described in one of the most influential theories of motivation: the psychologically oriented Self-Determination-Theory by Deci and Ryan (2000,

2015). Briefly, Deci and Ryan identified a spectrum of motivation from the autonomous to the controlled, with the two extremes being *intrinsic motivation* (most autonomous), when people find an activity naturally interesting and enjoyable, and *extrinsic motivation*, when people are not interested in the activity itself, but in a consequence of their engagement such as a reward (most controlled). In Deci and Ryan's analysis, establishing intrinsic motivation is preferred to extrinsic motivation in particular with respect to learning, as it is not dependent on a supporting framework. The framework furthermore claims that originally extrinsic motives may be internalized if they serve fulfillment of human's basic and innate psychological needs (Deci and Ryan, 1985; Ryan and Deci, 2000). Specifically, intrinsic motivation is increased by satisfaction of our natural need for *autonomy* (comprising freedom of choice and integration of values); our need for *competence* (the propensity to have an effect on the environment as well as to attain valued outcomes within it); and our need for *relatedness* (a natural sense of belonging to the environment and the people that are with us). In simple words: we won't enjoy – and therefore be naturally interested in – an activity, if we are coerced to do it, not convinced about its value, unable to master it, or if it deprives us of our feeling of belonging. On the other hand, the more autonomous, competent and related we feel when doing it, the stronger our intrinsic motivation can get.

Seen in this light, our findings suggest that creating the conditions to get playful means creating conditions to get intrinsically motivated. We found that participants intuitively used a contextual reinterpretation to modulate their degree of autonomy when asked to modulate their stance of playfulness. Further, they reported that the success of the modulation of playfulness depended on the success of the modulation of autonomy. In particular, when the outer context was perceived to oppose a feeling of autonomy, participants reported difficulties in becoming playful. Indeed, allowing for a playful stance crucially seems to depend on the creation of a sense of autonomy. This suggests that situations lacking autonomy will not be experienced as playful, and that designing the specifics of the context (the organizational environment etc.) may be crucial for allowing such autonomy.

Our findings further present a possible mechanism for how the experience of autonomy may evoke a feeling of competence: the diversity of ducks built in the playful condition was experienced as a result of the openness of the process, created by the autonomous position. Many participants reported a dynamic interaction with the building material, at times experienced as aesthetic, which allowed for an exploration of its possibilities. Many explicitly mentioned to be surprised by the results of their activity, stressing that they had not been aware of their own capacity to build so creatively. We suggest that this represents a concrete experience of competence, thus providing the second key component of intrinsic motivation.

Playfulness Versus Gamification

Finally, our findings may throw new light on key problems faced within the field of gamification. Briefly, gamification is the overarching term for the approach to introduce "*game design elements within non-game contexts*" with the explicit aim to achieve levels of motivation "*as high as for playing video games*" for a task that is considered difficult to initiate (Deterding et al., 2011, p. 1). However, a number of studies have indicated that the approach may face some fundamental issues. Though the majority of research exploring the effect of gamification on motivation have found more positive than negative or null effects (Hamari et al., 2014; Seaborn and Fels, 2015), it has been claimed that the main kind of motivation established in gamification is extrinsic, rather than intrinsic, and that the effects may thus not transfer outside the specific context (Nicholson, 2015). It has even been warned that designs that are more extrinsically motivating might risk to replace intrinsic drives on the long run, thus creating a constant dependence on reward structures and other forms of extrinsic evaluation (Deci, 1971, 1972; Deci et al., 1999; Kohn, 1999). Thus, Hanus and Fox (2015) showed that gamification in a classroom led to lower performance compared to a group attending non-gamified tuition and that this result was mediated by the lower level of intrinsic motivation of the gamified tuition group. Most interestingly for us, it has been suggested that the main cause of this circumstance may be that gamification, by focusing on game design elements, focuses on creating a with play associated activity rather than a "playful stance" *per se* (Deterding, 2010; Nicholson, 2015). However, only the latter might be effective in enhancing intrinsic motivation. Indeed, Deterding (2016) showed that some aspects of games, such as playing in a team, might lead to the feeling of obligation, a loss of playfulness and therefore reduced motivation to play (Deterding, 2016).

Our results support and extend on this idea: they provide strong support for the hypothesis that allowing for a stance of playfulness may be an effective way to increase intrinsic motivation. Furthermore, they generate specific hypotheses about what processes one should pay attention to, if intending to design playful experiences.

Experiencing autonomy, that is, the feeling of freedom and meaningfulness of own actions, seems key to adopting a playful stance. We found that our participants had internal means to modulate this feeling – but there may be constraints to this, due to differences of context as well as personal capacity.

There appears to be a thin line between empowering self-determination and the experience of stress due to self-imposed expectations. It is possible, that different capacities to adapt own expectations accordingly affect the capacity of being playful and that training focusing on sustainable self-management might thus support playfulness.

The situational context seems critical for achieving the experience of autonomy. In particular, participants mentioned that the experimental setting in itself was a restriction. This may also apply in educational as well as professional environments. Our findings do not suggest a solution for this problem, though

they indicate that some individuals could overcome this in time. There is scope to research this further.

Our findings highlight the importance of the properties of the physical material involved in the experience – in our case exemplified by the LEGO bricks provided. It seems that a material might differentially foster explorative behavior due to (a) its abstractness, that is its resistance to be guided by mental imaginations and (b) its sensual, aesthetic properties and design. Our findings suggest that the abstract, modular qualities of the material provided made a direct translation from a mental image into the model difficult and that this increased the explorative potential. This was further supported by sensual pleasures that may enhance further exploration. However, due to the small sample size, these hypotheses should be followed up by research.

Lastly our findings indicate that autonomy and competence (as two critical components of intrinsic motivation) may in some instances stand in a causal relationship. Thus, a feeling of competence can be evoked by unforeseen products of the explorative process facilitated by an autonomic, playful stance. This proposes a putative looping effect of autonomy and surprise, which may be critical in supporting intrinsic motivation during a playful stance. Accordingly, designers of playful processes may first and foremost focus on establishing an autonomy component, while the feeling of competence, also constitutional for intrinsic motivation, may be elicited in and by the process itself. Thus, if one designs for learning through play, one may want to pay particular attention to the engagement of participants and evaluate their experience of playfulness rather than focus on their *a priori* competences and on the game-like qualities of the situation.

Limitations of Study and Directions for Future Research

We hope the results discussed above have demonstrated the usefulness of our approach for researchers interested in playfulness as a stance or state of mind. However, for a number of reasons, the study should be considered a pilot, and further empirical work may be required.

Firstly, our population sample comprises quite a homogenous group of students, all taking part of a schooling program (the Danish Hojskole) that in its creative curriculum might attract particularly playful participants. Unfortunately, we did not assess playfulness as a character trait specifically, e.g., using dedicated questionnaires (Proyer, 2012). Future research should explore if individuals, independent of their personal tendency to be playful, show behavior and experiential reports similar to the ones observed here.

Secondly, we may have primed some of the participants, by describing the non-playful condition as referring to a stance similar to "work." The work-play contrast is indeed one that has been suggested by existing literature on play and might have triggered certain cultural connotations impinging on the otherwise data-driven approach (see for example Bateson and Martin, 2013). This term should thus be avoided in future replications of this work.

Thirdly, interviews revealed that the building material chosen for the experiment may in itself have primed participant to be playful. This could be for cultural as well as material reasons: LEGO bricks are deeply embedded in Danish culture and may obviously trigger childhood related memories in participants. As explicated above, it might also enhance explorative behavior due to its abstractness and sensual aesthetics. Future research should thus further explore the influence of the material chosen by comparing the results of this study with one using a material more neutral (that is not intuitively associated with either play or work) and possibly also with one clearly associated with work.

Fourthly, the data-driven analysis did not provide results that could immediately be ascribed to "relatedness," the third component of intrinsic motivation according to Deci and Ryan's theory. In this framework, "relatedness" is understood as a natural sense of belonging to the environment and the people around. One reason that this aspect of intrinsic motivation may not have shown up in the data, that otherwise strongly reminded of Deci and Ryan's theory, is that the setup did not include obvious others – like building partners etc. It has been suggested that superiors like supervisors or teachers are important others too, with students being dependent on them to like, respect and value their work, and to develop intrinsic motivation for it. In this sense the experimenter might be interpreted as a distinct other. However, our data only gave indirect evidence of this, mainly with reference to an evaluative instance in the non-playful condition. Future research should engage in further exploring this aspect by modulating the setup to include partners in the task (see also Tylén et al., 2016).

Summary

To the best of our knowledge, this is the first empirical and data-driven study assessing people's experience of becoming playful. To our estimation our data shows that our experimental design and the chosen methods allow a deep and detail-rich insight into participants' capacity of voluntary modulating their own playful stance.

In particular, interview results indicate that participants are able to voluntarily modulate their playfulness to the degree that they are able to modulate their autonomy in the building process and trust the process elicited by that: higher autonomy then facilitates a dynamic interaction with the material given, that can be further enhanced by the properties of that material. The surprising products of that activity make people aware of their own creative competence, which positively affects their mood and motivates them to continue the process. It seems that only experiences that fulfill these looping processes of autonomy, surprise, feeling of competence and motivation are categorized as playful in hindsight.

We thus propose a new working definition. Playfulness may be conceptualized as an attitude of throwing off constraints, which facilitates an explorative interaction with materials and others. This allows for intrinsic motivation to arise, supported by the surprising results of that interaction and by the connected positive emotions and feeling of competence.

When designing for playfulness, internal and external variables that modulate these processes should be taken into account.

We hope that future research will be able to make further usage of these findings and propositions.

AUTHOR CONTRIBUTIONS

KH designed the research, conducted and analyzed the data and wrote the manuscript. AR supervised all stages and reviewed the manuscript.

FUNDING

The research was undertaken as part of the PLAYTrack project at the Interacting Minds Centre, Aarhus University, supported by a research grant from the LEGO Foundation.

ACKNOWLEDGMENTS

We thank Rikke Grinderslev Rasmussen for her most valuable help in transcribing and coding the data.

REFERENCES

Apter, M. J. (1991). "A structural-phenomenology of play," in *Adult Play: A Reversal Theory Approach*, eds J. H. Kerr and M. J. Apter (Amsterdam: Swets & Zeitlinger), 13–29.

Arrasvuori, J., Korhonen, H., and Boberg, M. (2010). "Understanding playfulness: an overview of the revised playful experience (PLEX) framework," in *Proceedings of the Conference on Designing Pleasurable Products and Interfaces*, (New York, NY: ACM).

Barnett, L. A. (2017). The inculcation of adult playfulness: from west to east. *Int. J. Play* 6, 255–271. doi: 10.1080/21594937.2017.1383010

Bateson, P., and Martin, P. (2013). *Play, Playfulness, Creativity and Innovation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139057691

Bitbol, M., and Petitmengin, C. (2013). A defense of Introspection from within. *Constr. Found.* 8, 269–279.

Blurton Jones, N. G. (1972). "Categories of child-child interaction," in *Ethological Studies of Child Behaviour*, ed. D. Morris (Cambridge: Cambridge University Press), 3–39.

Boberg, M., Karapanos, E., Holopainen, J., and Lucera, A. (2015). *PLEXQ: Towards a Playful Experiences Questionnaire. Conference paper CHI PLAY 2015*. doi: 10.1145/2793107.2793124

Bojner Horwitz, E., Stenfors, C., and Osika, W. (2013). Contemplative inquiry in movement: managing writer's block in academic writing. *Int. J. Transpers. Stud.* 32, 16–26. doi: 10.24972/ijts.2013.32.1.16

Burghardt, G. M. (2005). *The Genesis of Animal Play: Testing the Limits*. Cambridge, MA: MIT Press.

Caillois, R. (1961). *Man, Play and Games*. New York, NY: Free Press.

Créno, L., and Cahour, B. (2015). Triangulation of methods for an ecological analysis of the experience of email management in overworked executives. *French Psych.* 2, 129–144.

De Kowen, B. L. (2017). *Defining Playfulness or not. Psychology Today*. Available at: https://www.psychologytoday.com/us/blog/having-fun/201708/defining-playfulness

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *J. Pers. Soc. Psychol.* 18, 105–115. doi: 10.1037/h0030644

Deci, E. L. (1972). Intrinsic motivation, extrinsic reinforcement, and inequity. *J. Pers. Soc. Psychol.* 22, 113–120. doi: 10.1037/h0032355

Deci, E. L., Koestner, R., and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychol. Bull.* 125, 627–668. doi: 10.1037/0033-2909.125.6.627

Deci, E. L., and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behaviour*. New York, NY: Plenum. doi: 10.1007/978-1-4899-2271-7

Deci, E. L., and Ryan, R. M. (2000). The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. *Psychol. Inq.* 11, 227–268. doi: 10.1080/08870440902783628

Deci, E. L., and Ryan, R. M. (2015). "Self-determination theory," in *International Encyclopedia of the Social & Behavioral Sciences*, 2nd Edn, Vol. 21, eds N. Smelser and P. Baltes (New York, NY: Elsevier), 486–491. doi: 10.1016/B978-0-08-097086-8.26036-4

Depraz, N., Gyemant, M., and Desmidt, T. (2017). A first-person analysis using third-person data as a generative method: a case study of surprise in depression. *Constr. Found.* 12, 190–203.

Deterding, S. (2010). *Just Add Points? What UX can (and cannot) Learn from Games. Slide Show presented at UX Camp*. Available at: https://www.slideshare.net/dings/just-add-points-what-ux-can-and-cannot-learn-from-games/ [accessed February 13, 2017]

Deterding, S. (2016). "Contextual autonomy support in video game play: a grounded theory," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, (San Jose, CA: ACM), 3931–3943. doi: 10.1145/2858036.2858395

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). *Gamification: Toward a Definition. Conference: CHI 2011 Gamification Workshop Proceedings*. Available at: http://gamification-research.org/wp-content/uploads/2011/04/02-Deterding-Khaled-Nacke-Dixon.pdf [accessed December 2, 2018]

Erikson, P. (1959). *Identity and the Life Cycle*. New York, NY: International Universities Press.

Fagen, R. (1981). *Animal Play Behavior*. New York, NY: Oxford University Press.

Glynn, M. A., and Webster, J. (1992). The adult playfulness scale: an initial assessment. *Psychol. Rep.* 71, 83–103. doi: 10.3389/fpsyg.2018.00421

Gray, P. (2013). *Free to Learn: Why Unleashing the Instinct to Play will Make our Children Happier, More Self-reliant, and Better Students for Life*. New York, NY: Basic Books.

Hamari, J., Koivisto, J., and Sarsa, H. (2014). Does gamification work? - a literature review of empirical studies on gamification. *Paper presented at the 47th Hawaii International Conference on System Sciences*, Waikoloa. doi: 10.1109/HICSS.2014.377

Hanus, M. D., and Fox, J. (2015). Assessing the effects of gamification in the classroom: a longitudinal study on intrinsic motivation, social comparison, satisfaction, effort and academic performance. *Comput. Educ.* 80, 152–161. doi: 10.1016/j.compedu.2014.08.019

Huizinga, J. (1955). *Homo Ludens; A Study of the Play-Element in Culture*. Boston: Beacon Press.

Jack, A. I., and Roepstorff, A. (2003). Why trust the subject? *J. Conscious. Stud.* 10, 5–20.

Johansson, P., Hall, L., Sikström, S., and Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science* 310, 116–119. doi: 10.1126/science.1111709

Kohn, A. (1999). *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes*. Boston: Houghton Mifflin.

Lieberman, J. N. (1965). Playfulness and divergent thinking: an investigation of their relationship at the kindergarten level. *J. Gen. Psychcol.* 107, 219–224. doi: 10.1080/00221325.1965.10533661

Lieberman, J. N. (1966). Playfulness: an attempt to conceptualize a quality of play and of the player. *Psychcol. Rep.* 19:1278. doi: 10.2466/pr0.1966.19.3f.1278

Millar, S. (1968). *The Psychology of Play*. London: Penguin Books, Ltd.

Murray, H. A. (1938). *Explorations in Personality*. Oxford: Oxford University Press.

Nicholson, S. (2015). "A RECIPE for meaningful gamification," in *Gamification in Education and Business*, eds T. Reiners and L. Wood (Cham: Springer), 1–20.

Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychcol. Rev.* 84, 231–259. doi: 10.1037/0033-295X.84.3.231

Petitmengin, C. (2001). *L'expérience Intuitive*. Paris: L'Harmattan.

Petitmengin, C. (2006). Describing one's subjective experience in the second person: an interview method for a science of consciousness. *Phenomenol. Cogn. Sci.* 5, 229–269. doi: 10.1007/s11097-006-9022-2

Petitmengin, C., and Bitbol, M. (2009). The validity of first-person descriptions as authenticity and coherence. *J. Conscious. Stud.* 16, 363–404.

Petitmengin, C. (ed.) (2009). *Ten Years of Viewing from Within: The Legacy of Francisco Varela*. Exeter: Imprint Acadamic.

Petitmengin, C., Remillieux, A., Cahour, B., and Carter-Thomas, S. (2013). A gap in Nisbett and Wilson's findings? A first-person access to our cognitive processes. *Conscious. Cogn.* 22, 654–669. doi: 10.1016/j.concog.2013.02.004

Petitmengin-Peugeot, C. (1999). "The intuitive experience," in *The View from Within*, eds F. Varela and J. Shear (Exeter: Imprint Academic), 43–77.

Proyer, R. T. (2012). Examining playfulness in adults: testing its correlates with personality, positive psychological functioning, goal aspirations, and multi-methodically assessed ingenuity. *Psychcol. Test Assess. Model.* 54, 103–127.

Proyer, R. T. (2018). Playfulness and humor in psychology: an overview and update. *Humor* 31, 259–271. doi: 10.1515/humor-2016-0080

Reynolds, P. (1976). "Play, language, and human evolution," in *Play: Its role in Development and Evolution*, eds J. Bruner, A. Jolly, and K. Sylva (New York: Basic Books), 621–637.

Rubin, K. H., Fein, G. G., and Vandenberg, B. (1983). "Play," in *Handbook of Child Psychology, Vol. 3: Social Development*, ed. E. M. Hetherington (New York, NY: Wiley), 694–774.

Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 68–78. doi: 10.1037/0003-066X.55.1.68

Sailer, M., Hense, J. U., Mayr, S. K., and Mandl, H. (2017). How gamification motivates: an experimental study of the effects of specific game design elements on psychological need satisfaction. *Comput. Hum. Behav.* 69, 371–380. doi: 10.1016/j.chb.2016.12.033

Sanderson, R. C. (2010). *Towards a New Measure of Playfulness: The Capacity to Fully and Freely Engage in Play*. Ph.D. Dissertations, Loyola University Chicago, Chicago, IL.

Seaborn, K., and Fels, D. I. (2015). Gamification in theory and action: a survey. *Int. J. Hum. Comp. Stud.* 74, 14–31. doi: 10.1016/j.ijhcs.2014.09.006

Shen, X. S., Chick, G., and Zinn, H. (2014). Playfulness in adulthood as a personality trait – a reconceptualization and a new measurement. *J. Leis. Res.* 46, 58–83. doi: 10.1080/00222216.2014.11950313

Sutton-Smith, B. (2009). *The Ambiguity of Play*. Cambridge: Harvard University Press.

Sutton-Smith, B., and Kelly-Byrne, D. (1984). "The idealization of play," in *Play in Animals and Humans*, ed. P. K. Smith (Oxford: Basil Blackwell), 305–321.

Tylén, K., Bjørndahl, J. S., Roepstorff, A., and Fusaroli, R. (2016). Trails of meaning construction: symbolic artifacts engage the social brain. *Neuroimage* 134, 105–112. doi: 10.1016/j.neuroimage.2016.03.056

Vermersch, P. (1994). *L'entretien D'explicitation en Formation Initiale et en Formation Continue*. Paris: ESF.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Heimann and Roepstorff. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 29 March 2018
doi: 10.3389/fpsyg.2018.00421

An Initial Cross-Cultural Comparison of Adult Playfulness in Mainland China and German-Speaking Countries

Dandan Pang[1] and René T. Proyer[2]*

[1] Personality and Assessment, Department of Psychology, University of Zurich, Zurich, Switzerland, [2] Personality and Assessment, Department of Psychology, Martin Luther University Halle-Wittenberg, Halle, Germany

Compared with playfulness in infants and children, playfulness in adults is relatively under-studied. Although there is no empirical research comparing differences in adult playfulness across cultures, one might expect variations between Western and Eastern societies such as China. While playfulness is typically seen as a positive trait in Western culture, there are hints in Chinese culture that being playful has negative connotations (e.g., associations with laziness and seeing play as the opposite of work). The aim of this study was to compare expressions of playfulness in one sample from German-speaking countries (n = 143) and two samples from China (Guangzhou: n = 176; Beijing: n = 100). Participants completed one playfulness scale developed in the West (Short Measure of Adult Playfulness, SMAP) and one from the East (Adult Playfulness Questionnaire, APQ). Additional ratings of the participants were collected to measure: (a) the level of playful behavior expressed by people in different situations (e.g., when being around family members, in public, or on social media), and (b) individuals' perceptions of society's expectations concerning the appropriateness of being playful in the given situations. Overall, the results of the comparisons were mixed. Although SMAP scores did not vary significantly across the three samples, people from German-speaking countries tended to score higher on some facets of the APQ and some situational ratings. Stronger effects were found when comparing only the German-speaking sample and the Guangzhou sample. In addition to the cross-cultural differences that we expected, we also detected Chinese regional variations (North vs. South). We conclude that societal rules and cultural factors may impact expressions of playfulness in a society.

Keywords: adult playfulness, cross-culture, situation-specific playfulness, positive traits, China

OPEN ACCESS

Edited by:
Anat Bardi,
Royal Holloway, University of London, United Kingdom

Reviewed by:
Jen-Ho Chang,
Academia Sinica, Taiwan
Peter Bevington Smith,
University of Sussex, United Kingdom

****Correspondence:***
Dandan Pang
d.pang@psychologie.uzh.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *15 October 2017*
Accepted: *13 March 2018*
Published: *29 March 2018*

Citation:
Pang D and Proyer RT (2018) An Initial Cross-Cultural Comparison of Adult Playfulness in Mainland China and German-Speaking Countries. Front. Psychol. 9:421. doi: 10.3389/fpsyg.2018.00421

INTRODUCTION

Theoretical Background and Current Studies in Western Culture

Play, as a component of human behavior, is an innate part of our nature, and a basic *need to play* has been described as a core human characteristic that can take many forms, defined for instance as "to relax, amuse oneself, seek diversion and entertainment; to 'have fun,' to play games; to laugh, joke and be merry; to avoid serious tension" (Murray, 1938; p. 83). Developmental

psychology has acknowledged the importance of play for the acquisition of different abilities and developmental transitions (Erikson, 1950; Piaget, 1951). Accordingly, infants and children have an intrinsic understanding of the importance of play (Yu et al., 2007). Previous studies suggest that play contributes to physical, cognitive, social, linguistic and emotional aspects of child development (Csikszentmihalyi, 1975; Lieberman, 1977; Isenberg and Quisenberry, 1988; Barnett, 1990; Blasi et al., 2002; Tamis-LeMonda et al., 2004; Ginsburg, 2007). For example, in cognitive development, play and games can assist children with creative thinking and behavioral flexibility (Piaget, 1951; Sutton-Smith, 1967), as well as widen their memory of factual knowledge (Lunzer, 1959). It has been argued that when we play, we are engaged in the purest expression of our humanity (Brown and Vaughan, 2009). Of course, play is not only limited to children. It can also be found in adults; even in comparatively serious situations (Bologh, 1976) such as when people are at work (Csikszentmihalyi, 1975; Csikszentmihalyi and LeFevre, 1989). For the present study, not only the actual behavior (*play*), but *playfulness* as a personality trait is of importance. Lieberman (1977) argues that "[. . .] playfulness as a quality of play would developmentally transform itself into a personality trait of the player in adolescence and adulthood" (Lieberman, 1977; p. 23).

Playfulness in adults is comparatively a rarely studied trait (Proyer, 2012a). It can be defined as: "[. . .] an individual differences variable that allows people to frame or reframe everyday situations in a way such that they experience them as entertaining, and/or intellectually stimulating, and/or personally interesting. Those on the high end of this dimension seek and establish situations in which they can interact playfully with others (e.g., playful teasing, shared play activities) and they are capable of using their playfulness even under difficult situations to resolve tension (e.g., in social interactions, or in work-type settings). Playfulness is also associated with a preference for complexity rather than simplicity and a preference for—and liking of—unusual activities, objects and topics, or individuals" (Proyer, 2017; p. 114). Previous research has shown that adult playfulness is associated with a large number of positive outcomes such as academic performance (Proyer, 2011); facilitation of positive emotions (Fredrickson, 2001); relationship satisfaction (Aune and Wong, 2002; Proyer, 2014b; Proyer et al., in press a); sexual selection (Chick, 2001; Chick et al., 2012; Proyer and Wagner, 2015); coping with stress (Qian and Yarnal, 2011; Magnuson and Barnett, 2013; Proyer, 2014a); and well-being (Barnett, 2012; Proyer, 2012c, 2013, 2014a)—to name but a few.

Cross-Cultural Aspects and Current Studies in Eastern Culture

To the best of our knowledge, play and playfulness in adults are mostly studied from a Western perspective. For example, studies have recently been conducted with samples from the United States (Barnett, 2007), United Kingdom (Aroean, 2012), Denmark (Hasse, 2008), Switzerland (Proyer, 2011), and Germany (Proyer and Wagner, 2015), while Eastern countries have only rarely been studied (e.g., Yu et al., 2003, 2007; Yue et al., 2016). Studies conducted in German-speaking countries have led to the development of a multifaceted model of playfulness; i.e., other-directed, lighthearted, intellectual, and whimsical (Proyer, 2017). A discussion of structural issues (Proyer, 2012a; Proyer and Jehle, 2013) has provided support for its importance in academic settings (Proyer, 2011) and romantic relationships (Proyer, 2014a,b; Proyer et al., in press a), and its association with virtuousness and positive psychological functioning (Proyer and Ruch, 2011; Proyer, 2013). An analysis of German-speaking laypersons' perceptions of how they use playfulness in their daily lives revealed seven main categories: well-being; humor and laughter; mastery orientation; creativity; relationships; coping strategies; and coping with specific situations (Proyer, 2014a). Overall, these findings provide support for the notion that people studied in German-speaking countries assign important functions to playfulness and that it is related to important outcome variables such as relationship satisfaction and academic success. Comparatively less knowledge exists about the role of playfulness in Eastern culture. In an effort to narrow this gap in the literature, the aim of this study was to compare measures developed in the West and the East, collect data from both Western culture (German-speaking countries) and Eastern culture (China), and see whether the findings differ.

This comparison is of particular interest since German-speaking countries are typically rated higher in *individualism* than China. On a 1 to 10-point scale, the country scores on Individualism-Collectivism are 7.90 for Switzerland, 7.35 for West Germany, 6.75 for Austria, and 2.00 for China (Suh et al., 1998). Hence German-speaking countries (Switzerland, Austria, and Germany) and China enable a cultural comparison along the Individualism-Collectivism dimension. People in individualistic countries display less conformity behavior (Hofstede, 2001; p. 236). One might argue that people in individualistic cultures utilize a larger variety of playfulness functions in different areas of life than those in collectivistic countries. For a better understanding of potential cultural differences, and given the absence of previous data, we discuss the Eastern perspective (Chinese, to be precise) on play and playfulness in more detail.

A common stereotype about the Chinese is that they are diligent (Smith, 1894). In one of the first chapters of his book "Chinese Characteristics" Smith (1894) concludes: "[. . .] there can be little doubt that casual travelers, and residents of the longest standings, will agree in a profound conviction of the diligence of Chinese" (p. 27). Smith also pointed out that this diligence is not characteristic of a single group within Chinese society, but rather that it can be applied to all residents of the country (Smith, 1894). Even nowadays, with a growing influence of globalization, it is still highly valued to be diligent in China. Aphorisms such as "玩物丧志" ("Excessive attention to plaything saps the will"), "勤有功, 嬉无益" ("Reward lies ahead of diligence, but nothing is gained by play"), and "业精于勤荒于嬉" ("Achievements are reached by hard work rather than play"), are taught to children when they start primary school. Overall, it seems as if many Chinese tend to have a negative bias toward play. One common belief is that play is the opposite of work [see Glynn and Webster (1992) for a Western representation of this idea] and is only reserved for children. Only by working hard, can happiness and success be achieved (Harrell, 1985).

Of course, there are also other variables in addition to the individualism vs. collectivism dimension that may contribute to cultural differences; for example, the autonomy vs. embeddedness dimension in Schwartz's (2006) theory of value orientation (for an overview see Sagiv et al., 2017). Playfulness shares characteristics with both intellectual and affective autonomy as it relates to intellectual striving and its pursuit enables positive experiences (e.g., Proyer, 2017). However, an emphasis on embeddedness does not seem to foster playful behaviors.

A dominant perception in Chinese culture seems to be that the intense competition of the education system requires their students to study hard without being distracted by play activities. There are only about twenty top-tier universities in mainland China and there are millions of students every year, all having only one chance in that year to get accepted via the national university entrance examination (Davey et al., 2007). It is also seen as one of the few chances for students from the rural areas in China to change their social class in a comparatively fast and low-cost way (Chen and Uttal, 1988). This competition forces children to study hard from the very first school day, so that they can get accepted into a better secondary school and eventually a better college. A big difference in comparison with competitive educational systems in the West (e.g., in the United States) seems to be that this national examination is the only criterion for Chinese high school students, whereas the United States system is characterized by a variety of criteria (i.e., in addition to achieving good grades, students are also encouraged to attend extracurricular activities). The idealized image of the hard-working student is culturally well-represented by paragons from earlier times. For example, a story tells us that Sun Jing (1425–1484), a student in Sichuan province, tied his hair to a house beam so that he could keep on learning and did not fall asleep despite his long working hours (Lin, 2012).

This relatively negative perspective on play and playfulness seems to have had an impact on the language, which has led to a basic problem for the present study. A term in Chinese that corresponds precisely to playfulness seems to be missing (see also Yu et al., 2007). It should be mentioned that the term "play" is avoided in the Chinese language in many cases. For example, instead of saying "playing football," one says "踢足球 (kicking football)," while "playing the piano" is "演奏钢琴 (performing on the piano)." Consequently, at the early stages of our study we asked Chinese students who study in Switzerland (and should have some understanding of the Western concept of play and playfulness) about their understanding and suggestions for translating the term "playfulness." Twenty-two students (13 female, 9 male) were asked: "How would you translate the sentence 'I am a playful person,' especially the word 'playful'?" The answers were diverse. Some of them referred to people as "not reliable," "playboy," "not nice," or that it should be expressed as the "opposite of study" and so forth. It was mainly the students who had been abroad only for 1 year or less who expressed these associations. Those who had been abroad for more than 5 years had different opinions. They would link playful to adjectives like "humorous," "witty," or "interesting." This may point to some cultural transmission in how playfulness is being perceived and in associations related to this individual differences variable (see also Barnett, 2017).

As mentioned above, China is a collectivistic country (Hofstede, 2001) with strong social hierarchies (Triandis et al., 1990; Markus and Kitayama, 1991). In Western culture, an individual's dominant behavior is positively reinforced and people are encouraged to climb the hierarchy (Triandis and Gelfand, 1998). In contrast, a collectivistic society prefers subordination (Triandis and Gelfand, 1998) and praises agreeable individuals rather than dominant ones (Moskowitz et al., 1994; Realo et al., 1997). In this sense, play could be considered as not obeying certain rules and to being self-centered, which is not approved by collectivistic cultures and may even lead to anxiety and insecurity for those in power. As Confucius himself once remarked: "each should behave appropriately according to his or her station" and "man has to be serious to be respected" (cited after Liao, 2007).

It should be mentioned that there is an important differentiation between the public and the private self when discussing play and playfulness in Chinese culture and tradition. Confucius himself allows for *proper* playfulness, which refers to a form of private, moderate, good-natured, tasteful and didactically useful mirth (cited after Milner Davis and Chey, 2011). This sense of propriety can also be found in a famous quote by Pu Songling (1640–1715), a writer of the Qing dynasty, who notes: "There is no one who does not laugh, but one must laugh at an appropriate time" (人罔不笑, 但须有时; Liaozhai zhi yi, p. 155). Additionally, Daoism, as an alternative view of life, has a tradition of the appropriate use of playfulness. Two of the main pieces of Daoist literature, Liezi and Zhuangzi, are both made up of legends, jokes, parables and allegorical tales, all laced with playfulness and paradoxes. Daoists such as Zhuangzhou criticized Confucian social conventions by being a "huaji-ist," "huaji" being an earlier indigenous term for humor. In addition, playfulness in China can also be found in many forms, both literary and conventional. For instance, Dayoushi (打油诗), a Chinese literary game between friends where each player picks up a thought or expression from the last player and twists the meaning in an unexpected and, therefore, funny way, is one source of evidence (cited after Milner Davis and Chey, 2011). In the Chinese Spring Festival Gala, a wide variety of puns are found in the cross talk, since the Chinese language is rich in homophones. Western influences on humor seem to be comparatively limited. However, selected works (e.g., by Henri Bergson; see Milner Davis, 2014) were translated into Chinese and comparatively well-received in academic circles.

To summarize, although there are ambivalent attitudes toward playfulness, the negative perception of play and playfulness still seems to be present in China. Thus we expected that the Chinese participants in our study would be less playful than the German-speaking participants. Likewise, we expected that the differences in situations with hierarchical communication in various forms, such as in a public situation or at the workplace, would be larger. To assess the participants' ratings of their level of playfulness in these different contexts, we developed a list of 14 different situations in daily life for this study: the *Brief Rating List of Playfulness in Different Situations* (BRLPS).

Additionally, participants provided ratings on the perceived *societal appropriateness* of being playful in the given situations. This will allow for a comparison of the two perspectives.

It has already been mentioned that there is a paucity of research on playfulness in Eastern countries (cf. Yue et al., 2016). However, a few studies exist that should be highlighted. The first translation of the word "playfulness" in Chinese emerged in Taiwan. Researchers used the word "玩兴 (wanxing)," which means "being in the mood to play" or "having an interest in playing". Yu and her colleagues (Yu et al., 2003) discuss the influence of traditional Chinese values on people's attitudes toward playfulness, such as "play only belongs to children" and "adults should work hard and be serious." However, they also noted that because of globalization and the impact of a post-materialist value system, playfulness is becoming more and more important among young Taiwanese (Yu et al., 2003). Those who have fun at work experience high spontaneity, concentration, relaxation and happiness, which contributes to creativity, team feelings and better work performance (Yu, 2004). Their definition of playfulness is: "... a personal characteristic of pleasantry temperament, combining physical, cognitive and social spontaneity, which shows the power to begin energetically or to concentrate on events or activities, and the ability to utilize resources in solving problems or in rising to the challenge of own competence" (Yu et al., 2007; p. 416). Based on this definition, Yu and her colleagues developed an *Adult Playfulness Questionnaire* (APQ, Yu et al., 2003) within the context of Eastern culture. In total, 755 Taiwanese adults from different occupations were consulted, and the items were derived from a literature review, group discussion, open questionnaires, and in-depth interviews. The results showed acceptable reliability and validity, and factor analysis yielded a six-factor model (Yu et al., 2003). Later, the authors favored a reduced three-factor model of adult playfulness; namely, "pleasantry," "initiative and concentration," and "creativity" (Yu et al., 2007). It is important to note that the term "pleasantry" is being used in a different sense here compared to the common understanding. Yu et al. (2007) argue that it is a combination of a sense of humor and a childlike manner. We kept the original translation by Yu and colleagues because we wanted to keep the terminology of the original authors.

In a review article, Li (2006) noted that playfulness contributes positively to the creativity of college students. Zhang (2011, Unpublished) developed a measure of playfulness for college students that consists of a seven-factor structure: namely, sense of humor, creativity, curiosity, activity, sociality, spontaneity, and pleasure. Differences in playfulness were found for gender (males scored higher than females in creativity, whereas females were higher in spontaneity, sociality, and pleasure); grades (e.g., first-years showed the highest level of playfulness); majors (e.g., literature and history students were higher than science and engineering students in sense of humor); and backgrounds (e.g., students from the city scored higher than students from rural areas). A recent study used two student samples from Hong Kong and Guangzhou (China) and showed the relationship between playfulness and their humor styles. The results suggested that highly playful Chinese students preferred using affiliative and self-enhancing humor to amuse themselves and others (Yue et al., 2016).

One recent study (Barnett, 2017) addressed the cultural aspect of playfulness by comparing three groups of Chinese female graduate students who varied in the length of time they had lived in the United States, and thus had been exposed to American culture, with a fourth group of American students who were born in the United States and had always lived there. Her findings suggest that playfulness can be culturally transmitted to Chinese women who are from a different culture. However, to the best of our knowledge, there are no direct comparisons of adult playfulness in Western and Eastern cultures.

The Present Study

The aim of the current study was threefold. First, we aimed to establish measurement equivalence of two playfulness instruments, one of which was developed in Switzerland (i.e., *Short Measure of Adult Playfulness*, SMAP; Proyer, 2012b), and one in Taiwan (*Adult Playfulness Questionnaire*, APQ; Yu et al., 2003). Second, we aimed to investigate cross-cultural playfulness by comparing mean level differences of playfulness between students from the West (German-speaking countries) and the East (mainland China). Chinese students were expected to be less playful in comparison to German-speaking students using both measures, Western and Eastern. Third, we aimed to explore the cross-cultural differences of playfulness in different situations and to estimate the social appropriateness of playfulness in these situations.

MATERIALS AND METHODS

Participants

Sample 1 consisted of 143 German-speaking students aged 18–48 years (M = 23.2, SD = 4.6) from Switzerland (n = 100), Germany (n = 31) and Austria (n = 12). Of these, 72.0% were female (n = 103). Approximately two-thirds were single (66.4%) and slightly less than a third were in a relationship or married (32.9%). About a third held a Bachelor of Science degree from a university (31.5%); of the rest, 67.1% held a school-leaving diploma qualifying for attending university, and 1.4% had completed compulsory education.

Sample 2 consisted of 176 university students who were aged 18–24 years (M = 19.8, SD = 1.2) and lived in Guangzhou, mainland China. Of these, 56.3% were female (n = 99). Three-quarters of the participants were single (n = 132, 75.0%) while 22.2% (n = 39) were in a relationship; the 5 remaining participants did not indicate their marital status. Almost all participants held a university degree (Bachelor of Science) or were currently enrolled at a university (n = 169, 96.0%).

Sample 3 consisted of 100 university students aged 18–27 years (M = 20.4, SD = 1.5) and living in Beijing, mainland China. Of these, 69% were female (n = 69). The majority of the participants (n = 83, 83.0%) were single. Almost all of them held a university degree (Bachelor of Science) or were currently enrolled at a university (n = 95, 95.0%).

Instruments

Short Measure of Adult Playfulness (SMAP)

The SMAP (Proyer, 2012b) consists of five items that allow for a global assessment of adult playfulness. Answers are given on a 7-point Likert scale ranging from 1 = "*strongly disagree*" to 7 = "*strongly agree*". All items are positively keyed. Previous data (e.g., Proyer, 2012b; Proyer and Ruch, 2011) showed a one-dimensional solution with satisfactory reliabilities (Cronbach's $\alpha > 0.80$). The SMAP also converges well with other measures of playfulness (Glynn and Webster, 1992, 1993; Barnett, 2007) and the need for play (Jackson, 1974). High scorers in the SMAP expressed higher approval and liking of an unstructured working environment and higher approval and liking of an abstract painting in comparison with low scorers who expressed greater disapproval of the unstructured work space and an abstract art piece; no differences were found in rating for an orderly work space and simple geometric figures (Proyer, 2012b). The Chinese version of the SMAP (SMAP-CN) was developed for the current study using the back-translation procedure (see below). It consists of the same items and scoring rules as the German version. A sample item is "我是一个乐玩的人 (I am a playful person)". We used the term "乐玩" as a translation for playful because it can reduce the negative linguistic bias of the current translation ("玩兴") by the Taiwanese scholars. The SMAP-CN can be found in the online Supplementary Materials of the study.

Adult Playfulness Questionnaire (APQ)

The APQ scale (Yu et al., 2003) consists of 29 items loading on three factors: "*pleasantry,*" "*initiative and concentration,*" and "*creativity*". All items are positively keyed and utilize a five-point Likert scale (1 = "*strongly disagree,*" 5 = "*strongly agree*"). Yu and her colleagues (Yu et al., 2003) reported a satisfying internal consistency (Cronbach $\alpha = 0.95$ for the total score) and acceptable construct, concurrent, and discriminant validities. One sample item of the Chinese version (Yu et al., 2003) is "从事喜欢的活动, 会觉得时间过得很快, 甚至忘了时间 (For whatever I love to do, time will fly by and I even forget about the time spent on it)". A German translation of the items has been used in a previous study (Proyer and Jehle, 2013).

Brief Rating List of Playfulness in Different Situations (BRLPS)

The BRLPS was developed for this study to assess playfulness in different situations. It consists of 14 different contexts with two perspectives: the self-perspective and the perceived society perspective. The self-perspective covers the level of playful behavior expressed by participants when they are with certain people (e.g., friends, family), or when they are in certain situations (e.g., at the workplace). The perceived society perspective covers the perception of how society would rate the appropriateness of being playful around these people or in the given situation. Answers are given on a 7-point scale ranging from 1 = "*Not at all*" to 7 = "*Very much*" and also include "*Not applicable.*" The Chinese version of the scale was developed for the current study and had the identical contexts as well as scoring rules with the German version. Two sample situations are: in German, "zusammen mit Grosseltern (together with grandparents)" and "in der Öffentlichkeit (in public)"; in Chinese, "跟爷爷奶奶一起 (together with grandparents)" and "在公开场合 (in public)". The German and Chinese versions used in this study are provided in the online Supplementary Materials.

Procedure[1]

Translation

The Short Measure of Adult Playfulness (SMAP; Proyer, 2012b) was translated from English into Chinese using Brislin's (1970) back-translation model. The first author of the current study did the initial translation from Chinese into German. Afterward a master student who studied psychology in mainland China back-translated all the items independently. The two versions of the instruments were compared for concept equivalence by another Chinese student who was studying for a Ph.D. in psychology at the University of Cambridge. Once an error or disagreement was found in the back-translated version, the first author tried to retranslate the item and discussed this with the original author of the scale (the second author of the current study). This procedure continued until all three translators agreed that the two versions of the instruments were identical and had no errors in meaning. As mentioned before, there was no corresponding term for playfulness in Chinese, and the Taiwanese translation "玩兴 (wanxing)" could not be used because the term is not used in daily language in mainland China. Hence participants would have more than one way of understanding its meaning (e.g., it could mean interested in playing, or a trait of playing), which would lead to confusion and stronger linguistic bias. Therefore, after discussing the issue with experts as well as laypeople, playfulness was translated as "乐玩派 (lewanpai)" in the current study, which means a person who enjoys playing. An explanation of playfulness was presented in the introduction to the SMAP for both German-speaking participants and Chinese. Thus we ensured that all participants had an identical understanding of the concept. The Adult Playfulness Questionnaire (APQ; Yu et al., 2003) was adapted into simplified Chinese accordingly and the word "玩兴 (wanxing)" was replaced with "乐玩派 (lewanpai)" for the participants from mainland China.

Recruitment

We trained two undergraduate students who were studying psychology at Sun Yat-sen University (Guangzhou) and Renmin University (Beijing) to recruit the participants in mainland China in paper-pencil form. Meanwhile, a German version and a Chinese version of the questionnaires were created online through a web-based survey solution (SurveyMonkey). Advertisements were placed on the Internet and via email (e.g., students' forums, social media, university mailing list, etc.), and in a public place such as a pin board to get as many participants as possible. As a result, we had access to students who studied in German-speaking countries (mainly Switzerland) or in mainland China. To motivate the participants, participants living in Guangzhou received a postcard as a gift, while participants

[1] This is part of a larger data collection; other instruments have been filled out using the same sample. However, they were not relevant to the current research question and the data presented in the manuscript have not been published elsewhere.

who studied psychology at the University of Zurich were given 0.75 experiment-hours or a piece of sushi as an incentive for participation. Participants were not paid for their service, but were given a written feedback of individual results when interest was expressed.

Data Collection

All questionnaires (paper–pencil form) collected in mainland China were delivered to Switzerland by DHL and Federal Express Corporation Inc., and were then scanned using the software Remark Office OMR (version 6).

RESULTS

Examination of Measurement Invariance

Although the questionnaires were translated using a translation-back-translation procedure, measurement equivalence must be established for enabling comparisons (see e.g., Mullen, 1995; van de Vijver and Tanzer, 2004). Metric measurement invariance was tested for the SMAP and APQ (testing each facet separately) using a multi-group CFA with the lavaan (Rosseel, 2012) and semTools packages (semTools Contributors, 2015) in *R*. It was tested by forcing all item[2] loadings to be equal across groups. This model was then compared with the baseline model that allows a free estimation of the item loadings, comparing the difference in the CFI and the RMSEA. Changes of $\leq|0.01|$ in the CFI and changes of $\leq|0.015|$ in the RMSEA were used as cut-offs to indicate measurement invariance, based on the recommendations by Cheung and Rensvold (1999) and Chen (2007). Metric measurement invariance was tested across the three samples. The results are displayed in **Table 1**, which depicts the fit indices of the baseline model (in which the item loadings were allowed to vary freely), the metric invariance model (in which the item loadings were constrained to be equal across groups), and the changes in the CFI and the RMSEA.

As shown in **Table 1**, the baseline model had an adequate fit to the data for SMAP and the creativity facet of the APQ. However, the remaining facets of the APQ had a rather weak fit to the data. The CFI changes were $<|0.01|$ for the SMAP and pleasantry and the RMSEA changes were $<|0.015|$ for creativity. Follow-up analyses were conducted for assessing partial measurement invariance of the APQ, comparing the metric invariance of each of the items in the three samples. The metric invariance was supported for each item in all three facets of the APQ, as the CFI change between the baseline model and the metric invariance model was $<|0.01|$ (with a range from $|0.000|$ to $|0.008|$). Thus, partial measurement invariance was supported in our study and this allows us to meaningfully compare the mean level differences between the playfulness scores across the samples[3].

Correlations Among the Playfulness Measures and Situational Ratings of Playfulness

In the next step, to test for overlaps and to establish validity of these measures, we correlated the scores obtained on the two measures of playfulness in each sample. The results are presented in **Table 2**. The correlation coefficient between the total score of the APQ and the SMAP was 0.61 for the German-speaking sample, and 0.46 for both Chinese samples. Of the subscales of the APQ, pleasantry correlated highest with the SMAP (coefficients ranged from 0.43 to 0.69), while the other two facets correlated numerically much lower with the SMAP. The

[2] All items of SMAP were used while, of the 29 items of the APQ (Yu et al., 2003), 11 were excluded from the current analysis due to high double loadings in the factor analysis.

[3] The Tucker's φ coefficient (Tucker, 1951; Lorenzo-Seva and ten Berge, 2006) was additionally computed out of Principal Component Analyses for the two measures and indicated excellent equivalence for the SMAP (all values of $\varphi \geq 0.997$). The Tucker's φ coefficients were excellent for the "*pleasantry*" and "*initiative and concentration*" facets of the APQ (all values of $\varphi \geq 0.92$) facets and adequate for the "*creativity*" facet. The coefficients were $\varphi = 0.94$ between Sample 1 and Sample 2; 0.97 between Sample 2 and Sample 3; and 0.88 between Sample 1 and Sample 3.

TABLE 1 | Fit indices of models assessing metric (fixed loadings) invariance of SMAP and APQ across three samples.

Measurement invariance models	*df*	χ^2	CFI	RMSEA	CFI change	RMSEA change
SMAP						
Baseline model	15	39.63	0.97	0.11	–	–
Metric invariance	23	47.45	0.98	0.09	0.000	0.021
APQ						
Pleasantry						
Baseline model	60	318.43	0.82	0.18	–	–
Metric invariance	74	337.70	0.82	0.17	0.004	0.017
Initiative and Concentration						
Baseline model	27	163.88	0.90	0.20	–	–
Metric invariance	37	189.52	0.88	0.18	0.012	0.020
Creativity						
Baseline model	6	20.96	0.97	0.14	–	–
Metric invariance	12	38.17	0.95	0.13	0.022	0.009

χ^2, chi square. CFI, comparative fit index, RMSEA, root mean square error of approximation. SMAP, Short Measure of Adult Playfulness. APQ, Adult Playfulness Questionnaire.

TABLE 2 | Correlations between SMAP and APQ in three samples (Controlled for Gender).

	SMAP		
	Sample 1 (*n* = 116)	**Sample 2 (*n* = 164)**	**Sample 3 (*n* = 91)**
APQ			
Total	0.61***	0.46***	0.46***
Pleasantry	0.69***	0.43***	0.52***
I&C	0.30**	0.37***	0.11
Creativity	0.21*	0.27***	0.36***

*SMAP, Short Measure of Adult Playfulness. APQ, Adult Playfulness Questionnaire. I&C, Initiative & Concentrating. Sample 1, German-speaking sample; Sample 2, Guangzhou sample, and Sample 3, Beijing sample. *p < 0.05, **p < 0.01, ***p < 0.001, two-tailed.*

correlation coefficients between 'initiative and concentration' and the SMAP varied across the three samples (from 0.11 to 0.37), while creativity demonstrated robust associations with the Guangzhou sample (0.30) and the German sample (0.37), but the coefficient was lower for the Beijing sample (0.11).

When correlating the situational ratings of playfulness with the two measures of playfulness, the coefficients for SMAP and APQ were largely around 0.30 (see **Table 3**). The self-ratings of different situations for the German-speaking sample showed a median of $r = 0.33$ for SMAP and 0.24 for APQ; the Beijing sample showed a median of $r = 0.31$ for both; and the Guangzhou sample was numerically smaller (median $r = 0.12$ for SMAP and 0.23 for APQ). When analyzing the perceived society perspective, they were uncorrelated in the German-speaking sample while there were some associations in the two Chinese samples (see **Table 4**; e.g., the "with parents" situation). These results reveal that in a collectivistic country like China, the perceived society norms had an impact on the associations with playfulness.

Descriptive Statistics of the Scales

Table 5 shows the descriptive statistics of the SMAP and APQ. An examination of each scale's skewness and kurtosis suggested that they were all normally distributed in three samples. Their internal consistency was high in all three samples (all ≥ 0.71). The mean scores were comparable (where previous data was available) to prior research (Yu et al., 2003; Proyer, 2012b; Proyer and Jehle, 2013; Yue et al., 2016). We also checked whether they correlated with gender, age, and collection mode (paper and pencil vs. online). Correlation coefficients with age and collection mode were negligible, but there were minor associations with gender (all < 5% overlapping variance). Nevertheless, we decided to control for the potential effects of gender in the analyses conducted subsequently.

Cross-Cultural Differences in Playfulness

In order to explore differences in playfulness between German-speaking participants and Chinese participants, a one-way analysis of covariance (ANCOVA) was conducted (covariate: gender). The independent variable "*Region*" involved three levels: German-speaking participants, participants from Guangzhou, and participants from Beijing. The dependent variables were the playfulness scores in the SMAP and APQ. The preconditions for the ANCOVA were met. In particular, the homogeneity of the regression effect was evident for the covariate, and the covariate was linearly related to the dependent measure. The results are displayed in **Table 6**.

The table shows that the main effect of the variable *Region* for the SMAP (Proyer, 2012b) was not significant ($F[2,412] = 1.59$, $p = 0.205$). The main effect of *Region* for the total score of the APQ (Yu et al., 2003) was significant ($F[2,378] = 4.22$, $p = 0.008$, $\eta_p^2 = 0.02$), as well as being significant for the subscales *Creativity* and *Pleasantry*. Comparisons revealed that the German-speaking participants scored higher in the total score of APQ than the

TABLE 3 | Correlations between SMAP, APQ and situational ratings of playfulness (self-perspective).

	Sample 1		Sample 2		Sample 3	
	SMAP	**APQ**	**SMAP**	**APQ**	**SMAP**	**APQ**
With grandparents	0.04	0.02	0.28**	0.15	0.22*	0.08
With parents	0.24**	0.20*	0.31**	0.23**	0.43***	0.31**
With siblings	0.27**	0.33**	0.28**	0.24**	0.41***	0.32**
With partner	0.35***	0.33**	0.29**	0.26**	0.21	0.15
With children	0.10	0.12	0.12	0.29**	0.29*	0.06
With friends	0.44***	0.40**	0.39***	0.41***	0.32**	0.27**
With classmates	0.33***	0.30**	0.38***	0.31***	0.40***	0.30**
With work colleagues	0.32**	0.33**	0.12	0.22*	0.33**	0.35**
With teachers	0.32***	0.25**	0.12	0.16*	0.35***	0.30**
With boss	0.35***	0.22*	−0.03	0.02	0.29*	0.34**
In public	0.42***	0.20*	0.11	0.29***	0.33**	0.32**
In business meeting	0.35***	0.18	−0.05	−0.01	0.27	0.56***
On online-forum	0.41***	0.20	0.06	0.19*	0.05	0.31**
On social media	0.20*	0.26**	0.08	0.23**	0.22*	0.36***

*SMAP, Short Measure of Adult Playfulness. APQ, Adult Playfulness Questionnaire. Sample 1, German-speaking sample; Sample 2, Guangzhou sample, and Sample 3, Beijing sample. *p < 0.05, **p < 0.01, ***p < 0.001, two-tailed.*

TABLE 4 | Correlations between SMAP, APQ and situational ratings of playfulness (society-perspective).

	Sample 1		Sample 2		Sample 3	
	SMAP	**APQ**	**SMAP**	**APQ**	**SMAP**	**APQ**
With grandparents	−0.12	−0.06	0.21**	0.12	0.28**	0.10
With parents	0.02	−0.07	0.21**	0.23**	0.34**	0.11
With siblings	0.03	−0.01	0.28***	0.26**	0.19	0.15
With partner	0.10	0.08	0.08	0.25*	0.18	0.06
With children	−0.04	0.03	0.08	0.24*	0.09	0.03
With friends	0.08	0.11	0.20**	0.31***	0.20	0.15
With classmates	−0.09	−0.17	0.17*	0.25**	0.34**	0.17
With work colleagues	0.06	0.00	−0.05	0.04	0.25*	0.37**
With teachers	0.10	0.08	0.06	0.21**	0.37**	0.30**
With boss	0.10	0.07	0.04	0.15	0.33**	0.31*
In public	0.10	−0.06	0.03	0.21**	0.22*	0.18
In business meeting	0.08	−0.03	0.22	0.21	0.33*	0.44**
On online-forum	0.04	−0.01	0.04	0.18*	−0.09	0.07
On social media	−0.02	0.05	0.12	0.24**	−0.01	0.11

*SMAP, Short Measure of Adult Playfulness. APQ, Adult Playfulness Questionnaire. Sample 1, German-speaking sample; Sample 2, Guangzhou sample, and Sample 3, Beijing sample. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, two-tailed.*

Guangzhou participants (Cohen's $d = 0.32$), but did not differ from the Beijing participants. Similar results were found for the subscales *Creativity* and *Pleasantry* (Cohen's $d = 0.23$ and 0.44, respectively) when comparing the German-speaking sample and the Guangzhou sample. A *post hoc* test (Fisher's LSD) was conducted to explore the potential difference within mainland China (Guangzhou sample vs. Beijing sample). We found that the Beijing sample scores were higher on the total score of APQ (Cohen's $d = 0.23$) and on the *Creativity* subscale (Cohen's $d = 0.23$).

Mean Level Differences of Playfulness in Different Situations

We averaged the responses of the *Brief Rating List of Playfulness in Different Situations* (BRLPS; Pang and Proyer, 2013a,b, Unpublished). In order to group the situations into different categories, we conducted a principal component analysis (PCA) on the 15 situations with oblique rotation. The eigenvalues for the first five components were 4.53, 2.63, 1.66, 1.12, and 0.89 (self-reported playfulness in given situations) and 5.20, 2.96, 1.54, 0.98, and 0.86 (perceived societal perspective). Three factors were extracted in both analyses and tentatively labeled as (a) *private situations* (e.g., with relatives and friends); (b) *formal situations* (e.g., with work colleagues and teachers); and (c) *university/online settings* (e.g., online forum and social media). A one-way analysis of covariance (ANCOVA) was conducted (covariate: gender) and *post hoc* tests (Fisher's LSD) were used for pairwise comparisons after obtaining significant differences. **Table 7** (self-perspective) and **Table 8** (perceived society perspective) show the sample size (n), mean score (M), standard deviation (SD), and findings of the ANCOVA (F score and p-value).

Self-Perspective

As displayed in **Table 7**, the three samples differed in their ratings of private situations ($F[2,405] = 8.91$, $p < 0.001$) and of university/online settings ($F[2,387] = 29.08$, $p < 0.001$). However, no differences were found in the formal situations ($p = 0.075$, one tailed). *Post hoc* tests showed that, in comparison with the Guangzhou sample, participants from the German-speaking sample and the Beijing sample seemed to be more playful in their private life (Cohen's $d = 0.52$ and 0.29 respectively). Additionally, the German-speaking sample scored lower in playfulness than both the Guangzhou sample (Cohen's $d = 0.79$) and the Beijing sample (Cohen's $d = 0.89$) when they were in university/online situations, whereas no differences were found between the two Chinese samples.

Perceived Society Perspective

As displayed in **Table 8**, from the perceived perspective of society, differences across the three samples were found in private situations ($F[2,391] = 6.80$, $p < 0.001$), formal situations ($F[2,391] = 49.58$, $p < 0.001$), and in university/online settings ($F[2,377] = 18.05$, $p < 0.001$). *Post hoc* tests showed that, in comparison with the Guangzhou sample, the German-speaking sample rated that it would be more appropriate from society's perspective to behave playfully when they were in private situations (Cohen's $d = 0.45$). The Beijing sample did not differ from the other two samples in their private situations from the perceived society perspective. Interestingly and unexpectedly, in comparison with both Chinese samples, participants in the German-speaking sample indicated that it would be *less* appropriate from the perspective of society to behave playfully when in formal situations (Cohen's $d = 1.00$ for the Guangzhou sample, and Cohen's $d = 1.13$ for the Beijing sample). No differences were found across the two Chinese samples in the formal situations from the perceived society perspective. Also, in comparison with the German-speaking sample, both Chinese samples rated that it would be more appropriate from society's perspective to behave playfully in university/online settings (Cohen's $d = 0.52$ for

TABLE 5 | Psychometric characteristics and correlations with gender of SMPA and APQ in three samples.

	Sample 1						Sample 2						Sample 3					
	M	*SD*	α	*S*	*K*	r_{sex}	*M*	*SD*	α	*S*	*K*	r_{sex}	*M*	*SD*	α	*S*	*K*	r_{sex}
SMAP																		
Total	5.03	1.10	0.85	−0.99	0.85	−0.17*	4.86	1.13	0.82	−0.83	1.00	−0.09	5.02	1.13	0.82	−0.48	−0.61	−0.04
APQ																		
Total	3.87	0.47	0.84	−0.73	0.46	−0.17	3.69	0.65	0.91	−0.30	−0.08	0.01	3.83	0.63	0.90	−0.60	0.85	−0.13
Pleasantry	3.95	0.64	0.83	−0.88	0.65	−0.04	3.63	0.81	0.88	−0.29	−0.40	−0.03	3.75	0.80	0.88	−0.52	−0.34	−0.10
I&C	4.09	0.57	0.79	−0.48	−0.32	−0.20*	4.10	0.73	0.91	−0.85	0.58	0.12	4.23	0.72	0.87	−0.85	0.25	−0.01
Creativity	3.36	0.71	0.71	−0.19	0.32	−0.20*	3.18	0.87	0.83	−0.28	−0.34	−0.04	3.39	0.94	0.86	−0.06	−0.66	−0.22*

*SMAP, Short Measure of Adult Playfulness. APQ, Adult Playfulness Questionnaire. I&C, Initiative & Concentrating. Sample 1, German-speaking sample (116 ≤ n ≤ 125); Sample 2, Guangzhou sample (165 ≤ n ≤ 168), and Sample 3, Beijing sample (92 ≤ n ≤ 94). M, mean; SD, standard deviation; α, Cronbach alpha; S, Skewness; K, Kurtosis; r_{sex}, Correlation with gender (1 = "male," 2 = "female"). *p < 0.05, **p < 0.01, ***p < 0.001, one-tailed.*

the Guangzhou sample, and Cohen's d = 0.79 for the Beijing sample).

DISCUSSION

Recent years have seen a growing interest in the study of adult playfulness as a personality trait. To the best of our knowledge, only a few studies have taken a cross-cultural perspective into account (see Barnett, 2017), and a direct comparison of different cultures was missing. We aimed to narrow this gap in the literature by collecting data from German-speaking countries and an Eastern country (China), and by analyzing data using an instrument developed in a German-speaking country and one that has been developed in Taiwan. This allows for a more comprehensive analysis of potential differences, in contrast to using only one instrument that has been developed from a certain cultural perspective. Our expectations derived from previous literature were only partially supported as there were only a few differences across the two tested regions. Hence the differences were smaller than expected at the trait level. One might argue that future research should probably focus on the identification and analysis of inter-individual differences in playful behavior in specific situations or personal relationships, and in the perception of expectations found in societal norms, as such findings are potentially more informative about cultural differences compared to our initial study. The results, however, provide an initial overview not only of cross-cultural diversity but also of cross-cultural similarities, both of which contribute toward a better understanding of the nature of playfulness.

As expected, mean level differences in playfulness can be observed between the German-speaking sample and the Guangzhou sample with small to middle effect sizes. This could be explained by the negative bias toward play in Chinese culture. As mentioned above, play is traditionally considered as not obeying the rules and is mostly negatively connoted. However, given the effect sizes, the differences should not be over-interpreted. An observation that may be of interest for follow-up studies is that the playfulness scores of the Beijing sample were always located between the other two samples: In certain scales and certain situations (e.g., the pleasantry subscale and in university/online settings), they rated themselves similarly to the Guangzhou sample, but in other scales and other contexts they rated themselves similarly to the German-speaking sample (e.g., the creativity subscale and in private situations). This might be due to differences in the mindset between South China and North China and, therefore, within-country differences may also provide a fruitful area for future research. For example, people in North China have a flourishing tradition of enjoying "cross talk" (相声; xiangsheng), which concentrates on language and word play, such as using puns, homonyms, dialects, idioms, and double entendre (Chey, 2014). This may also reflect a somewhat playful nature of people who live in the north of China and may have led to higher subjective ratings of playfulness than of people who live in the south. Hence within-country differences need consideration when thinking about playfulness in China.

TABLE 6 | Mean level differences of playfulness in three samples.

	Sample 1			Sample 2			Sample 3			Variance	
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*F*	*p*
SMAP	143	5.03	1.10	175	4.85	1.13	98	5.01	1.13	1.59	0.205
APQ	120	3.86_{a}	0.47	168	3.68_{ab}	0.65	94	3.83_{b}	0.63	4.22	0.015
Pleasantry	119	3.95_{a}	0.64	168	3.63_{a}	0.81	96	3.75_{b}	0.80	6.52	0.002
I&C	120	4.09	0.57	168	4.11	0.73	96	4.25	0.72	1.80	0.166
Creativity	120	3.36_{a}	0.71	168	3.18_{ab}	0.87	96	3.39_{b}	0.93	3.37	0.035

SMAP, Short Measure of Adult Playfulness. APQ, Adult Playfulness Questionnaire. I&C, Initiative & Concentrating. Sample 1, German-speaking sample; Sample 2, Guangzhou sample; Sample 3, Beijing sample. Means in a row sharing subscript are statistically different from each other at $p < 0.05$ (two-tailed) utilizing planed contrast (when comparing German-speaking sample with the two Chinese sample separately) and the Fisher's least significant difference (LSD) procedure (when comparing the two Chinese samples). For all measures, higher means indicate higher playfulness scores.

TABLE 7 | Mean level differences of playfulness in different situations in three samples (self-perspective).

	Sample 1			Sample 2			Sample 3			Variance	
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*F*	*p*
Private situation	139	5.43_{a}	0.78	175	4.95_{ab}	1.05	95	5.24_{b}	0.94	8.91	<0.001
With grandparents	122	3.87	1.74	166	3.58	1.75	90	3.88	1.70	0.90	0.205
With parents	137	4.87_{a}	1.51	175	4.35_{a}	1.60	95	4.62_{b}	1.68	3.20	0.021
With siblings	121	5.92_{ab}	1.09	160	5.33_{a}	1.38	87	5.55_{b}	1.11	8.36	<0.001
With partner	104	6.21_{a}	1.11	108	5.86_{a}	1.29	64	6.14_{b}	0.99	2.58	0.039
With children	103	5.82	1.19	91	5.48	1.37	57	5.79	1.24	1.52	0.110
With friends	135	5.99	0.97	173	5.99	1.06	95	6.09	0.90	0.44	0.321
Formal situation	139	3.38	1.18	174	3.66	1.06	95	3.67	1.19	2.26	0.105
With work colleagues	117	4.40	1.54	109	4.47	1.24	60	4.47	1.32	0.62	0.470
With teachers	129	2.39_{ab}	1.38	174	3.37_{a}	1.42	94	3.49_{b}	1.47	21.28	<0.001
With boss	122	2.73	1.54	106	3.09	1.55	61	3.18	1.51	2.21	0.056
In public	139	4.40	1.48	174	4.15	1.29	94	4.09	1.33	2.06	0.065
In business meeting	99	2.35_{ab}	1.35	82	2.90_{a}	1.64	50	3.02_{b}	1.55	4.28	0.008
University/online settings	139	4.63_{ab}	1.22	175	5.12_{a}	1.37	95	5.16_{b}	1.01	8.30	<0.001
With classmates	136	5.44_{ab}	1.15	174	5.22_{a}	1.08	95	5.18_{b}	1.19	2.68	0.035
On online-forum	85	3.64_{a}	1.78	127	4.87_{a}	1.59	74	5.23_{a}	1.31	22.99	<0.001
On social media	119	4.08_{ab}	1.61	170	5.19_{a}	1.44	93	5.19_{b}	1.33	22.69	<0.001

Sample 1, German-speaking sample; Sample 2, Guangzhou sample; Sample 3, Beijing sample. The different participant number is due to the fact that the situation items did not apply for everyone in the sample and we offered them the option "it doesn't apply to me" for each situation. For instance, because most of the participants are students and it is difficult for some of them to judge the situation with children, they would decide to answer "it doesn't apply to me." Therefore, different situations end up with different participant numbers. Means in a row sharing subscript are statistically different from each other at $p < .05$ (one-tailed) according to Fisher's least significant difference (LSD) procedure. For all measures, higher means indicate higher playfulness scores.

Recent work has suggested that the south-north difference in mainland China mirrors the differences between collectivistic East Asia and the more individualistic Western world. Talhelm et al. (2014) proposed the so-called "*Rice Theory*" and argued that the differences seem to appear because southern China has grown rice for 1000s of years, whereas the north has grown wheat. They argue that a history of farming rice makes cultures more interdependent whereas farming wheat makes cultures more independent, and that these agricultural legacies continue to affect people in the modern world. Their findings, based on 1,162 Han Chinese participants, confirmed their assumption that rice-growing southern China is more interdependent (Talhelm et al., 2014). This is also in accordance with the current findings. One might argue that people living in South China (e.g., Guangzhou), where they have grown rice for 1000s of years, are more collectivistic and interdependent and, therefore, rated themselves lower in playfulness in this study. However, both Renmin University and Sun Yat-sen (Zhongshan) University belong to the high-status universities in China (both ranked in the top 10 in various university rankings in China) and have students from all over China. Consequently, we checked the admission numbers of the two universities at the year of data-collection for each province in China. About 53% of the admitted students at Sun Yat-sen University were from the Guangdong province, while students from the other provinces were less well represented (about 2% on average). About 11% of the admitted students at Renmin University were from Beijing, and about 3% on average from other provinces. We also checked the number of students in the wheat-rice categorization (Talhelm et al., 2014). The portion of students from the wheat culture (north of China)

TABLE 8 | Mean level differences of playfulness in different situations in three samples (perceived society perspective).

	Sample 1			Sample 2			Sample 3			Variance	
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*F*	*p*
Private situation	132	5.63_{a}	1.02	171	5.15_{a}	1.10	92	5.45_{b}	0.92	6.80	<0.001
With grandparents	120	4.67_{a}	1.66	161	4.12_{ab}	1.75	87	4.68_{b}	1.58	3.61	0.014
With parents	131	5.03_{a}	1.34	168	4.64_{ab}	1.65	92	5.02_{b}	1.43	2.51	0.042
With siblings	120	6.06_{ab}	1.15	156	5.49_{a}	1.47	83	5.72_{b}	1.05	6.44	0.001
With partner	121	5.76	1.22	104	5.82	1.17	62	5.89	1.07	0.34	0.358
With children	115	6.31_{ab}	1.17	92	5.68_{a}	1.24	59	5.81_{b}	1.27	6.87	0.001
With friends	128	6.03	1.08	170	5.90	1.12	91	6.01	1.01	0.47	0.312
Formal situation	132	2.75_{ab}	1.15	171	4.03_{a}	1.18	92	4.28_{b}	1.24	47.94	<0.001
With work colleagues	124	3.81_{ab}	1.61	105	4.67_{a}	1.44	63	4.94_{b}	1.28	15.57	<0.001
With teachers	125	2.06_{ab}	1.47	169	3.98_{a}	1.68	91	4.20_{b}	1.55	67.66	<0.001
With boss	120	2.08_{ab}	1.46	105	3.64_{a}	1.76	61	3.89_{b}	1.72	36.47	<0.001
In public	132	3.52_{ab}	1.53	171	4.27_{a}	1.48	90	4.49_{b}	1.50	15.16	<0.001
In business meeting	111	1.81_{ab}	1.25	83	3.41_{a}	2.05	49	3.76_{b}	1.94	32.29	<0.001
University/online settings	131	4.79_{ab}	1.25	171	5.28_{a}	1.16	92	5.17_{b}	1.02	12.70	<0.001
With classmates	131	5.24	1.31	170	5.42	1.21	92	5.52	1.24	1.48	0.115
On online-forum	105	4.12_{a}	1.75	124	5.01_{a}	1.61	72	5.61_{a}	1.30	19.96	<0.001
On social media	121	4.83_{ab}	1.40	166	5.39_{a}	1.40	90	5.66_{b}	1.29	10.30	<0.001

Sample 1, German-speaking sample; Sample 2, Guangzhou sample; Sample 3, Beijing sample. The different participant number is due to the fact that the situation items did not apply for everyone in the sample and we offered them the option "it doesn't apply to me" for each situation. For instance, because most of the participants are students and it is difficult for some of them to judge the situation with children, they would decide to answer "it doesn't apply to me." Therefore, different situations end up with different participant numbers. Means in a row sharing subscript are statistically different from each other at $p < 0.05$ (one-tailed) according to Fisher's least significant difference (LSD) procedure. For all measures, higher means indicate higher playfulness scores.

at Sun Yat-sen University was only about 18%, but students from the rice culture (south of China) was 63%. At Renmin University, the portion of students from the wheat culture (north of China) was 43%, while the portion of students from the rice culture (south of China) was 23%. (The percentages exclude students from the three major herding provinces and the rice-wheat border provinces.) Therefore, we could conclude that the students are still representative of the north-south difference.

Contrary to our expectations, the German-speaking participants indicated a *lower* acceptance in society to playfulness in formal situations in comparison with the Chinese participants; in particular, in work situations (e.g., business meetings) or when interacting with teachers. The differences were even larger from the perceived society perspective. This may reflect actual differences (e.g., in implicit agreements on how business meetings are conducted), but a limitation of our study must be noted at this point as we have only tested students with potentially limited experiences of business settings. Additionally, the infrastructure and atmosphere of the universities in mainland China and in German-speaking countries differ (e.g., almost all students in China live on the campus which is separated from the outside world, whereas students in German-speaking countries often live with their parents or in a shared flat in the city). This may help to explain the differences. Students in German-speaking countries may *perceive* a business setting (based on their limited experience) as more formal and structured than their Chinese counterparts. Additionally, one might argue that the rules for such meetings in German-speaking countries are potentially more implicit and take experience to understand, whereas the setting in China is more structured. Given that many students in German-speaking countries work at least part-time to help finance their education, they are presumably in low-status jobs with little room for expressions of playfulness. Additionally, it must be noted that the students from mainland China in our sample were younger and most of them were studying full time. Therefore, when the Chinese students were questioned about general situations such as "with work colleagues" or "in business meetings," they seemed to have even less work-related experience and therefore extrapolated experience from their university lives (e.g., when thinking of colleagues from students' associations or meetings with students' assignment groups). In contrast, the students from German-speaking countries would potentially recall experiences from real workplaces, such as their job as barkeeper, waiter/waitress, intern, etc., which were rather low in the hierarchy of a company. Accordingly, playfulness was not encouraged in these situations because of the low level in the hierarchy. Overall therefore, these findings need to be interpreted with caution as a replication is needed involving participants with more work experience.

In our study, we used a new translation for playfulness (i.e., "乐玩派 (lewanpai)") and, therefore, had to provide a description of what this term means in the introduction to the questionnaires. This was done to ensure that all participants had an identical understanding of the concept and to avoid cultural bias. Contrariwise, however, the usage of such an explanation might reduce potential cultural differences too much in the sense that the description could have been too narrow. Although we obtained satisfactory psychometric data, more validity studies (e.g., divergent/convergent validity) of the instruments are needed in the future. This is of particular importance given that

many of the measures currently in use seem to have a bias in terms of an unwanted overlap with broader personality traits—mainly emotional stability and extraversion (Proyer and Jehle, 2013; Proyer, 2017) —and lack conceptual distinctiveness from potentially related traits such as humor or creativity (e.g., Proyer et al., in press b).

Our expectation was that playfulness would be more prevalent in individualistic than in collectivistic cultures, because people in collectivistic cultures tend to display more conformity behavior than those in individualistic countries (Hofstede, 2001; p. 236). However, there are also other cultural dimensions that might play a role in explaining why German-speaking countries would score higher in playfulness than China. One candidate is the tight vs. loose culture dimension (Gelfand et al., 2011). "Tight" cultures refer to those that have strong norms and a low tolerance of deviant behavior, whereas "loose" cultures have weak norms and a high tolerance of deviant behavior. Yet both China and German-speaking countries were not on the extreme end of the scale. Future studies might aim at comparing countries with extreme dimensions (such as Pakistan vs. Netherlands). Hence, future studies might consider this variable as well as others (e.g., a comparison with English-speaking countries as they are the highest on the "loose" dimension). Our initial study shows that we can expect from such studies a contribution to the understanding of playfulness from a cross-cultural perspective.

A further limitation must be noted due to a potential confounding from an acquiescence bias. Both measures employed in this study consisted of positively worded items only. One might argue that there are differences with respect to acquiescence in the two tested countries. Smith (2004) found that acquiescence was positively related to collectivism, which supports the idea that acquiescence bias may be higher in China and may interact with (or counterbalance) the initially expected lower expressions in playfulness. It is possible therefore that our findings are biased by country-level differences in acquiescence bias, leading to an underestimation of the actual differences. However, it must be noted that data on the self-other agreement in playfulness (Ostendorf et al., 1986; Fekken et al., 1987; Proyer, 2017; Proyer and Brauer, 2018) suggest good convergence. Hence, while acquiescence may play a role and should be controlled for, self-ratings seem to reflect the perception of (well) acquainted others. There is even evidence that people can gather information on a person's playfulness in zero-acquaintance settings (Proyer and Brauer, 2018). Nevertheless, future studies should contain reverse coded items.

Of the 29 items in the APQ (Yu et al., 2003), 11 were excluded from the current analysis mainly due to high double loadings in the factor analysis. This could be due to issues with the translation and adaptation of the items or to cultural differences. In any case, the exclusion of such a large number of items limits the interpretation of the findings. However, it must be noted that the APQ facets seem more culturally bound than might be the case in other measures (cf. Proyer and Jehle, 2013). According to an article on the construction of APQ (Yu et al., 2003), published in a Taiwanese journal, "pleasantry" is a combination of "sense of humor" and "childlike manner," "initiative and concentration" means "flow because of intrinsic motivation," and "creativity" stands for "solving problems with creativity" (all translations by the first author), which are all essential parts of playfulness but appear more difficult to understand in the West than in the East. The pleasantry facet is closely related to the Western understanding of playfulness (correlation coefficients ranged from 0.43 to 0.69), whereas the other two facets seem to be more embedded in Eastern thinking; nevertheless, they show some overlap. This in itself may be of interest as it points to potential differences in the understanding of the trait across more distant cultures (e.g., individualistic vs. collectivistic) and more related countries (e.g., China and Taiwan).

Participants were asked to rate their playfulness (self/perceived society perspective) generally in different situations. However, the samples were all students for the sake of comparison. Hence, their experience of workplace situations was somewhat limited, as mentioned above. Answers to this question may, therefore, refer to imagined behaviors and rules at work. There are also some specific factors that could have an impact on a person's display of playfulness, such as the working atmosphere, the size of the company, the organizational culture, etc. Such variables have not been controlled for in this study, but may have had an impact. Additionally, there might also be shifts in the general perception of the roles that play and playfulness may have at the workplace (e.g., Petelczyc et al., 2017), and how this may permeate into different cultures. It seems more common today than in previous years to relate innovativeness and creativity to companies that foster and allow for play at the workplace; e.g., when thinking of labeling Google employees in Zurich as *Zooglers* and related newspaper headlines such as "Zooglers: Why staff are paid to play in Google's Zurich office" (The Guardian, 2018). Along with the other suggestions for future research, it would be interesting to study such changes from a longitudinal perspective and to analyze potential differences among age groups with varying exposure to Western culture from data collected in the East. As for humor, it has been argued that this has become more appreciated by people of all ages and different backgrounds (see Yue, 2014), and a similar transition may perhaps be expected for play and playfulness. Additionally, future studies should include working professionals for a further verification of cross-cultural differences and the contribution play and playfulness may have at work (Yu et al., 2007; Petelczyc et al., 2017). Instead of using only subjective instruments, some objective measurements could be added as well, such as uploading a picture of the work desk, which can be an indicator of playfulness.

We used only a single question for being playful in online situations. However, given the rise of social media and online communication, a more fine-grained analysis of such settings seems warranted, especially as the standards of living in China are growing and the entertainment sector is starting to flourish. Phenomena such as spoofing (恶搞; e'gao), which became more accepted in Chinese culture from early 2000 (for an overview see Yu, 2014), use parody, irony, and satire to mock those in power or make social comments. Moreover, other Internet media platforms and programs (such as PapiTube, U Can U Bibi, and Mars Intelligence Agency) enable collaboration and the production, circulation, and consumption of entertainment to be much faster, easier, and more convenient. Expressing one's

playfulness more privately on the Internet also seems to be in line with the Confucian tenets to express humor and play(fulness) in one's daily life.

Aside from what has been mentioned earlier, this study has several limitations. Firstly, the sample sizes are comparatively small and imbalanced with respect to certain demographics. Secondly, the *Brief Rating List of Playfulness in Different Situations* was developed for this study and further studies on its validity are needed. Additionally, it represents only a selected number of persons and situations that would be worth studying in the future. Thirdly, given the size of China, it would be desirable to have even more samples to represent regional differences. Fourthly, we have tested university students only and they are probably more diligent than the general population because the two Chinese samples were from top-tier universities, which means that they scored very high during their national entrance exams, while diligence also benefits students from the University of Zurich. This hinders the generalizability of the findings and future studies should consider controlling the results for diligence. Nevertheless, one might still assume that there will be differences across, for example, certain age groups (e.g., moderated by exposure to Western culture). Finally, measurement invariance was only established for the SMAP and the creativity facet of the APQ, while there was only partial measurement invariance for the other facets. Hence findings for the APQ must be interpreted with some reservations. We have already mentioned difficulties for the cross-cultural understanding of the pleasantry facet due to translation problems; consequently, more balanced measures from a cultural perspective will be needed in follow-up studies.

CONCLUSION

This study shows that it is of interest to study adult playfulness from a cross-cultural perspective (see also Barnett, 2017) and the findings have the potential to contribute toward a better understanding of the nature of playfulness. While the findings warrant replication, it seems safe to note that it would be fruitful to encourage further research on playfulness in Eastern countries (cf. Yu et al., 2003, 2007; Yue et al., 2016).

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of "the ethical guidelines of the ethics committee of the Faculty of Arts and Social Science, University of Zurich." Participants provided either online or written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

Both authors initiated the project, designed the concepts, and analyzed the data. DP collected the data. Both authors contributed to the writing of the manuscript, read it critically and gave consent to its publication.

ACKNOWLEDGMENTS

The authors thank Gu Li and Suqing Tang for their help with the back translation of the questionnaires and Chenna Yuang and Shan Jiang for their support in the data collection in China. The authors also would like to thank Alexander Stahlman, Barry Slaff and Sascha Kaelin for their comments and corrections on an earlier version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00421/full#supplementary-material

REFERENCES

Aroean, L. (2012). Friend or foe: in enjoying playfulness, do innovative consumers tend to switch brand? *J. Consum. Behav.* 11, 67–80. doi: 10.1002/cb.370

Aune, K. S., and Wong, N. C. (2002). Antecedents and consequences of adult play in romantic relationships. *Pers. Relat.* 9, 279–286. doi: 10.1111/1475-6811.00019

Barnett, L. A. (1990). Developmental benefits of play for children. *J. Leis. Res.* 22, 138–153. doi: 10.1080/00222216.1990.11969821

Barnett, L. A. (2007). The nature of playfulness in young adults. *Pers. Individ. Dif.* 43, 949–958. doi: 10.3390/children1030280

Barnett, L. A. (2012). Playful people: fun is in the mind of the beholder. *Imagin. Cogn. Pers.* 31, 169–197. doi: 10.2190/IC.31.3.c

Barnett, L. A. (2017). The inculcation of adult playfulness: from west to east. *Int. J. Play* 6, 255–271. doi: 10.1080/21594937.2017.1383010

Blasi, M., Hurwitz, S. C., and Hurwitz, S. C. (2002). For parents particularly: to be successful—let them play! *Child. Educ.* 79, 101–102. doi: 10.1080/00094056.2003.10522779

Bologh, R. W. (1976). On fooling around: a phenomenological analysis of playfulness. *Ann. Phenomenol. Sociol.* 1, 113–125.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *J. Cross Cult. Psychol.* 1, 185–216. doi: 10.1177/135910457000100301

Brown, S., and Vaughan, C. (2009). *Play: How it Shapes the Brain, Opens the Imagination, and Invigorates the Soul*. New York, NY: Penguin.

Chen, C., and Uttal, D. H. (1988). Cultural values, parents' beliefs, and children's achievement in the United States and China. *Hum. Dev.* 31, 351–358. doi: 10.1159/000276334

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Modeling* 14, 464–504. doi: 10.1080/10705510701301834

Cheung, G. W., and Rensvold, R. B. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *J. Manag.* 25, 1–27. doi: 10.1177/014920639902500101

Chey, J. (2014). "History of xiangsheng," in *Encyclopedia of Humor Studies*, Vol. 2, ed. S. Attardo (Los Angeles, CA: Sage), 811–813.

Chick, G. (2001). What is play for? Sexual selection and the evolution of play. *Play Cult.* 3, 3–25.

Chick, G., Yarnal, C., and Purrington, A. (2012). Play and mate preference: testing the signal theory of adult playfulness. *Am. J. Play* 4, 407–440.

Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *J. Humanist. Psychol.* 15, 41–63. doi: 10.1177/002216787501500306

Csikszentmihalyi, M., and LeFevre, J. (1989). Optimal experience in work and leisure. *J. Pers. Soc. Psychol.* 56, 815–822. doi: 10.1037/0022-3514.56.5.815

Davey, G., De Lian, C., and Higgins, L. (2007). The university entrance examination system in China. *J. Furth. High. Educ.* 31, 385–396. doi: 10.1080/03098770701625761

Erikson, E. H. (1950). *Childhood and Society*. New York, NY: W. W. Norton & Company.

Fekken, C. G., Holden, R. R., Jackson, D. N., and Guthrie, G. M. (1987). An evaluation of the validity of the personality research form with Filipino university students. *Int. J. Psychol.* 22, 399–407. doi: 10.1080/00207598708246781

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: the broaden-and-build theory of positive emotions. *Am. Psychol.* 56, 218–226. doi: 10.1037//0003-066X.56.3.218

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., et al. (2011). Differences between tight and loose cultures: a 33-nation study. *Science* 332, 1100–1104. doi: 10.1126/science.1197754

Ginsburg, K. R. (2007). The importance of play in promoting healthy child development and maintaining strong parent-child bonds. *Pediatrics* 119, 182–191. doi: 10.1542/peds.2006-2697

Glynn, M. A., and Webster, J. (1992). The adult playfulness scale: an initial assessment. *Psychol. Rep.* 71, 83–103. doi: 10.2466/pr0.1992.71.1.83

Glynn, M. A., and Webster, J. (1993). Refining the nomological net of the adult playfulness scale: personality, motivational, and attitudinal correlates for highly intelligent adults. *Psychol. Rep.* 72, 1023–1026. doi: 10.2466/pr0.1993.72.3.1023

Harrell, S. (1985). Why do the Chinese work so hard? Reflections on an entrepreneurial ethic. *Mod. China* 11, 203–226. doi: 10.1177/009770048501100203

Hasse, C. (2008). Learning and transition in a culture of playful physicists. *Eur. J. Psychol. Educ.* 23, 149–164. doi: 10.1007/BF03172742

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. Thousand Oaks, CA: Sage Publications.

Isenberg, J., and Quisenberry, N. L. (1988). Play: a necessity for all children. *Child. Educ.* 64, 138–145. doi: 10.1080/00094056.1988.10521522

Jackson, D. N. (1974). *Personality Research Form Manual*. Port Huron, MI: Research Psychologists Press.

Li, Y. (2006). 玩兴的研究及其对大学生创造力培养的启示 [Playfulness and its implications for training the creativity of college students]. *J. China Univ. Geosci.* 6, 77–80.

Liao, C. C. (2007). One aspect of Taiwanese and American sense of humour: attitudes toward pranks. *J. Hum. Res.* 2, 289–324.

Lieberman, J. N. (1977). *Playfulness: Its Relationship to Imagination and Creativity*. New York, NY: Academic Press.

Lin, H. (2012). 悬梁刺股 [Xuan liang ci gu]. *Quan Guo You Xiu Zuo Wen Xuan* 9, 33–33.

Lorenzo-Seva, U., and ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodol. Eur.* 2, 57–64. doi: 10.1027/1614-1881.2.2.57

Lunzer, E. A. (1959). Intellectual development in the play of young children. *Educ. Rev.* 11, 205–224. doi: 10.1080/0013191590110305

Magnuson, C. D., and Barnett, L. A. (2013). The playful advantage: how playfulness enhances coping with stress. *Leis. Sci.* 35, 129–144. doi: 10.1080/01490400.2013.761905

Markus, H. R., and Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychol. Rev.* 98, 224–253. doi: 10.1037/0033-295X.98.2.224

Milner Davis, J. (2014). "Bergson's theory of the comic," in *Encyclopedia of Humor Studies*, Vol. 1, ed. S. Attardo (Los Angeles, CA: Sage), 77–80.

Milner Davis, J., and Chey, J. (2011). *Humour in Chinese Life and Letters: Classical and Traditional Approaches*. Hong Kong: Hong Kong University Press.

Moskowitz, D. S., Suh, E. J., and Desaulniers, J. (1994). Situational influences on gender differences in agency and communion. *J. Pers. Soc. Psychol.* 66, 753–761. doi: 10.1037/0022-3514.66.4.753

Mullen, M. R. (1995). Diagnosing measurement equivalence in cross-national research. *J. Int. Bus. Stud.* 26, 573–573. doi: 10.1057/palgrave.jibs.8490187

Murray, H. A. (1938). *Explorations in Personality*. Oxford: Oxford University Press.

Ostendorf, F., Angleitner, A., and Ruch, W. (1986). *Die Multitrait-Multimethod Analyse. Konvergente und Diskriminante Validität der Personality Resesarch Form [The Multitrait-multimethod Analysis. Convergent and Discriminant Validity of the Personality Research Form]*. Göttingen: Hogrefe.

Petelczyc, C. A., Capezio, A., Wang, L., Restubog, S. L. D., and Aquino, K. (2017). Play at work: an integrative review and agenda for future research. *J. Manage.* 44, 161–190. doi: 10.1177/0149206317731519

Piaget, J. (1951). *Play, Dreams and Imitation in Childhood*. New York, NY: Norton.

Proyer, R. T. (2011). Being playful and smart? The relations of adult playfulness with psychometric and self-estimated intelligence and academic performance. *Learn. Individ. Differ.* 21, 463–467. doi: 10.1016/j.lindif.2011.02.003

Proyer, R. T. (2012a). A psycho-linguistic study on adult playfulness: its hierarchical structure and theoretical considerations. *J. Adult Dev.* 19, 141–149. doi: 10.1007/s10804-012-9141-7

Proyer, R. T. (2012b). Development and initial assessment of a short measure for adult playfulness: the SMAP. *Pers. Individ. Dif.* 53, 989–994. doi: 10.1016/j.paid.2012.07.018

Proyer, R. T. (2012c). Examining playfulness in adults: testing its correlates with personality, positive psychological functioning, goal aspirations, and multi-methodically assessed ingenuity. *Psychol. Test Assess. Model.* 54, 103–127.

Proyer, R. T. (2013). The well-being of playful adults: adult playfulness, subjective well-being, physical well-being, and the pursuit of enjoyable activities. *Eur. J. Humour Res.* 1, 84–98. doi: 10.7592/EJHR2013.1.1.proyer

Proyer, R. T. (2014a). Perceived functions of playfulness in adults: Does it mobilize you at work, rest, and when being with others? *Eur. Rev. Appl. Psychol.* 64, 241–250. doi: 10.1016/j.erap.2014.06.001

Proyer, R. T. (2014b). To love and play: testing the association of adult playfulness with the relationship personality and relationship satisfaction. *Curr. Psychol.* 33, 501–514. doi: 10.1007/s12144-014-9225-6

Proyer, R. T. (2017). A new structural model for the study of adult playfulness: assessment and exploration of an understudied individual differences variable. *Pers. Individ. Dif.* 108, 113–122. doi: 10.1016/j.paid.2016.12.011

Proyer, R. T., and Brauer, K. (2018). Exploring adult playfulness: examining the accuracy of personality judgments at zero-acquaintance and an LIWC analysis of textual information. *J. Res. Pers.* 73, 12–20. doi: 10.1016/j.jrp.2017.10.002

Proyer, R. T., and Jehle, N. (2013). The basic components of adult playfulness and their relation with personality: the hierarchical factor structure of seventeen instruments. *Pers. Individ. Dif.* 55, 811–816. doi: 10.1016/j.paid.2013.07.010

Proyer, R. T., and Ruch, W. (2011). The virtuousness of adult playfulness: the relation of playfulness with strengths of character. *Psychol. Well Being* 1, 1–12. doi: 10.1186/2211-1522-1-4

Proyer, R. T., and Wagner, L. (2015). Playfulness in adults revisited: the signal theory in German speakers. *Am. J. Play* 7, 201–227.

Proyer, R. T., Brauer, K., Wolf, A., and Chick, G. (in press a). Beyond the ludic lover: individual differences in self- and partner-rated playfulness and love styles in heterosexual relationships. *Am. J. Play*.

Proyer, R. T., Tandler, N., and Brauer, K. (in press b). "Playfulness and creativity: a selective review," in *Creativity and Humor*, eds S. R. Luria, J. Baer, and J. C. Kaufman (San Diego, CA: Academic Press).

Qian, X. L., and Yarnal, C. (2011). The role of playfulness in the leisure stress-coping process among emerging adults: an SEM analysis. *Leisure* 35, 191–209. doi: 10.1080/14927713.2011.578398

Realo, A., Allik, J., and Vadi, M. (1997). The hierarchical structure of collectivism. *J. Res. Pers.* 31, 93–116. doi: 10.1006/jrpe.1997.2170

Rosseel, Y. (2012). Lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Sagiv, L., Roccas, S., Cieciuch, J., and Schwartz, S. H. (2017). Personal values in human life. *Nat. Hum. Behav.* 1, 630–639. doi: 10.1038/s41562-017-0185-3

Schwartz, S. H. (2006). A theory of cultural value orientations: explication and applications. *Comp. Sociol.* 5, 137–182. doi: 10.1163/156913306778667357

semTools Contributors (2015). *semTools: Useful Tools for Structural Equation Modeling. R Package Version 0.4-9*. Available at: http://cran.r-project.org/package=semTools

Smith, A. H. (1894). *Chinese Characteristics*. New York, NY: Fleming H. Revell.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *J. Cross Cult. Psychol.* 35, 50–61. doi: 10.1177/0022022103260380

Suh, E., Diener, E., Oishi, S., and Triandis, H. C. (1998). The shifting basis of life satisfaction judgments across cultures: emotions versus norms. *J. Pers. Soc. Psychol.* 74, 482–493. doi: 10.1037/0022-3514.74.2.482

Sutton-Smith, B. (1967). The role of play in cognitive development. *Young Child.* 22, 361–370.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., et al. (2014). Large-scale psychological differences within china explained by rice versus wheat agriculture. *Science* 344, 603–608. doi: 10.1126/science.1246850

Tamis-LeMonda, C. S., Shannon, J. D., Cabrera, N. J., and Lamb, M. E. (2004). Fathers and mothers at play with their 2- and 3-year-olds: contributions to language and cognitive development. *Child Dev.* 75, 1806–1820. doi: 10.1111/j.1467-8624.2004.00818.x

The Guardian (2018). *Zooglers: Why Staff are Paid to Play in Google's Zurich Office*. Available at: https://www.theguardian.com/technology/shortcuts/2018/jan/15/zooglers-why-staff-are-paid-to-play-in-googles-zurich-office [accessed January 22, 2018].

Triandis, H. C., and Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *J. Pers. Soc. Psychol.* 74, 118–128. doi: 10.1037/0022-3514.74.1.118

Triandis, H. C., McCusker, C., and Hui, C. H. (1990). Multimethod probes of individualism and collectivism. *J. Pers. Soc. Psychol.* 59, 1006–1020. doi: 10.1037/0022-3514.59.5.1006

Tucker, L. R. (1951). *A Method for Synthesis of Factor Analysis Studies*. Personnel Research Section Report No. 984. Washington, DC: Department of the Army. doi: 10.21236/AD0047524

van de Vijver, F., and Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *Eur. Rev. Appl. Psychol.* 54, 119–135. doi: 10.1016/j.erap.2003.12.004

Yu, H. (2014). "E'gao: culture of internet spoofing in China," in *Encyclopedia of Humor Studies*, Vol. 1, ed. S. Attardo (Los Angeles, CA: Sage), 210–212.

Yu, P. (2004). 樂在工作–組織玩興研究 [Enjoy fun at work: an explorative study of organizational playfulness]. *Kaohsiung Norm. Univ. J.* 6, 19–37.

Yu, P., Wu, J., Lin, W., and Yang, G. (2003). 成人玩興量表與組織玩興氣氛量表之發展 [Development of adult playfulness and organizational playfulness climate scales]. *Psychol. Test.* 50, 73–110.

Yu, P., Wu, J. J., Chen, I. H., and Lin, Y. T. (2007). Is playfulness a benefit to work? Empirical evidence of professionals in Taiwan. *Int. J. Technol. Manag.* 39, 412–429. doi: 10.1504/IJTM.2007.013503

Yue, X. D. (2014). "History of humor: modern and contemporary China," in *Encyclopedia of Humor Studies*, Vol. 1, ed. S. Attardo (Los Angeles, CA: Sage), 302–304.

Yue, X. D., Leung, C. L., and Hiranandani, N. A. (2016). Adult playfulness, humor styles, and subjective happiness. *Psychol. Rep.* 119, 630–640. doi: 10.1177/0033294116662842

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Pang and Proyer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 27 November 2018
doi: 10.3389/fpsyg.2018.02255

Assessing the Temperamental Basis of the Sense of Humor: Adaptation of the English Language Version of the State-Trait Cheerfulness Inventory Long and Standard Form

Jennifer Hofmann[1]*, *Hugo Carretero-Dios*[2] and *Amy Carrell*[3]

[1] Department of Psychology, University of Zurich, Zurich, Switzerland, [2] Department of Research Methods in Behavioral Sciences, University of Granada, Granada, Spain, [3] Department of English, University of Central Oklahoma, Edmond, OK, United States

OPEN ACCESS

Edited by:
René T. Proyer,
Martin Luther University of Halle-Wittenberg, Germany

Reviewed by:
Danny Hinton,
University of Wolverhampton, United Kingdom
William Larry Ventis,
College of William and Mary, United States
Jason S. Wrench,
Suny New Paltz, United States

***Correspondence:**
Jennifer Hofmann
j.hofmann@psychologie.uzh.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: 19 July 2018
Accepted: 30 October 2018
Published: 27 November 2018

Citation:
Hofmann J, Carretero-Dios H and Carrell A (2018) Assessing the Temperamental Basis of the Sense of Humor: Adaptation of the English Language Version of the State-Trait Cheerfulness Inventory Long and Standard Form. Front. Psychol. 9:2255. doi: 10.3389/fpsyg.2018.02255

The State-Trait Model of Cheerfulness assesses the temperamental basis of the sense of humor with the traits and respective states of cheerfulness, seriousness, and bad mood. Cheerfulness is a dominant factor in current measures of the sense of humor and explains both, the disposition to engaging in smiling and laughter, as well as humor behaviors, and trait seriousness and bad mood are antagonistic to the elicitation of amusement (albeit for different reasons). Several studies have shown the validity and reliability of the STCI questionnaire in German and other language versions (i.e., Spanish). In this study, the English language version with 106 items (STCI-T<106>) was translated, checked for its item and scale characteristics, and tested with a confirmatory factor analysis approach ($N = 1101$) to investigate the factorial validity of the STCI-T<106> scale. Results show good psychometric characteristics, good internal consistencies, and a fit to the postulated underlying structure of the STCI-T. Then, the standard form with 60 items (STCI-T<60>) was developed and the psychometric characteristics initially tested. In an independent sample ($N = 169$), the characteristics of the standard form were compared to the parent form and German equivalent. It showed good psychometric characteristics, internal consistencies, as well as a good self- and peer-report congruence. To conclude, the STCI-T<106> is the measure of choice for the assessment of the temperamental basis of the sense of humor and the separate facets of the traits, while the standard form (60 items) allows of an economic assessment of cheerfulness, seriousness, and bad mood, free of context-saturated items and humor preferences.

Keywords: bad mood, cheerfulness, humor, sense of humor, seriousness, STCI

Ruch and colleagues found intra- as well as inter-individual differences in humor-related behaviors, thoughts, motivation, and responses (Ruch, 1993; Ruch et al., 1996, 1997; Ruch and Köhler, 2007). In particular, they found that individuals differ habitually in the likelihood of engaging in humor, the frequency and intensity of amusement responses, the quality and quantity of humor production, and the appreciation of humorous interactions (for an overview, see Ruch and Hofmann, 2012). High scorers of those habitual differences are usually nominated to have a good "sense of humor." When looking at individuals with a "good sense of humor" more closely, the authors also observed

that variations in the individuals' readiness to engage in humor occurred across situations and time, indicating that there might be humor-related *states* that enhance or lower ones threshold for amusement (for an overview see Ruch and Hofmann, 2012). To explain these inter- and intra-individual differences, Ruch et al. (1996) put forward a model defining the temperamental basis of the sense of humor: the State-Trait Model of Cheerfulness.

The State-Trait Model of Cheerfulness is a multidimensional model that assesses the affective and cognitive foundation of the sense of humor, assuming this foundation to have trait-like qualities. Therefore, the model is largely free of specific contents and preferences for certain humor materials (such as "dark humor" or "nonsense humor," certain comedians or comedy formats) but rather describes the underlying traits that predispose individuals to humor (or "humorlessness"). To account for intra-individual differences as well, the same concepts were used as traits and states, which allows for the study of mood states and their influence on humor elicitation (Ruch, 1993; Ruch et al., 1996, 1997; Ruch and Köhler, 2007; Ruch and Hofmann, 2012). The three relevant states and traits are: state and trait cheerfulness, state and trait seriousness and state and trait bad mood. While cheerfulness lowers the threshold for amusement, the latter two dimensions heighten the threshold for amusement (Ruch et al., 1996).

Cheerfulness goes along with a low threshold for amusement and liking to engage in humorous interactions (Ruch et al., 1996). Whereas both seriousness and bad mood may be perceived as forms of humorlessness (cf. McGhee, 1996), they do indeed heighten the threshold for amusement, but they are not the opposite of it. In fact, there are forms of humor that require certain degrees of seriousness (i.e., the ability to laugh at oneself, McGhee (1996) or "Ernstheiterkeit," see Proyer and Rodden, 2013) or bad mood (e.g., cynicism; see Ruch et al., 1996). In the case of seriousness, there is lowered interest in humor and playfulness; i.e., more volition is needed for individuals to switch into a playful frame of mind or engage humorously. In the case of bad mood, negative affective states are predominant and hinder the elicitation of amusement. Within trait bad mood reasons for humorlessness differ as well: Whereas ill-humornedness may also lead to not wanting to engage in humor, sadness may lead to an inability to engage in humor (e.g., Ruch and Hofmann, 2012).

MEASUREMENT OF THE STATE-TRAIT CHEERFULNESS MODEL

The aim was to create a questionnaire largely free of content saturated items and specific humor materials/stimuli categories (Ruch et al., 1996, 1997), as this lowers the generalizability and adequate use of the scale (i.e., item judgments might be biased by the fact that certain age or cultural groups do or do not know certain materials or contents; a former criticism toward many questionnaires attempting to assess the "sense of humor," cf. Ruch, 2007). The original German trait long form the State-Trait Cheerfulness Inventory (STCI-T<106>) provides scores for the three traits of cheerfulness (STCI-T CH; 38 items), seriousness (STCI-T SE; 37 items), and bad mood (STCI-T BM; 31 items), as well as a separate analysis of the facets of each trait. For cheerfulness, five inter-correlated facets were derived: a prevalence of cheerful mood (CH1), a low threshold for smiling and laughter (CH2), a composed view of adverse life circumstances (CH3), a broad range of active elicitors of cheerfulness and smiling or laughter (CH4), and a generally cheerful interaction style (CH5; Ruch et al., 1996). Trait bad mood (BM) is composed of the predominance of three mood states and their respective behaviors: generally being in a bad mood (BM1), sadness (i.e., despondent and distressed mood; BM2), and ill-humoredness (i.e., sullen and grumpy or grouchy feelings; BM4). Two further facets are specifically related to the sad (BM3) and ill-humored (BM5) individual's behavior in cheerfulness evoking situations, their attitudes toward such situations and the objects, persons, and roles involved (Ruch et al., 1996). Trait seriousness consists of six inter-correlated facets: the prevalence of serious states (SE1), a perception of even everyday happenings as important and considering them thoroughly and intensively (SE2), the tendency to plan ahead and set long-range goals (SE3), the tendency to prefer activities for which concrete, rational reasons can be produced (SE4), the preference for a sober, object-oriented communication style (SE5), and a "humorless" attitude about cheerfulness-related behavior, roles, persons, stimuli, situations, and actions (SE6; see Ruch et al., 1996).

The STCI-T<106> long form was constructed for three reasons: to provide an assessment of the facets, to be able to test hypotheses that link to the facets, as well as to empirically evaluate the facet model. A rational-theoretical construction strategy was applied, with the facet model serving as the basis for the generation of items (see Ruch et al., 1996). The 106 items were chosen from a pool and designed to be (1) short and understandable, (2) of diverse content, (3) covering the construct-related behavior and attitudes comprehensively, (4) be free of extreme levels of social desirability, (5) suitable for adolescents and adults, and (6) not biased toward particular populations and (7) the items needed to be logically related to the target constructs but not overlap with similar but irrelevant constructs (Ruch et al., 1996). The questionnaire utilizes a four-point answering format.

Because of the antithetical nature of the three traits (i.e., cheerfulness denominating a lowered threshold for amusement, whereas seriousness and bad mood go along with a heightened threshold for amusement), most items were positively poled, as negatively poled items could be viewed as a (positive) indicator for another trait (Ruch et al., 1996, 1997). Ruch et al. (1996) confirmed the facet model by means of factor analyses and reported mostly satisfactory to high internal consistencies for the facets ($\alpha = 0.64$ to $\alpha = 0.91$) and high internal consistencies for the trait total scores (CH $\alpha = 0.93$, SE $\alpha = 0.91$, BM $\alpha = 0.93$).

To arrive at a questionnaire that is more economic for the use in research and practice, a standard trait form with 60 items was derived from the STCI-T<106>. This version is not considered for scoring facets but only total scores for the three traits of cheerfulness, seriousness, and bad mood are derived. In general, a concept-guided strategy in item reduction was coupled with an empirically guided selection of items (Ruch et al., 1996, 1997).

The following criteria were considered (Ruch et al., p. 316): (a) the best corrected item to total correlation (CITC), (b) consideration of items content and representation of items of all facets, (c) roughly equal representation of the facets (where this was not possible, core facets got more weight), and (d) avoidance of very similar items as regards to content or linguistic usage (Ruch et al., 1996, p. 316). Ruch and Köhler (2007) reported high internal consistencies for the traits (CH $\alpha = 0.93$, SE $\alpha = 0.88$, and BM $\alpha = 0.94$; $N = 600$) and the one-month retest-stability was high for the traits (between 0.77 and 0.86), in line with the expectations (Ruch et al., 1997). The three-factor structure was replicable and showed to be generalizable across samples of different nationalities and language groups.

The *state* version of the STCI initially consisted of 45 items, cheerfulness, seriousness, and bad mood as actual feeling states. To capture the mood quality, items were included that allow for a sensitive assessment of mood alternations (Ruch et al., 1997). Different facets of cheerfulness as mood states were distinguished (Ruch et al., 1997): A cheerful mood (tranquil, composed) and hilarity (more shallow, outward; Ruch et al., 1997). In state seriousness, soberness, pensiveness, and earnestness were differentiated and in state bad mood, melancholy and ill-humor were distinguished. There is also a four-point answer format, like in the trait version (Ruch et al., 1997). In an iterative process spanning over several samples, the scale was finally reduced to consist of ten items for each scale. Ruch et al. (1997) report satisfactory internal consistencies (alpha coefficients from $\alpha = 0.85$ to $\alpha = 0.94$) and low test-retest correlation in line with the expectations.

Language Adaptations, Peer and Childrens' Versions

Since the first publication of the STCI-T in 1996 and the STCI-S in 1997 (Ruch et al., 1996, 1997) several translations into different languages and adaptations to other target groups (other than self-reports) were done. **Table 1** gives an overview on the available versions (adapted from Ruch and Hofmann, 2012).

As **Table 1** shows, the STCI exists in over ten languages and can be applied in various settings, with various versions for self- and peer-reports (e.g., general peer-report, peer-report for parents, peer-report for the workplace; see **Table 1**). The instruments typically yielded comparable psychometric characteristics and correlational patterns (see Ruch and Hofmann, 2012).

Correlations Among the Traits and States

Cheerfulness and bad mood are affective concepts with an antagonistic valence, supposedly leading to a negative correlation between the two. Seriousness is also a factor increasing the threshold for humor, though not on an affective level, but on the level of cognition: Seriousness refers to a frame of mind (cf. Ruch et al., 1996). Thus, correlations of seriousness to cheerfulness should be negative, but weaker as compared to bad mood, as the latter is conceptually closer, as it also refers to an affective concept. Seriousness and bad mood should be correlated positively, as they both refer to concepts potentially hindering the induction of amusement or the engagement with humor (cf. Ruch et al., 1996). For both, the STCI-T<106>, as well as the STCI-T<60> and the standard form of the state STCI-S<30> questionnaire, the results showed that homologous states and traits are separable and correlations between the converging states and traits were expectedly positive, correlations with heterologous concepts were lower (for an overview, see Ruch and Hofmann, 2012). Cheerfulness in state and trait was negatively related to trait and state seriousness and trait and state bad mood (and the latter two were positively correlated themselves). In line with the expectations, correlations among the three traits were numerically lower than among the three states (e.g., Ruch and Hofmann, 2012).

VALIDATION

Trait

With respect to the *factorial* validity of the trait form, factor analyses of the facet model of the STCI-T<106> trait version supported the model by Ruch et al. (1996) with three correlated higher order factors and their five to six facets in the German version of the questionnaires, as well as the Spanish version with 104 items (Carretero-Dios et al., 2011, 2014). For the STCI-T<60> and STCI-S<30>, typically a three-factor structure could be confirmed (e.g., Ruch et al., 1996, 1997; Tapia-Villanueva et al., 2014; Chen et al., 2017).

With respect to the *convergent* and *discriminant* validity, Carretero-Dios et al. (2011) applied a multi-trait multi-method method approach (MTMM) to data of the STCI. The MTMM approach allows for the separation of different sources of individual differences (influences due to trait, method, error components). By means of confirmatory factor analysis, the convergent validity (self-reports, peer-reports, aggregated states) and discriminant validity (relationships among cheerfulness, seriousness, and bad mood) of the trait form of STCI-T<104> were tested (the Spanish version contains 104, not 106 items; see Carretero-Dios et al., 2011) and confirmed: cheerfulness, seriousness and bad mood, as both state and traits are homogeneous factors (across self reports and peer-reports). Also, aggregated states measures were correlated to their traits and these correlations were higher than for single state measures, in line with the expectations (Carretero-Dios et al., 2011). Finally, the expected patterns of correlations between the three dimensions were confirmed and the data provided support for the hypothesis that traits represent the dispositions for their respective states. Furthermore, **Table 2** shows correlations of the STCI cheerfulness scale to relevant measures assessing aspects of the sense of humor (convergent validity; i.e., Ruch et al., 2011) as well as indicators of *predictive* validity (e.g., Ruch, 1997, **Table 2** is adapted and updated from Ruch and Hofmann, 2012).

To summarize, **Table 2** shows that trait cheerfulness correlates positively to convergent measures of the sense of humor (e.g., Ruch et al., 2011). For example, trait cheerfulness correlates positively to *coping humor* (measured by the Situational Humor Response Questionnaire, SHRQ, Martin and Lefcourt, 1984; or the Coping Humor Scale, CHS, Martin and Lefcourt, 1983), *humor styles* (measures by the Humor Styles Questionnaire, HSQ, Martin et al., 2003), the *facets of the sense of humor* (Sense of

TABLE 1 | Overview of the different versions of the STCI-T and STCI-S (adapted from Ruch and Hofmann, 2012).

Version	Target	Facet structure	Language
TRAIT			
STHI-T <106>	Self, peer	5 facets for cheerfulness (38 items), 6 facets for seriousness (37 items), 5 facets for bad mood (31 items)	German, English
STHI-T <104>		5 facets for cheerfulness (38 items), 6 facets for seriousness (37 items), 5 facets for bad moods (29 items)	Spanish
STHI-T <60>	Self, peer, workplace	1 score for each trait (20 items each)	Chinese (Hong Kong; Mainland China), English, French (Québec), German, Hebrew, Italian, Japanese Polish, Romanian, Russian, Slovene, Spanish (Chile)
STHI-T <30>	Self, peer	1 score for each trait (10 items each)	German, English, Italian
STHI-T <30> children	Self, peer, parent, teacher	1 score for each trait (10 items each); items adapted for children	German, Spanish
STATE			
STHI-S <45>	Self		German, Spanish[a]
STHI-S <30>	Self	1 score for each state (10 items each)	German, English
STHI-S <20>	Self	1 score for each state (8 items for cheerfulness, 6 seriousness, 6 bad mood)	English
STHI-S <18>	Self	1 score for each state (6 items each)	German, English, Hebrew
STHI-S <20> children	Self, peer	1 score each state (8 items for cheerfulness, 6 seriousness, 6 bad mood)	German

Further information on the different versions and authors involved in translation and adaptation can be obtained from the authors. [a]López-Benítez et al. (2017c).

Humor Scale, SHS, McGhee, 1996), and *styles of everyday humor conduct* (e.g., Humorous Behavior Q-Sort Deck, HBQD, Craik et al., 1996; and the HUMOR, Manke, 2007). With respect to the *predictive validity*, a range of studies have shown the power of trait and state cheerfulness in the prediction of responses to humor and amusement eliciting stimuli (see **Table 2**). **Table 2** shows the influence of state and trait cheerfulness on the experimental induction of amusement and external criteria (i.e., pain tolerance, see Zweyer et al., 2004; for a more detailed discussion, please see the original sources named in **Table 2**).

For seriousness and bad mood, sound correlations to convergent and discriminant measures could be established too. For example, trait bad mood and trait seriousness go along with gelotophobia (Ruch et al., 2009), less socially warm humor, and less competent humor in the HBQD, and less affiliative, self-enhancing humor in the HSQ (see Ruch et al., 2011). Moreover, trait bad mood with using less benevolent humor (that bases on a non-judgmental, cheerful outlook on the world; Hofmann et al., in press), and trait seriousness with less use of aggressive humor in the HSQ (see Ruch et al., 2011), as well as correlating negatively to most global assessments of playfulness, as well as playfulness facets (Proyer and Rodden, 2013).

Also for trait seriousness and bad mood, results on predictive validity and further aspects are in line with the expectations. For example, low trait seriousness was found to predict greater substance use, indicating that taking a less serious outlook on life also goes along with a more liberal attitude with respect to substance use (and maybe health related habits in general, see Edwards, 2012). Moreover, trait cheerfulness correlated positively with the wittiness of punch lines in a humor production task, whereas trait bad mood correlated negatively (Ruch et al., 2009). Interestingly, the numerically highest correlations were found for traits seriousness, both with the quality, as well as the quantity of humor production (with negative correlations see Ruch et al., 2009). Therefore, low seriousness predicts quantitative and qualitative aspects of humor production. With respect to positive emotional responses to situations where laughing at oneself is possible, trait cheerfulness predicted greater frequency and intensity of smiles in response to ones funnily distorted photo, whereas negative correlations of the frequency and intensity of smiling to trait bad mood were found (see Beermann and Ruch, 2011).

Looking at general evaluations of one's life (well-being, stress) as well as the dispositions of resilience and optimism, results of a recent study showed that trait cheerfulness correlates positively with resilience, optimism and well-being, whereas being negatively correlated to stress. For trait bad mood, the exact opposite pattern was found. For trait seriousness, a positive correlation to resilience was reported (Lau et al., 2018).

State

A range of studies has undertaken the assessment of the three states of cheerfulness, seriousness and bad mood including mood changes due to natural phenomena (e.g., weather), experimental variations (e.g., experimenter personality or social behavior) and chemical substances (i.e., inhalation of "laughing gas"

TABLE 2 | State and trait cheerfulness and the experimental induction of amusement and cheerful mood (adapted from Ruch and Hofmann, 2012).

Individuals high in trait cheerfulness (compared to individuals low in trait cheerfulness)
…laugh more often and have higher increases in state cheerfulness after inhaling nitrous oxide (Ruch and Stevens, 1995)
…stay in a cheerful mood when having to elaborate proverbs with negative, misanthropic contents (Wancke, 1996)
…show facial signs of exhilaration more frequent and intense, when interacting with a clowning experimenter for 10 min (Ruch, 1997)
…have higher increases in cheerfulness after listening to funny tapes (in comparisons to tapes containing neutral contents; Ruch, 1997)
…keep a cheerful state, even when having to sit in a depressing room while working on several tasks (Ruch and Köhler, 1999)
…show more smiling and laughter (higher contraction of the zygomatic major muscle) when looking at video clips of simple news or news speaker's slips of the tongues (Beyler, 1999)
…report higher state cheerfulness, and no more physical symptoms, even when facing negative life events and stress (Hausser, 1999; Ruch and Köhler, 1999; Ruch and Zweyer, 2001)
…report using humor as a coping strategy (Ruch and Zweyer, 2001)
…have a higher pain tolerance (in the cold pressure test) after watching a funny film and producing humor to it, or smiling and laughing voluntarily at it (Zweyer et al., 2004)
…have higher rises in state cheerfulness after consuming kava extract (Thompson et al., 2004)
…report more emotional intelligence (Yip and Martin, 2006*)
…display BOLD activation in the inferior parietal lobule of the right hemisphere. This might be associated with a general readiness/tendency to be amused by jokes. Regions previously shown to be activated in humor appreciation studies seem more likely to be related to the understanding of individual jokes and the momentary emotion and the momentary emotional reaction of exhilaration (Rapp et al., 2008)
…score higher on all of the sense of humor facets measured by the SHS (Ruch and Carrell, 1997*; Ruch et al., 2009)
…report less fear of being laughed at (gelotophobia, Ruch et al., 2009)
…report less need for structure (Hodson et al., 2010)
…are higher in socially warm, competent, earthy humor of the HBQD (Ruch et al., 2011)
…report more humor behavior (Ruch et al., 2011)
…show more Duchenne smiling in response to seeing distorted photographs of themselves (Beermann and Ruch, 2011)
…are higher in affiliative and self-enhancing humor styles, report less self-defeating humor of the HSQ (Martin et al., 2003; Ruch et al., 2011)
…experience an increase in state cheerfulness and show more facial displays of joy when watching funny videos alone or with a virtual companion (Hofmann et al., 2015)
…respond with more positive emotions to a clinic clowning intervention (Auerbach, 2017)
…are more sensitive to the emotional environment (López-Benítez et al., 2017a,b)
…report more resilience, mindfulness, optimism, well-being and less stress (Lau et al., 2018; Hofmann et al., in press)

*Studies are presented ordered by date of publication. *These studies used the pilot version of the English STCI basing on its initial translation from German.*

or "kava-kava" extract; see Ruch and Hofmann, 2012 for an overview). The results showed that one's current mood indeed alters the threshold for amusement and the manipulation of mood states might heighten or lower this threshold. For example, cheerfulness decreased after being exposed to situations inducing bad mood and was high when assessing female visitors of a carnival event (Ruch et al., 1997; Ruch and Köhler, 1999). Seriousness increased when being confronted with a 2 h mental work task, when listening to audiotapes of a serious (but also bad mood) quality and decreased in some cheerful situations, such as carnival and due to humor trainings (e.g., Falkenberg et al., 2011a,b; Ruch et al., 2018). Bad mood increased when being exposed to an adverse room environment (Ruch, 1997) and decreased after watching funny films (Hofmann et al., 2015), the inhalation of nitrous oxide (Ruch and Stevens, 1995), and sessions of humor trainings (Ruch et al., 2018). Furthermore, it was shown that that the STCI-S is a sensitive instrument for assessing longer lasting states too: As expected, depressive patients were shown to be lower in state cheerfulness and higher in state seriousness and state bad mood in comparison to the construction sample, and similarly for schizophrenic patients compared to the construction sample (Krantzhoff and Hirsch, 2001; Hirsch et al., 2010; Falkenberg et al., 2011a; Ruch et al., 2011; on depressed patients and on schizophrenic patients by Falkenberg et al., 2007).

Recently, the state-trait cheerfulness influence on self-reported disease activity levels in rheumatoid arthritis patients (Delgado-Domínguez et al., 2016) was investigated in a cross-sectional study. State cheerfulness and trait cheerfulness were assessed at the same time as a blood sample was taken from patients in order to analyze the corresponding biochemical parameters (Erythrocyte sedimentation rate and C-reactive protein), and just before measuring patient-reported disease activity. Higher state cheerfulness was observed in rheumatoid arthritis patients with lower scores in self-reported disease activity. Moreover, higher state cheerfulness was associated with lower values of C-reactive protein. Finally, results showed that the relationship between the biochemical parameters of rheumatoid arthritis and patient-reported disease activity partially depended (i.e., mediation analysis) on cheerful mood at the moment of assessment (Delgado-Domínguez et al., 2016).

Aims of the Current Study

Although the STCI-T questionnaire has been used in the original German language versions and has been adapted to different languages (e.g., Carretero-Dios et al., 2014; Chen et al., 2017), an English language version, both in the long trait form with 106 items and an economic version with 60 items has not been tested and validated for research and

practice. Therefore, the aim of the current study was 2-fold: Firstly, a long form with 106 items was translated, adapted, and initially validated. Secondly, the more economic short form with 60 items was adapted and initially tested (as well as being tested in an independent sample, including self-and peer-reports).

METHODS

Participants

Construction Sample

The sample consisted of 1,101 English speaking adults (36.2% men, 56.1% women and 7.7% indicating no gender) aged from 15 to 70 years ($M = 24.85$, $SD = 10.11$) from four different universities.

Replication Sample

The sample consisted of 85 English speaking adults (71.1% female, 24.4% male, and 4.4% not indicating their gender), age from 18 to 78 years ($M = 43.05$, $SD = 14.33$).

Peer-Report Sample

For the *Replication Sample*, a sample of peer-raters was collected, consisting of 84 individuals (69% female, 19% male, 12% not indicting their gender), with ages ranging from 18 to 67 ($M = 53.82$, $SD = 14.68$). On average, the peer-raters spent $M = 41.40$ h with the person they had rated and they indicated that they were very familiar with the rated person ($M = 6.40$, $SD = 1.01$, Min = 4.00, Max = 7.00; scale ranging from 1 to 7).

Instruments

STCI-T<106>

Cheerfulness (CH), seriousness (SE), and bad mood (BM) were assessed by the English language version of the *State-Trait-Cheerfulness-Inventory* (STCI; Ruch et al., 1996). The facet version of the STCI-T with 106 items was utilized to measure the three respective traits (and their respective facets) on a four-point scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*. "I am often in a joyous mood" is an indicator for CH, "I am a rather sad person" an indicator for BM, and "one of my principles is: first work, then play" an indicator for SE. Because of the antithetical nature of the concepts a negatively keyed cheerfulness item, for example, could also be seen prototypical for seriousness or bad mood. Whereas the sentence "I feel like laughing" might indicate cheerfulness, its negation "I don't feel like laughing" might well indicate sadness. Therefore, negations were only used when they represented standing expressions used in everyday language (cf. Ruch et al., 1996).

STCI-T<60>

Cheerfulness (CH), seriousness (SE), and bad mood (BM) were assessed by the *State-Trait-Cheerfulness-Inventory* standard short form and peer-report form (Ruch et al., 1996). The STCI-T<60> self- and peer-report measure the respective traits (60 items) on a four-point scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*.

Procedure

Translation Procedure

In step 1, all 106 items were translated into English by two persons (experts of humor) independently. Step 2 included a comparison of both translations, discussions about linguistic peculiarities and the intent of several items, and ended in a first list of suitable translations (coordinated by the senior author of the scale; see also Ruch and Carrell, 1997). In step 3 this list was sent to two American researchers (experts of humor) who checked it for orthographical and/or grammatical errors. Their corrections were checked for correspondence regarding the items' content and retained to a large extent. In step 4, this modified list was discussed by further two American researchers and the senior author of the original scale (WR), all familiar with the State-Trait Model of Cheerfulness. This resulted in the pilot version that was used for the current study. This procedure ensured a high level of expertise in humor research and sensitivity of the translators for the challenges of measuring humor (i.e., use of negations on an item level may make items indicative of other traits than the target trait, etc.).

Participant Recruitment Construction Sample

Participants were recruited over various channels at four universities and were given a paper-pencil version of the STCI-T<106>. They returned it after completion at home or testing in the class room. After handing in the completed questionnaire, participants were thanked for their participation.

Participant Recruitment Replication Sample and Peer-Report Sample

Participants were recruited over various channels, including universities and were given a paper-pencil version of the STCI-T<60>. They returned it after completion at home or testing in the class room. All participants were encouraged to give a peer-version of the STCI-T<60> to a good friend or relative. After handing in the completed questionnaire (self-report) participants were thanked for their participation. Peer reports could be sent back via post.

All procedures complied with the ethical guidelines of the local ethics committee at the University of Zurich, Faculty of Philosophy, Department of Psychology. All participants took part in the study voluntarily and could refrain from participation at any time without any consequences to them and consent was obtained by virtue of survey completion. The anonymity of participants was ensured. An ethics approval as per institutional and national guidelines was not required.

RESULTS

Psychometric Characteristics of the STCI-T<106> in English

Means, standard deviations, skewness, and kurtosis of the facets and total scores are given in **Table 3** (all analyses conducted in SPSS 25). Also, the internal consistencies (Cronbach Alpha; α), mean CITC for each facet and the total scores, as well as means,

TABLE 3 | Descriptive statistics, corrected item to total correlation and reliability of the scales and facets of the STCI-T<106> in the construction sample.

									German construction sample				
	Ni	*M*	*Md*	*SD*	*Sk*	*K*	*M* CITC	α	α	*M*	*SD*	*M* CITC	*t (1100)*
CH1	8.00	24.09	24.00	4.68	−0.45	0.01	0.66	0.88	0.91	24.67	4.69	0.72	−4.12*
CH2	5.00	16.51	17.00	2.85	−0.66	0.01	0.55	0.76	0.76	15.73	2.77	0.54	9.04*
CH3	8.00	22.57	23.00	3.24	−0.03	0.15	0.34	0.65	0.76	23.96	3.84	0.46	−14.27*
CH4	8.00	26.14	26.00	3.55	−0.32	−0.21	0.43	0.72	0.68	24.29	3.74	0.38	17.35*
CH5	9.00	29.84	30.00	4.01	−0.44	−0.06	0.52	0.82	0.84	28.54	4.21	0.54	10.73*
CH Total	38.00	119.15	120.00	15.69	−0.33	−0.08	0.50	0.89	0.93	117.2	15.9	0.52	4.11*
SE1	6.00	14.21	14.00	2.84	−0.03	0.08	0.32	0.57	0.65	15.27	3.05	0.39	−12.37*
SE2	7.00	17.42	17.00	3.47	0.03	−0.07	0.42	0.70	0.75	18.56	3.73	0.47	−10.91*
SE3	7.00	19.05	19.00	3.84	−0.15	−0.14	0.48	0.76	0.76	18.72	3.98	0.48	2.86
SE4	5.00	11.53	12.00	2.49	0.18	0.02	0.29	0.52	0.64	12.43	2.93	0.41	−12.06*
SE5	6.00	14.49	14.00	2.62	−0.02	0.11	0.26	0.51	0.70	14.65	3.4	0.44	−0.91
SE6	6.00	10.53	10.00	3.20	0.59	−0.11	0.48	0.75	0.74	12.95	3.69	0.48	−25.07*
SE total	37.00	87.22	87.00	13.64	0.03	0.01	0.38	0.87	0.91	92.58	15.79	0.44	−13.03*
BM1	6.00	12.19	12.00	3.35	0.48	−0.02	0.48	0.74	0.78	12.21	3.38	0.53	−0.17
BM2	8.00	16.76	16.00	4.73	0.50	−0.30	0.55	0.82	0.85	17.56	4.89	0.59	−5.61*
BM3	5.00	8.72	8.00	2.95	0.79	0.25	0.53	0.76	0.73	9.91	2.78	0.49	−13.35*
BM4	7.00	14.58	14.00	4.14	0.43	−0.06	0.55	0.81	0.81	14.71	4.04	0.55	−1.08
BM5	5.00	8.83	8.00	2.95	0.71	0.07	0.49	0.73	0.71	9.49	2.88	0.47	−7.38*
BM total	31.00	61.09	59.00	15.83	0.60	0.03	0.52	0.94	0.93	63.89	15.35	0.56	−5.88*

N = 1101. Ni, number of items per facet or scale; Sk, skewness; Ku, kurtosis; α, Cronbach Alpha; CITC, corrected item to total correlation. t (1100) = mean comparison between construction sample (German language version) and current sample. p < 0.01 (Bonferroni corrected).

standard deviations, Cronbach Alpha and mean CITC of the original German scale are reported in **Table 3**.

The internal consistencies for the total scores of cheerfulness, seriousness, and bad mood were high (CH $\alpha = 0.89$; SE $\alpha = 0.87$; BM $\alpha = 0.94$) and comparable to the internal consistencies of the German version of the STCI-T<106> and an American sample that had completed an English pilot version (CH $\alpha = 0.93$; SE $\alpha = 0.89$; BM $\alpha = 0.92$; see Ruch and Carrell, 1997). With respect to the facets, the five facets of trait cheerfulness all yielded satisfactory reliabilities ranging from $\alpha = 0.72$ to $\alpha = 0.88$, apart from facet CH3, with an $\alpha = 0.65$, see **Table 3**. Looking at the facets of trait seriousness, the internal consistencies ranged from $\alpha = 0.51$ to $\alpha = 0.76$. The facets of bad mood all reached satisfactory internal consistencies, between $\alpha = 0.73$ and $\alpha = 0.82$. Overall, the internal consistencies of the facets were highly comparable to the scores reported for the German version of the STCI-T<106> (see Ruch et al., 1996), apart from the facet SE5, which yielded a lower α in the English version ($\alpha = 0.51$ compared to $\alpha = 0.70$ in the German version), see **Table 3**. The skewness values (ranging between −0.66 to 0.79) were numerically comparable to the German language version questionnaire and the kurtosis values (ranging −0.30 to 0.25) were numerically slightly lower as compared to the German language version.

When looking at the items of the facets of the three traits, the corrected item to total correlations (CITC) were generally satisfactory to high. For the facets of trait cheerfulness, the mean CITC for the items of a facet ranged between 0.34 to 0.66 (CH1: $r_m = 0.66$; CH2: $r_m = 0.55$; CH3: $r_m = 0.34$; CH4: $r_m = 0.43$; CH5: $r_m = 0.52$; see **Table 3**), with six items having a low CITC (i.e.,>0.30). For the facets of trait seriousness, the mean CITC for the items of a facet ranged between 0.26 to 0.48 (SE1: $r_m = 0.32$; SE2: $r_m = 0.42$; SE3: $r_m = 0.48$; SE4: $r_m = 0.29$; SE5: $r_m = 0.26$; SE6: $r_m = 0.48$), with six items not reaching a minimal CITC of > 0.30. Lastly, with respect to the facets of trait bad mood, the mean CITC for the items of a facet ranged between 0.48 to 0.55 (BM1: $r_m = 0.48$; BM2: $r_m = 0.55$; BM3: $r_m = 0.53$; BM4: $r_m = 0.55$; BM5: $r_m = 0.49$), with one item not reaching a minimal CITC of > 0.30. Overall, the mean CITC were highly comparable to the ones reported for the German version of the STCI-T<106>, apart from a lower score for SE4 in the English language version, see **Table 3**.

Mean Comparisons and Correlations

Next, we compared the means of the current English version to the means of the German construction sample (*t*-tests). **Table 3** shows that the means generally differed between the two language versions in nearly all of the facets. With respect to the total scores, the means in the English language version indicated that the sample reported to higher scores in trait cheerfulness, less trait seriousness and less trait bad mood. Moreover, we computed Pearson correlations between the total scores and the facets of the three traits. The correlations can be seen in **Table 4**.

As expected, the numerically highest correlations were found for the facets of each respective trait with the total score (see **Table 4**). Also, correlations of homologous scales were higher than correlations to heterologous scales. In line with former findings, cheerfulness correlated negatively with seriousness and

TABLE 4 | Correlations of the facets and the total scores of the STCI–T<106> in the construction sample.

	CH total	CH1	CH2	CH3	CH4	CH5	SE total	SE1	SE2	SE3	SE4	SE5	SE6	BM total	BM1	BM2	BM3	BM4	BM5
CH total	1	0.89***	0.85***	0.79***	0.84***	0.89***	−0.46***	−0.40***	−0.37***	−0.17***	−0.29***	−0.19***	−0.62***	−0.73***	−0.64***	−0.60***	−0.72***	−0.61***	−0.65***
CH1		1	0.72***	0.67***	0.62***	0.71***	−0.36***	−0.33***	−0.32***	−0.13***	−0.19***	−0.13***	−0.48***	−0.71***	−0.65***	−0.60***	−0.66***	−0.60***	−0.58***
CH2			1	0.56***	0.64***	0.74***	−0.37***	−0.29***	−0.31***	−0.13***	−0.24***	−0.15***	−0.52***	−0.59***	−0.52***	−0.44***	−0.61***	−0.48***	−0.57***
CH3				1	0.59***	0.59***	−0.33***	−0.31***	−0.34***	−0.16***	−0.16***	−0.08	−0.39***	−0.57***	−0.50***	−0.52***	−0.54***	−0.48***	−0.46***
CH4					1	0.75***	−0.47***	−0.39***	−0.32***	−0.18***	−0.36***	−0.26***	−0.62***	−0.58***	−0.49***	−0.47***	−0.62***	−0.46***	−0.55***
CH5						1	−0.44***	−0.39***	−0.31***	−0.13***	−0.29***	−0.19***	−0.64***	−0.63***	−0.53***	−0.49***	−0.65***	−0.53***	−0.61***
SE total							1	0.75***	0.80***	0.75***	0.76***	0.65***	0.71***	0.54***	0.42***	0.42***	0.49***	0.50***	0.55***
SE1								1	0.55***	0.46***	0.45***	0.35***	0.53***	0.47***	0.39***	0.40***	0.41***	0.43***	0.44***
SE2									1	0.55***	0.51***	0.44***	0.44***	0.50***	0.42***	0.43***	0.40***	0.47***	0.45***
SE3										1	0.51***	0.34***	0.30***	0.20***	0.14***	0.17***	0.18***	0.18***	0.21***
SE4											1	0.47***	0.52***	0.32***	0.23***	0.21***	0.33***	0.29***	0.39***
SE5												1	0.38***	0.28***	0.21***	0.20***	0.25***	0.28***	0.30***
SE6													1	0.63***	0.49***	0.45***	0.62***	0.56***	0.67***
BM total														1	0.88***	0.88***	0.87***	0.90***	0.82***
BM1															1	0.72***	0.70***	0.77***	0.66***
BM2																1	0.71***	0.72***	0.59***
BM3																	1	0.70***	0.73***
BM4																		1	0.71***
BM5																			1

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

trait bad mood, with the latter two being positively correlated. Overall, the correlations replicated the patterns reported for the German and Spanish language version of the STCI-T (see Ruch et al., 1996; Carretero-Dios et al., 2014).

Testing the Underlying Structure of the STCI-T<106>

Six alternative models regarding the disposition of the facet model were tested by structural equation modeling: a confirmatory factor analysis (CFA; in SPSS AMOS 20) was based on the STCI facets theoretically derived for cheerfulness (CH), seriousness (SE), and bad mood (BM), and empirically isolated by exploratory factorial analysis (Ruch et al., 1996). Alternative models on the disposition of the facet model were also tested, referring to different postulates of cheerfulness. For example, Schneider (1950) hypothesized that cheerfulness and bad mood form a bipolar dimension. If this held true, two factors would be extracted: One seriousness factor and a bipolar cheerfulness- bad mood factor (model 2).

The following models were tested:

Model 1, one factor: all facets on a general factor (CH-SE-BM);
Model 2, two factors: cheerfulness-bad mood (CH-BM); and seriousness (SE);
Model 3, two factors: cheerfulness-seriousness (CH-SE); and bad mood (BM);
Model 4, two factors; cheerfulness (CH); and seriousness-bad mood (SE-BM);
Model 5, three factors: cheerfulness (CH); bad mood (BM) and seriousness (SE);
Model 6, three factors: cheerfulness (CH); bad mood (BM) and seriousness (SE); and second loadings for several facets (CH1 and CH3 on BM; SE6 on CH; BM3 and BM5 on CH).

The CFA analysis was based on a correlation matrix and maximum likelihood estimation (ML) was employed. A multifaceted approach was used to evaluate the model fit (see Tanaka, 1993; Hu and Bentler, 1999). The reported chi-square denominates the difference between the observed data and the data implied by the specified model. Yet, the chi-square test usually produces a significant value in large samples ($N > 1{,}000$), even if the difference between the observed and implied data is trivial. Several goodness-of-fit indices were used to evaluate the models, including the root-mean-square error of approximation (RMSEA), standardized root mean square residual (SRMR), normed fit index (NFI), and Tucker-Lewis coefficient (TL, known as well like non-normed fit index). In general, it is considered that a fit index above 0.90 for NFI and TL as well as RMSE and SRMR values lower 0.1, are indicators of an acceptable fit (Bollen and Long, 1993; Browne and Cudeck, 1993). Cut off values of 0.95 or higher for NFI and TL, and of 0.05 or lower for RMSEA and SRMR signify a good model fit (Hu and Bentler, 1999). Before performing the analysis, descriptive statistics were checked (see **Table 3**). **Table 3** shows that none of the facets deviated from normal distribution. Average absolute levels of the skewness and kurtosis of the facets were in the acceptable range for structural equation model analysis using likelihood estimation (Muthen and Kaplan, 1985), see **Table 3**.

TABLE 5 | Loadings of the STCI–T<106> facets on the three unrotated and three obliquely rotated factors in the construction sample.

Facets	F1	F2	F3	Obl 1	Obl 2	Obl 3	h^2
CHEERFULNESS							
CH1	−0.78	0.31	0.09	*−0.63*	0.12	−0.33	0.72
CH2	−0.73	0.25	0.28	*−0.80*	0.04	−0.04	0.67
CH3	−0.65	0.23	0.10	*−0.54*	0.05	−0.23	0.49
CH4	−0.75	0.10	0.34	*−0.82*	−0.13	0.06	0.69
CH5	−0.80	0.21	0.37	*−0.92*	−0.03	0.04	0.82
SERIOUSNESS							
SE1	0.48	0.39	−0.06	0.13	*0.55*	0.03	0.39
SE2	0.58	0.47	0.14	−0.09	*0.64*	0.30	0.58
SE3	0.33	0.58	−0.03	−0.07	*0.70*	−0.04	0.44
SE4	0.47	0.58	−0.12	0.12	*0.76*	−0.10	0.58
SE5	0.37	0.47	0.01	−0.04	*0.59*	0.04	0.36
SE6	0.75	0.19	−0.14	0.45	*0.41*	0.11	0.62
BAD MOOD							
BM1	0.78	−0.18	0.35	0.06	−0.02	*0.84*	0.76
BM2	0.74	−0.15	0.35	0.02	0.00	*0.82*	0.68
BM3	0.84	−0.14	0.11	0.34	0.07	*0.55*	0.73
BM4	0.79	−0.08	0.38	−0.02	0.09	*0.86*	0.78
BM5	0.80	−0.03	0.09	0.29	0.20	*0.48*	0.65

N = 1101. Expected loadings were italicized. F, unrotated factors; Obl, rotated factors; h^2, communality.

For the exploratory analysis, the Kaiser-Meyer-Olkin (KMO) and Bartlett's sphericity tested the sampling adequacy for applying factorial analysis. KMO value was 0.93, and the Bartlett's test showed statistical significance ($\chi^2 = 12041.43$, $df = 120$, $p < 0.001$), indicating that the samples met the criteria for factor analysis. A principal axis analysis performed on the facet inter-correlations revealed three factors exceeding unity (Eigenvalues were 7.83, 2.12, 1.19, 0.84, 0.62, 0.58, and 0.44) and also the Scree-test suggested the retention of three factors, which explained 69.05% of the variance. Moreover, we computed a parallel analysis (Horn, 1965) to verify the retention of the three factors (using the SPSS syntax provided by O'Connor, 2000). In this analysis, the eigenvalues obtained in the dataset were compared to generated eigenvalues from PAF of 100 datasets (random data generated by permutations of the original raw dataset). The first three eigenvalues met the criterion for retention (i.e., their mean exceeded the randomly generated mean across 100 datasets, with the first four random means being 1.21, 1.16, 1.13, 1.10) and thus exceeding the upper 95th percentile of the distribution of the eigenvalues retrieved from the 100 random datasets. The location of the centroids indicated that the concepts were not orthogonal. An oblique rotation was undertaken, and the reference structure of the factors is given in **Table 5**.

The factors were identified as cheerfulness (1), seriousness (3) and bad mood (2). Each facet loaded highest on the factor it belongs to. However, it was also observed that important second loadings appeared for CH1, CH3, and SE2 on the bad mood factor, and SE6, BM3, and BM5 loaded also on cheerfulness.

TABLE 6 | Assessment of fit of the STCI–T<106> data.

Models	Chi–square	*df*	RMSEA	SRMR	NFI	TL
Model 1	4696.13	104	0.21	0.11	0.73	0.70
Model 2	3089.60	103	0.16	0.10	0.81	0.79
Model 3	3756.63	103	0.18	0.11	0.78	0.76
Model 4	3490.43	103	0.17	0.10	0.79	0.77
Model 5	1829.67	101	0.10	0.08	0.88	0.86
Model 6	1180.47	96	0.09	0.06	0.92	0.90

N = 1101. RMSEA, root–mean–square error of approximation; SRMR, standardized root mean square residual; NFI, Normed Fit Index; TL, Tucker–Lewis coefficient.

The loading of SE6 on cheerfulness (−0.45) exceeded the one obtained for the factor of seriousness (0.41). The inter-correlations among the factors showed that the cheerfulness factor correlated mildly negatively with seriousness ($r = -0.46$, $p < 0.001$) and highly negatively with the bad mood factor ($r = -0.73$, $p < 0.001$), and the two forms of humorlessness were positively correlated ($r = 0.54$, $p < 0.001$). Next, confirmatory factor analysis was performed on the facets (ML estimation). The measures of fit obtained with the different models are shown in **Table 6**.

Table 6 shows that model 1 (all facets on a general factor) yielded the worst fit. Although all two factor models (model 2: cheerfulness-bad mood and seriousness; model 3: cheerfulness-seriousness and bad mood; model 4: cheerfulness and seriousness-bad mood) showed a poor fit and none of goodness-of-fit indices considered were over the limit of a reasonable fit, it should be pointed out that the model 2 presented the best fit of these two-factorial options (which would be in line with Schneider, 1950). Third, model 5 (three factors: cheerfulness, seriousness, and bad mood) showed an acceptable fit index, with a RMSEA of 0.1 and a SRMR of 0.08. Nevertheless, the NFI and Tucker-Lewis coefficient indicated the fit of the model to the data was not acceptable (NFI = 0.88; TL = 0.89). Finally, the best fit was observed for model 6. Although the fit of the expected model (model 6) was not exceptionally good, a TL of 0.90 and a NFI of 0.92, were acceptable fit indices, and in line with Bollen and Long (1993), a RMSEA of 0.09 SRMR of 0.06 would denominate the limit of a reasonable error. The inspection of residuals showed that the fit for model 6 would improve if residuals were allowed to correlate (particularly the one among the facets of each factor). This result converges with previous research at item level, and reflects that facets forming each scale are not logically independent from each other (due to the antithetical nature of the traits). Additionally, higher modification indices would appear if a relation between SE3 and bad mood, or between SE3 and cheerfulness would be introduced. The standardized pattern coefficients obtained for model 6 are shown in **Table 7**.

The coefficients shown in **Table 7** could be taken as indices of the precision with which the corresponding facets measures the factor and these correspond to the reliability analysis presented in **Table 3**, reinforcing confidence in the model 6 estimations. The standardized coefficients ranged from 0.59 (CH1) to 0.90 (CH5) for cheerfulness; from 0.43 (SE6) to 0.76 (SE2) for seriousness; and from 0.57 (BM5) to 0.88 (BM4) for bad mood. The second loadings ranged from −0.29 (CH3 in bad mood) to −0.50 (SE6 in cheerfulness).

TABLE 7 | Standardized coefficients for Model 6.

Facets	Cheerfulness	Seriousness	Bad mood
CH1	0.59		−0.30
CH2	0.82		
CH3	0.54		−0.20
CH4	0.82		
CH5	0.90		
SE1		0.63	
SE2		0.76	
SE3		0.65	
SE4		0.72	
SE5		0.59	
SE6	−0.50	0.43	
BM1			0.87
BM2			0.83
BM3	−0.33		0.60
BM4			0.88
BM5	−0.30		0.57

Model 6: three factors, (CH); Bad Mood (BM), and Seriousness (SE); and second loadings for several facets (CH1 and CH3 in BM; SE6 in CH; and BM3 and BM5 in CH).

Adaptation of the Standard Trait Form STCI-T<60> in the Construction Sample

Next, we aimed at deriving a 60 item standard form of the STCI-T<106> long form, parallel to the standard form in German and other languages. Following criteria were applied for the item selection: (a) the best corrected item to total correlation (CITC) with the own scale, (b) consideration of items content in order to preserve the content domains, (c) balanced representation of the facets (if impossible, core facets got more weight), and (d) avoidance of items with similar content or linguistic usage, (e) a good convergence with the item content of the German version (i.e., if item characteristics were similar, the paralleled item was chosen), and (f) there should be 20 items per scale. In general, a concept-guided strategy in item reduction was preferred to a purely empirical selection of items (as for the German standard form, see Ruch et al. (1996), although indices derived from PAF and item analyses were considered. Descriptive statistics (mean, standard deviation, skewness, and kurtosis), CITC and Cronbach Alpha (α) of the STCI-T<60> are given in **Table 8**.

Table 8 shows that none of the factors deviated from normal distribution. Cronbach Alpha ranged from α = 0.84 (seriousness) to α = 0.93 (cheerfulness) With respect to the CITC, all correlations were as expected, with means of the CITC being $r_m = 0.61$ for cheerfulness, $r_m = 0.42$ for seriousness, and $r_m = 0.59$ for bad mood, ranging from $r = 0.39$ to $r = 0.75$ for the items of cheerfulness to the cheerfulness scale, $r = 0.25$ to $r = 0.52$ for seriousness and $r = 0.42$ to $r = 0.72$ for bad mood. With respect to the correlations of the items with the other scales, all correlations

TABLE 8 | Descriptive statistics, corrected item to total correlation and reliability of the standard English version STCI–T<60> in the construction sample.

Scale	*Ni*	*M*	*SD*	*Sk*	*Ku*	α	CITC		
							M	Min	Max
CH	20	64.39	9.78	−0.49	−0.01	0.93	0.61	0.39	0.75
SE	20	50.10	8.37	−0.16	0.03	0.84	0.42	0.25	0.52
BM	20	39.28	11.04	0.57	0.02	0.92	0.59	0.42	0.72

N = 1101. CH, cheerfulness; SE, seriousness; BM, bad mood; Ni, number of items per facet; Sk, skewness; Ku, kurtosis; α, Cronbach Alpha; CITC, corrected item to total correlation.

were numerically lower than the CITC correlations and in the expected direction ($r = -0.61$ to $r = -0.28$ and $r = -35$ to $r = 0.00$ for the bad mood and seriousness items to the cheerfulness scale respectively; $r = -0.33$ to $r = -0.01$ and $r = 0.28$ to $r = 0.13$ for the cheerfulness and bad mood items to the seriousness scale respectively; $r = -0.62$ to $r = -0.31$ and $r = 0.01$ to $r = 0.40$ for the cheerfulness and seriousness items to the bad mood scale respectively).

Structure of the STCI-T<60>

Next, we checked the structure of the STCI-T<60> by means of factor analysis. As the STCI-T utilizes a four-point Likert format, several problems may arise when using confirmatory factor analysis at item level. It is recommended to consider Likert responses as continuous variables without normal distribution (Bentler, 1995), and work on the asymptotic matrix of covariance in order to estimate the fit. However, this model makes strong assumptions which are difficult to verify. Furthermore, it requires very large samples to obtain accurate results in large models ($N > 20$ items). For this reason, we conducted an exploratory factor analysis in MPlus (6.11; Muthén and Muthén, 2005) with a robust least squares estimator (WLSMV) and by means of using polychoric correlations to analyze the STCI-T<60>. The main goal was to have the most parsimonious solution that represents the data well. Consequently, models with one to five factors were compared and several fit indices were used to evaluate the model fit (CFI ≥0.90; RMSEA/SRMR ≤ 0.8; Browne and Cudeck, 1993; Hu and Bentler, 1999)≥0.9. **Table 9** shows the Chi square values, CFI, RMSEA, and SRMR for the five different factor solutions.

As a result of the factor analysis, the first seven Eigenvalues were: 19.23, 5.14, 3.07, 1.88, 1.42, 1.23, and 1.24. As **Table 9** shows, the CFI increases from a one to a three factor solution (with the CFI of the three factor solution meeting the criterion) and does not increase much more in a four factor solution. Thus, the extraction of a forth factor would not lead to a big increase in the fit indices. Therefore, three factors were extracted and rotated obliquely (Oblimin-criterion; delta = 0; see **Table 10**).

The three factors were clearly identified as the three theoretically expected factors of cheerfulness (1), seriousness (2), and bad mood (3), in line with other recent findings on the pilot version of the STCI-T<60> (see Lau et al., 2018). All items loaded highest on their theoretically expected factor and no important second loadings occurred, see **Table 10**. The size of the intercorrelations for CH vs. SE ($r = -0.28$, $p < 0.01$) and SE vs. BM ($r = 0.28$, $p < 0.001$) was reduced compared with to the STCI-T<106>. The correlation for CH vs. BM ($r = -0.65$, $p < 0.001$) was similar to the STCI-T<106>. The STCI-T<60> can be seen in **Appendix A**.

TABLE 9 | Exploratory factor analysis on the STCI–T<60> in the construction sample.

Models	Chi–square	*df*	*p*	CFI	RMSEA	SRMR
1 Factor	13482.96	1710	<0.001	0.824	0.079	0.093
2 Factor	7655.97	1651	<0.001	0.907	0.057	0.058
3 Factor	4382.69	1593	<0.001	0.957	0.040	0.039
4 Factor	3394.04	1536	<0.001	0.967	0.033	0.032
5 Factor	2967.52	1480	<0.001	0.972	0.030	0.029

PSYCHOMETRIC CHARACTERISTICS OF THE STCI-T<60> IN ENGLISH IN THE REPLICATION SAMPLE

Next, we checked the scale characteristic of the STCI-T<60> in an independently collected sample where also peer-reports were available. The scale characteristics of the STCI-T<60> in self-and peer report can be seen in **Table 11**.

Table 11 shows that the descriptive statistics and the high reliabilities of the STCI-T<60> were replicated for the self-reports in the Replication Sample and were also high in the peer-report version (Peer Report Sample). The CITC were sufficient, with only few exceptions, and generally replicating the results from the Construction Sample. When looking at the means of the self-reports as compared to the peer-reports, no significant differences were detected (p-values ranging from 0.10 to 0.93, all n.s.) between self-and peer-reports. Moreover, the correlations between the self- and peer-reported traits indicated a moderate to good convergence ($r = 0.74$ for CH, $r = 0.33$ for SE and $r = 0.65$ for BM, all $p < 0.001$).

DISCUSSION

The aim of the current article was to test the factorial structure of the STCI-T<106> and provide the psychometric characteristics in the English language version, as well as the adaptation of the short form with 60 items. Most importantly, the postulated facet structure of the STCI-T<106> has been confirmed in the English language version. Using a structural equation model approach, six alternative models on the disposition of the facet model were tested in the Construction Sample. Results confirmed that the theoretically derived model with three factors agreed acceptably well with the data, and presented the best fit among the tested models. Specifically, this model consisted in three related factors of cheerfulness, seriousness and bad mood, with second theoretically derived loadings for CHl (prevalence of cheerful mood) and CH3 (composed view of adverse life circumstances) on bad mood factor; SE6 (a humorless attitude about cheerfulness-related matters) on cheerfulness; and BM3 (sad individual's prototypical behavior in cheerfulness evoking

TABLE 10 | Loadings of the STCI–T<60> items on the three obliquely rotated factors in the construction sample.

Items	Obl 1	Obl 2	Obl 3	Items	Obl 1	Obl 2	Obl 3	Items	Obl 1	Obl 2	Obl 3
1BM	*0.48*			2CH		*0.43*		3SE			*0.37*
6BM	*0.57*			4CH		*0.69*		5SE			*0.36*
11BM	*0.54*			9CH		*0.85*		7SE			*0.39*
8BM	*0.75*			14CH		*0.56*		12SE			*0.65*
13BM	*0.77*			16CH		*0.68*		10SE	0.23		*0.23*
17BM	*0.66*			19CH		*0.74*		20SE			*0.39*
21BM	*0.36*	0.22		22CH		*0.79*		18SE			*0.35*
24BM	*0.48*			57CH		*0.70*	−0.22	23SE			*0.64*
40BM	*0.73*			25CH		*0.72*		28SE			*0.49*
27BM	*0.65*			26CH		*0.71*		33SE			*0.52*
31BM	*0.55*	0.26		30CH		*0.86*		36SE			*0.31*
29BM	*0.76*			32CH		*0.74*		42SE	0.25		*0.38*
34BM	*0.76*			41CH		*0.69*		39SE			*0.57*
43BM	*0.78*			38CH		*0.66*		47SE			*0.62*
37BM	*0.71*	0.21		35CH		*0.44*		52SE			*0.55*
45BM	*0.61*			46CH	−0.26	*0.66*		15SE			*0.50*
54BM	*0.64*	0.25		44CH		*0.58*		55SE			*0.49*
48BM	*0.71*			50CH		*0.71*		60SE			*0.58*
51BM	*0.40*	0.21		53CH		*0.60*		58SE			*0.50*
56BM	*0.64*			59CH		*0.66*		49SE	0.25		*0.51*

N = 1101. The items are grouped by factor. CH, cheerfulness; SE, seriousness; BM, bad mood. Expected loadings were italicized. Obl, rotated factors. All listed loadings > 0.20.

TABLE 11 | Descriptive statistics, corrected item to total correlation and reliability of the standard English version STCI–T<60> (self– and peer–reports) in the replication sample and peer–report sample.

								CITC		
	M	*SD*	*Sk*	*K*	Min	Max	α	Min	Max	*M*
SELF–REPORT										
Cheerfulness	65.46	10.07	−0.59	−0.01	35	80	0.94	0.44	0.77	0.63
Seriousness	57.86	8.42	0.21	−0.43	39	77	0.84	0.21	0.61	0.42
Bad mood	35.13	10.71	0.58	−0.45	20	61	0.92	0.42	0.74	0.59
PEER–REPORT										
Cheerfulness	66.02	11.04	−0.79	−0.31	39	80	0.95	0.10	0.76	0.46
Seriousness	59.15	10.47	−0.43	0.28	30	80	0.85	0.16	0.63	0.22
Bad mood	36.14	13.58	0.98	0.21	20	73	0.94	0.08	0.78	0.43

N = 79–87. Sk, skewness; Ku, kurtosis; α, Cronbach Alpha; CITC, corrected item to total correlation; Each facet contains 20 items.

situations) and BM5 (ill-humored individual's prototypical behavior in cheerfulness evoking situations) on cheerfulness.

Moreover, in line with the expectations, the psychometric characteristics of the STCI-T<106> were sufficient to good and comparable with the German parent version. All scales and subscales were normally distributed and had an adequate internal consistency. However, facets SE1, SE4, and SE5 presented a low internal consistency, diverging from internal consistency values reported by Ruch and colleagues (they were higher: SE1 $\alpha = 0.65$; SE4 $\alpha = 0.64$; SE5 $\alpha = 0.70$; Ruch et al., 1996). This may be explained by the restricted variance of seriousness in the Construction Sample (i.e., under-representation of individuals over 40 years of age). In the present sample, participants aged from 15 to 70 years ($M = 24.85$; $SD = 10.11$), whereas in the German sample used to develop the STCI-T participants aged from 14 to 83 ($M = 33.90$; $SD = 15.09$), being older on average. Therefore, further research is needed to investigate the reliability of the STCI-T facets for the 106 item long form. Yet, the lower reliabilities are not problematic for the current version when being used in research. Also, it is not suggested to use the facets for diagnostic purposes, but refer to the total scores (which all show high reliabilities). Moreover, we found mean differences between the German and English samples, indicating that English participants were more cheerful and less serious, and less habitually in a bad mood as compared to German participants. To summarize, the English language version of the

STCI-T<106> shows sufficient item and scale characteristics and good reliabilities for the total scores, comparable to the German version. Future studies will need to focus on replicating and providing more evidence of the scales' external validity and criterion validity for the English language version.

After looking at the characteristics of the long form of the STCI, we also adapted the standard form with 60 items in two samples. Most importantly, the item and scale characteristics were as expected and comparable to the German language version. Also, additional peer-reports showed good comparability of self- and peer-reports on cheerfulness, seriousness and bad mood.

The current study has several limitations. First, following the guidelines by Smith et al. (2000) for short form constructions, the STCI-T<60> still needs to be validated in independently collected sample that allows performing confirmatory factor analyses (i.e., having a sufficiently large *N*). This could help showing that the parent form and the short form show a good congruence of factor solutions. Second, future studies should include a sample of individuals that complete the long, as well as the short form. Third, the translation procedure utilized in this approach deviated from a classical translation-back translation procedure. Forth, future studies will need to investigate the suitableness of this English language version across nations (USA, UK, Australia, etc.). Despite those limitations, the English STCI is ready for use in research. Whereas the long form (106 items) allows a fine-grained analysis of the facets, the standard form with 60 items serves as a more economic assessment tool of cheerfulness, seriousness, and bad mood.

To conclude, the STCI-T<106>, as well as the short form with 60 items (STCI-t<60>) are ready for further validations and use. Future studies should aim at investigating the incremental validity of the *State-Trait Cheerfulness Inventory* in the prediction of humor related outcomes when controlling for broader personality traits (i.e., the "Big Five," especially extraversion). Also, future studies should investigate the relationship of different models describing the sense of humor and related traits (such as playfulness, see Proyer, 2012, 2018), as well as looking more deeply into cheerfulness interventions (see Papousek and Schulter, 2008, 2010). Moreover, future studies may opt for more balanced samples in terms of gender ratio.

AUTHOR'S NOTE

JH is at the Department of Psychology at the University of Zurich, Switzerland. HC-D is at the Department of Research Methods in Behavioral Sciences, University of Granada, Spain. AC is at the Department of English at the University of Central Oklahoma, Edmond, United States of America. Our thanks go to Lambert Deckers and Willibald Ruch for helping with the data collection and Willibald Ruch for providing helpful comments on a prior version of the manuscript.

AUTHOR CONTRIBUTIONS

JH, HC-D: concept; AC: data collection; JH, HC-D: analyses; JH, HC-D: writing draft; JH, AC, HC-D: revisions and feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02255/full#supplementary-material

REFERENCES

Auerbach, S. (2017). Are clowns good for everyone? The influence of trait cheerfulness on emotional reactions to a hospital clown intervention. *Front. Psychol.* 8:1973. doi: 10.3389/fpsyg.2017.01973

Beermann, U., and Ruch, W. (2011). Can people ever "laugh at themselves"? Experimental and correlational evidence. *Emotion* 11, 492–501. doi: 10.1037/a0023444

Bentler, P. M. (1995). *EQS Structural Equations Program Manual*. Encino, CA: Multivariate Software.

Beyler, M. (1999). *Ist Erheiterung entspannend? [Is exhilaration relaxing?]* Unpublished Master Thesis, Department of Psychology, University of Düsseldorf, Düsseldorf, Germany.

Bollen, K. A., and Long, J. S. (1993). *Testing Structural Equation Models*. Sage: London.

Browne, M. W., and Cudeck, R. (1993). "Alternative ways of assessing model fit," in *Testing Structural Equation Models*, eds K. A. Bollen and J. S. Long (Sage: London), 136–162.

Carretero-Dios, H., Benítez, I., Delgado-Rico, E., Ruch, W., and López-Benítez, R. (2014). Temperamental basis of sense of humor: the Spanish long form of the trait version of the State-Trait-Cheerfulness-Inventory. *Personality Individ. Diff.* 68, 77–82. doi: 10.1016/j.paid.2014.03.045

Carretero-Dios,. H., Eid, M., and Ruch, W. (2011). Temperamental basis of sense of humor: a multilevel confirmatory factor analysis of multitrait-multimethod data. *J. Res. Personal.* 45, 153–164. doi: 10.1016/j.jrp.2010.12.007

Chen, G.-H., Ruch, W., and Li, Y.-N. (2017). Temperamental basis of sense of humor: validating the state-trait-cheerfulness-inventory in Mainland China. *Humor Int. J. Humor Res.* 29, 77–97. doi: 10.1515/humor-2015-0036

Craik, K. H., Lampert, M. D., and Nelson, A. J. (1996). Sense of humor and styles of everyday humorous conduct. *Humor Int. J. Humor Res.* 9, 273–302. doi: 10.1515/humr.1996.9.3-4.273

Delgado-Domínguez, C. J., Escudera-Contrras, A., Font-Ugalde, P., Ruiz-Vilchez, D., Collantes-Estévez, E., and Carretero-Dios, H. (2016). Upswings in cheerful mood and disease activity in patients with rheumatoid arthritis. *Int. J. Behav. Med.* 23, 606–610. doi: 10.1007/s12529-016-9550-2

Edwards, K. R. (2012). Do humorous people take poorer care of their health? Associations between humor styles and substance use. *Eur. J Psychol.* 8, 523–534.

Falkenberg, I., Buchkremer, G., Bartels, M., and Wild, B. (2011a). Implementation of a manual-based training of humor abilities in patients with depression: a pilot study. *Psychiatry Res.* 186, 454–457. doi: 10.1016/j.psychres.2010.10.009

Falkenberg, I., Jarmuzek, J., Bartels, M., and Wild, B. (2011b). Do depressed patients lose their sense of humor? *Psychopathology* 44, 98–105. doi: 10.1159/000317778

Falkenberg, I., Klügel, K., Bartels, M., and Wild, B. (2007). Sense of humor in patients with schizophrenia. *Schizophrenia Res.* 95, 259–261. doi: 10.1016/j.schres.2007.06.006

Hausser, S. (1999). *Heiterkeit, kritische Lebensereignisse und Wohlbefinden [Cheerfulness, critical life events and well-being]*. Unpublished Master Thesis, Department of Psychology, Heinrich-Heine University, Düsseldorf.

Hirsch, R. D., Junglas, K., Konradt, B., and Jonitz, M. F. (2010). Humortherapie bei alten menschen mit einer depression. Ergebnisse einer empirischen Untersuchung [Humortherapy in elderly depressed patients: empirical results]. *Zeitschrift für Gerontologie Geriatrie* 43, 42–52. doi: 10.1007/s00391-009-0086-9

Hodson, G., MacInnis, C. C., and Rush, J. (2010). Prejudice-relevant correlates of humor temperaments and humor styles. *Personality Individ. Diff.* 49, 546–549. doi: 10.1016/j.paid.2010.05.016

Hofmann, J., Heintz, S., Pang, D., and Ruch, W. (in press). Differential relationships of light and darker forms of humor with mindfulness. Applied Research in Quality of Life, Special Issue "Character strengths, well-being, and health in educational and vocational settings".

Hofmann, J., Platt, T., Ruch, W., and Niewiadomski, R., Urbain, J. (2015). The influence of a virtual companion on amusement when watching funny films. *Motiv. Emot.* 39, 434–447. doi: 10.1007/s11031-014-9461-y

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrica* 30, 179–185. doi: 10.1007/BF02289447

Hu, L., and Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equation Modelling* 6, 1–55. doi: 10.1080/10705519909540118

Krantzhoff, E. U., and Hirsch, R. D. (2001). "Humor in der gerontopsychiatrischen klinik: ergebnisse einer therapiebegleitenden studie [Humor in geriatric psychiatric hospitals: results of a therapy evaluation study]," in *Heiterkeit und Humor im Alter*, eds R. D. Hirsch, J. Bruder, and H. Radebold. Kassel: Chudeck Druck. p 139–162.

Lau, C., Chiesi, F., Hofmann, J., Ruch, W., and Saklofske, D. H. (2018). The Italian Version of the State-Trait Cheerfulness Inventory Trait Form: Psychometric Validation and Evaluation of Measurement Invariance. Manuscript submitted for publication.

López-Benítez, R., Acosta, A., Lupiáñez, J., and Carretero-Dios, H. (2017a). Are you ready to have fun? The Spanish state form of the state–trait–cheerfulness inventory. *J. Pers. Assess.* 21, 1–12. doi: 10.1080/00223891.2017.1368022

López-Benítez, R., Acosta, A., Lupiáñez, J., and Carretero-Dios, H. (2017b). High trait cheerfulness individuals are more sensitive to the emotional environment. *J. Happiness Stud.* 19, 1589–1612.

López-Benítez, R., Carretero-Dios, H., Acosta, A., and Lupiáñez, J. (2017c). Trait Cheerfulness Does Not Influence Switching Costs But Modulates Preparation and Repetition Effects in a Task-Switching Paradigm. *Front. Psychol.* 8:1013. doi: 10.3389/fpsyg.2017.01013

Manke, B. (2007). "Genetic and environmental contributions to children's interpersonal humor," in *The Sense of Humor: Explorations of a Personality Characteristic*, ed W. Ruch (Berlin: Mouton de Gruyter), 361–384.

Martin, R. A., and Lefcourt, H. M. (1983). Sense of humor as a moderator of the relation between stressors and moods. *J. Personality Soc. Psychol.* 45, 1313–1324. doi: 10.1037/0022-3514.45.6.1313

Martin, R. A., and Lefcourt, H. M. (1984). Situational humor response questionnaire: quantitative measure of sense of humor. *J. Soc. Personality Psychol.* 47, 145–155. doi: 10.1037/0022-3514.47.1.145

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., and Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: development of the humor styles questionnaire. *J. Res. Personality* 37, 48–75. doi: 10.1016/S0092-6566(02)00534-2

McGhee, P. E. (1996). *Humor, Healing, and the Amuse System: Humor as Survival Training*, 2nd Edition. Dubuque: Kendall/Hunt Publishing.

Muthen, B., and Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *Br. J. Mathematic Stat. Psychol.* 38, 171–189. doi: 10.1111/j.2044-8317.1985.tb00832.x

Muthén, L. K., and Muthén, B. O. (2005). *Mplus: Statistical Analysis With Latent Variables: User's Guide*. Los Angeles, CA: Muthén and Muthén.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav. Res. Methods Instrumentation Computers* 32, 396–402. doi: 10.3758/BF03200807

Papousek, I., and Schulter, G. (2008). Effects of a mood-enhancing intervention on subjective well-being and cardiovascular parameters. *Int. J. Behav. Med.* 15, 293–302. doi: 10.1080/10705500802365508

Papousek, I., and Schulter, G. (2010). "DON'T TAKE AN X FOR A U: why laughter is not the best medicine, but being more cheerful has many benefits," in *Psychological Well-Being*, ed I. E. Wells (New York, NY: Nova Science Publishers, Inc), 1–75.

Proyer, R. T. (2012). Development and initial assessment of a short measure for adult playfulness: the SMAP. *Personality Individ. Differences* 53, 989–994. doi: 10.1016/j.paid.2012.07.018

Proyer, R. T. (2018). Playfulness and humor in psychology: an overview and update. *Humor Int. J. Humor Res.* 31, 259–271. doi: 10.1515/humor-2016-0080

Proyer, R. T., and Rodden, F. A. (2013). Is the homo ludens cheerful and serious at the same time? An empirical study of Hugo Rahner's notion of Ernstheiterkeit. *Arch. Psychol. Religion* 35, 213–231.

Rapp, A. M., Wild, B., Erb, M., Rodden, F. A., Ruch, W., and Grodd, W. (2008). Trait cheerfulness modulates BOLD-response in lateral cortical but not limbic brain areas - a pilot fMRI study. *Neurosci. Lett.* 445, 242–245. doi: 10.1016/j.neulet.2008.09.017

Ruch, W. (1993). "Exhilaration and humor (Chapter 42)," in *The Handbook of Emotions*, eds M. Lewis and J. M. Haviland (New York, NY: Guilford Publications), 605–616.

Ruch, W. (1997). State and trait cheerfulness and the induction of exhilaration: A FACS study. *Eur. Psychol.* 2, 328–341. doi: 10.1027/1016-9040.2.4.328

Ruch, W. (Ed.) (2007). *The Sense of Humor: Explorations of a Personality Characteristic*. Berlin: Mouton de Gruyter.

Ruch, W., Beermann, U., and Proyer, R. T. (2009). Investigating the humor of gelotophobes: does feeling ridiculous equal being humorless? *Humor Int. J. Humor Res.* 22, 111–143. doi: 10.1515/HUMR.2009.006

Ruch, W., and Carrell, A. (1997). Trait cheerfulness and the sense of humor. *Pers. Individ. Diff.* 24, 551–558. doi: 10.1016/S0191-8869(97)00221-3

Ruch, W., and Hofmann, J. (2012). *A Temperament Approach to Humor*, in *Humor and Health Promotion*, ed P. Gremigni (Hauppauge, NY: Nova Science Publishers), 79–113.

Ruch, W., Hofmann, J., Rusch, S., and Stolz, H. (2018). The impact of the 7 humor habits program on the sense of humor and satisfaction with life. *Humor Int. J. Humor Res. Festschrift for Paul McGhee Humor Across the Lifespan, Theory, Measurement, and Applications*.

Ruch, W., and Köhler, G. (1999). The measurement of state and trait cheerfulness. Pers. Psychol. Europe Theor. Empir. Dev. 7, 67–83.

Ruch, W., and Köhler, G. (2007). "A temperament approach to humor," in *The Sense of Humor: Explorations of a Personality Characteristic*, ed W. Ruch (Berlin: Mouton de Gruyter), 203–230. doi: 10.1515/9783110804607-011

Ruch, W., Köhler, G., and van Thriel (1996). "Assessing the "humorous temperament": Construction of the facet and standard trait forms of the State-Trait-Cheerfulness-Inventory — STCI," in *Measurement of the Sense of Humor [special issue]*, ed W. Ruch, *Humor Int. J. Humor Res.* 9, 303–339. doi: 10.1515/humr.1996.9.3-4.303

Ruch, W., Köhler, G., and van Thriel (1997). To be in good or bad humor: construction of the state form of the State-Trait-Cheerfulness-Inventory—STCI. *Pers. Individ. Dif.* 22, 477–491. doi: 10.1016/S0191-8869(96)00231-0

Ruch, W., Proyer, R. P., Esser, C., and Mitrache, O. (2011). "Cheerfulness and everyday humorous conduct," in *Romanian Academy, "George Barit" Institute of History, Department of Social Research, Studies and Researches in Social Sciences* (Cluj-Napoca: Argonaut Press), 67–87.

Ruch, W., and Stevens, M. (1995). "The differential effects of nitrous oxide on mood level: the role of trait-cheerfulness," in *7th Meeting of the International Society for the Study of Individual Differences-ISSID, July* 15-19, 1995, (Warsaw).

Ruch, W., and Zweyer, K. (2001). "Heiterkeit und humor: ergebnisse der forschung," in (Hrsg) *Heiterkeit und Humor im Alter, Schriftenreihe der Deutschen Gesellschaft für Gerontopsychiatrie und –Psychotherapie Band 2*, eds R. D. Hirsch, J. Bruder, and H. Radebold (Bornheim: Chudeck-Druck), 9–43.

Schneider, K. (1950). *Klinische Psychopathologie*. Thieme, Stuttgart.

Smith, G. T., McCarthy, D. M., and Anderson, K. G. (2000). On the sins of short-form development. *Psychol. Assess.* 12, 102–111. doi: 10.1037/1040-3590.12.1.102

Tanaka, J. S. (1993). "Multifaceted conceptions of fit in structural models," in *Testing Structural Equation Models*, eds K. A. Bollen and J. S. (Long Sage: London), 10–39.

Tapia-Villanueva, L., Armijo, I., Pereira, X., and Molina, M. E. (2014). Chilean Spanish version of the State Trait Cheerfulness Inventory (STCI-T-60, trait form): individual and couple forms. *Humor Int. J. Humor Res.* 27, 481–498. doi: 10.1515/humor-2014-0074

Thompson, R., Ruch, W., Hasenöhrl R., U. (2004). Enhanced cognitive performance and cheerful mood by standardized extracts of piper methysticum (Kava-kava). *Human Psychopharmacol. Clin. Exp.* 19, 243–250. doi: 10.1002/hup.581

Wancke, C. U. (1996). *Der Einfluss unterschiedlicher emotionaler Qualitäten von Sinnsprüchen und des Umgangs mit dem Material auf die aktuelle Befindlichkeit von Probanden [The influence of sayings of different emotional qualities and the handling of the material on the mood of participants]*. Unpublished Master Thesis, Department of Psychology, University of Düsseldorf.

Yip, J. A., and Martin, R. A. (2006). Sense of humor, emotional intelligence, and social competence. *J. Res. Pers.* 40, 1202–1208. doi: 10.1016/j.jrp.2005.08.005

Zweyer, K., Velker, B., and Ruch, W. (2004). Do cheerfulness, exhilaration and humour production moderate pain tolerance? A FACS study; sense of humor and health (special issue). *Humor Int. J. Humor Res.* 17, 67–84. doi: 10.1515/humr.2004.009

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Hofmann, Carretero-Dios and Carrell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 June 2017
doi: 10.3389/fpsyg.2017.01013

Trait Cheerfulness Does Not Influence Switching Costs But Modulates Preparation and Repetition Effects in a Task-Switching Paradigm

Raúl López-Benítez[1], Hugo Carretero-Dios[2], Alberto Acosta[1] and Juan Lupiáñez[1]*

[1] Department of Experimental Psychology, Mind, Brain and Behavior Research Center, Faculty of Psychology, University of Granada, Granada, Spain, [2] Department of Methodology of Behavioral Sciences, Mind, Brain and Behavior Research Center, Faculty of Psychology, University of Granada, Granada, Spain

Many studies have shown the beneficial effect of positive emotions on various cognitive processes, such as creativity and cognitive flexibility. Cheerfulness, understood as an affective predisposition to sense of humor, has been associated with positive emotions. So far, however, no studies have shown the relevance of this dimension in cognitive flexibility processes. The aim of this research was to analyze the relationship between cheerfulness and these processes. To this end, we carried out two studies using a task-switching paradigm. Study 1 aimed at analyzing whether high trait cheerfulness was related to better cognitive flexibility (as measured by reduced task-switching costs), whereas Study 2 aimed at replicating the pattern of data observed in Study 1. The total sample was composed of 139 participants (of which 86 were women) selected according to their high versus low scores in trait cheerfulness. In a random way, participants had to judge whether the face presented to them in each trial was that of a man or a woman (gender recognition task) or whether it expressed anger or happiness (expressed emotion recognition task). We expected participants with high versus low trait cheerfulness to show a lower task-switching cost (i.e., higher cognitive flexibility). Results did not confirm this hypothesis. However, in both studies, participants with high versus low trait cheerfulness showed a higher facilitation effect when the stimuli attributes were repeated and also when a cue was presented anticipating the demand to perform. We discuss the relevance of these results for a better understanding of cheerfulness.

Keywords: sense of humor, trait cheerfulness, task switching, cognitive flexibility, attribute repetition, preparation

OPEN ACCESS

Edited by:
Willibald Ruch,
University of Zurich, Switzerland

Reviewed by:
Ursula Beermann,
University of Innsbruck, Austria
Ilona Papousek,
University of Graz, Austria
Sarah Gaither,
Duke University, United States

****Correspondence:***
Raúl López-Benítez
raullopezbenitez@ugr.es

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *23 March 2017*
Accepted: *01 June 2017*
Published: *22 June 2017*

Citation:
López-Benítez R, Carretero-Dios H, Acosta A and Lupiáñez J (2017) Trait Cheerfulness Does Not Influence Switching Costs But Modulates Preparation and Repetition Effects in a Task-Switching Paradigm. Front. Psychol. 8:1013. doi: 10.3389/fpsyg.2017.01013

INTRODUCTION

Nowadays, one of the main areas of interest in the sense of humor field has been to provide a global theoretical framework to guide research. In this sense, Ruch et al. (1996, 1997) developed a theoretical model focused on isolating the temperamental basis of sense of humor: cheerfulness, seriousness, and bad mood.

Cheerfulness, the subject of this research, is understood as a predisposition to smile/laugh and express positive emotions in response to humorous stimuli, alongside a general tendency to show a positive and a joy affective state. This dimension comprises five facets: the prevalence of a cheerful mood, a low threshold for smiling and laughter, a composed view of adverse life circumstances, a broad range of active elicitors of cheerfulness and smiling/laughter, and a generally cheerful interaction style. In the model (Ruch and Hofmann, 2012) only cheerfulness encourages hilarity[1].

Ruch and colleagues developed an inventory to assess the individual differences and connections that may exist between the affective and cognitive basis laid out in the model from both a trait perspective [State-Trait Cheerfulness Inventory-Trait Version (STCI-T); Ruch et al., 1996] and a state perspective [State-Trait Cheerfulness Inventory-State Version (STCI-S); Ruch et al., 1997]. This fact, along with the extensive body of knowledge obtained on cheerfulness over the last 20 years, has contributed to its development from both a theoretical and empirical point of view.

Previous research has shown that cheerfulness plays an important role in humor. In this sense, it has been pointed out that cheerfulness affects dispositions of the exhilaration response (Ruch, 1997), predicts most of sense-of-humor facets, contributes to the use of humor as a recovery strategy, and is associated with affiliative and self-enhancing humor styles (Ruch and Hofmann, 2012). Moreover, other research support the applicability and relevance of cheerfulness in areas as diverse as personality, health, or emotion (e.g., Ruch et al., 1996, 1997; Yip and Martin, 2006; Ruch and Köhler, 2007; Papousek and Schulter, 2010; Carretero-Dios et al., 2011; Ruch and Hofmann, 2012; Delgado-Domínguez et al., 2016).

Thus, the concept of cheerfulness can be granted similar virtues to those attributed to positive emotions (see Lyubomirsky et al., 2005, for a review). For instance, it has been established that trait cheerfulness is closely associated with better physical and psychological well-being, an increased manifestation and expression of positive emotions, satisfaction, and quality of life, better resilience, ability to cope, and recovery from stressful situations, a greater ability to use creative thinking, and high interpersonal skills (Papousek and Schulter, 2010; Ruch and Hofmann, 2012).

Within the area of research on positive emotions, several studies have highlighted the influence of such emotions on cognitive flexibility (e.g., Wadlinger and Isaacowitz, 2006). The results obtained can be included in Fredrickson's (2001) broaden-and-build theory, which suggests that positive emotions expand our mental and behavioral repertoire. As a consequence, after being exposed to positive affective states our scope of attention broadens (see, for example, Johnson et al., 2010) and aspects of cognition such as cognitive flexibility increase, leading to an adaptation to changes in the environment. In this regard, it should be noted that the conceptualization of cheerfulness as a positive affective dimension linked to sense of humor leads us to wonder how relevant this factor is for the study of cognitive flexibility.

[1]Ruch proposed the term *exhilaration* or *amusement* (hilarity, joy, euphoria, or rejoicing) as a name for this emotion, which is used to denote either the process of making cheerful or the temporary rising and fading out of a cheerful state (Ruch and Köhler, 2007, p. 205).

Cognitive Flexibility and Control Processes

Control processes are related to individuals' ability to select relevant information and ignore irrelevant information when performing a task (Posner and Rothbart, 2007). They are also related to cognitive flexibility (Davidson et al., 2006), understood as the ability to modify one's way of thinking or acting in accordance with changing demands.

Some authors argue that cognitive control has three central components: the inhibition of whatever is irrelevant to the fulfillment of our goals, the updating and monitoring of the information, and the switch between mindsets to activate the relevant material for the particular demand at hand (Miyake et al., 2000). When we perform two or more tasks alternately, we must constantly reconfigure our mindset to respond to the new demand (Crone et al., 2006). The ease with which these readjustments are carried out is the key defining characteristic of cognitive flexibility, which is a fruitful process for adapting to the environment.

Studies on control processes and cognitive flexibility have used numerous tasks (e.g., Stroop, 1935; Simon, 1969; Eriksen and Eriksen, 1974). Recently, one of the most widely used experimental procedures to explore cognitive flexibility has been task switching (Monsell, 2003; Kiesel et al., 2010). In task-switching tasks, participants are instructed to perform one of two possible tasks in each trial. In some consecutive trials the same demand is repeated, while in others it is different. This makes it possible to determine the task-switching cost, measured as the difference in performance when the task changes in two consecutive trials, compared to when it is repeated.

It has additionally been proved that, in this type of task, the amount of stimuli attributes that either repeats or changes on consecutive trials can also affect behavior and the typical effects of task switching costs. When an individual is exposed to a stimulus, a mental file is created about this event, including the attributes of the stimulus as well as the response to it. This representation is subsequently reactivated in the presence of similar stimuli, thus affecting the performance of tasks involving these stimuli (Hommel, 2004). In this regard, it has been reported that total attribute repetition only has a beneficial effect if the response is the same in two consecutive trials (Kahneman et al., 1992). However, the performance is worse when there is partial attribute repetition than when there is no attribute repetition (or when all the attributes are repeated). This is because, although in some cases this repetition may help solve the demand, it normally requires reconfiguring the previously created mental file (Hommel, 1998, 2004). Additionally, some studies have included cognitive or affective demands, or between two different cognitive demands in the presence of the same stimuli, which have made it possible to determine the task-switching cost between two consecutive trials depending on the type of demand (e.g., Egner et al., 2008; Ochsner et al., 2009; Schuch et al., 2012). Importantly, both the repetition of attributes and the type of task

interact with task switching (Marzecová et al., 2013) and therefore should be considered when studying task-switching costs.

Despite the lack of any existing literature on the modulation of cognitive flexibility processes by cheerfulness, some studies are beginning to offer clues on their possible relationship. Previous research has pointed out that the induction of positive affective states, which are related to cheerfulness, are associated with a better cognitive flexibility (Baumann and Kuhl, 2005; Yang and Yang, 2014). From a correlational perspective, it has been established that cheerfulness is linked to some personality variables of interest for the current research (Ruch and Köhler, 2007). For example, Carretero-Dios et al. (2014) observed positive relationships among trait cheerfulness, extraversion, openness, and agreeableness, and negative relationships between trait cheerfulness and neuroticism. And, importantly, some studies have found that such personality characteristics may modulate performance on tasks that requires cognitive flexibility (Murdock et al., 2013). For example, while positive associations among openness (DeYoung et al., 2005), agreeableness (Jensen-Campbell et al., 2002), and cognitive flexibility has been observed, extraversion (Campbell et al., 2011) and neuroticism (Compton, 2000) seem to contribute to reduce it.

Links between cognitive flexibility and sense of humor could also contribute to explain the possible modulation by cheerfulness. Some studies derived from clinical populations, such as Asperger's syndrome (Weiss et al., 2013) or Schizophrenia (Tsoi et al., 2008; Polimeni et al., 2010) have found reduced sensitivity to recognize or discriminate humor in these populations, perhaps reflecting a deficit in cognitive functions such as cognitive switching. In fact, one important component of humor response has been related to cognitive processes related to re-interpretation of evidence and congruity resolution, which involves cognitive flexibility (Suls, 1972). Furthermore, it has been established that cognitive flexibility and the use of emotion regulation strategies are positively related (e.g., Gul and Khan, 2014). For example, Malooly et al. (2013) found that a lower task-switching cost predicted the success to use reappraisal strategies to down-regulate negative emotions.

More specifically, cheerfulness has been specifically associated to cognitive flexibility; in its third facet -composed view of adverse life circumstances- it is assumed that high trait cheerfulness individuals are good in re-interpreting events (e.g., "Most problems turn out to be not as bad as all that when considered calmly and composedly"). Therefore, given that trait cheerfulness is an important key to understand and produce humor, and trait cheerfulness has been associated to a high ability to cope with negative events (Ruch and Hofmann, 2012), people scoring high in trait cheerfulness might also have better executive functioning.

To test this hypothesis, we conducted a study in our laboratory (López-Benítez et al., unpublished) in which participants differentiated in trait cheerfulness (assessed with the STCI-T) were required to perform the following task-switching paradigm: in a random way, in each trial, they had to say whether the face presented to them on a screen was that of a man or a woman (gender recognition task) or if the face expressed anger or happiness (expressed emotion recognition task). The task could change, or not, between two consecutive trials. The various conditions of repetition of the stimuli attributes were also analyzed (Kahneman et al., 1992; Hommel, 1998, 2004). With the additional goal of studying interference effects, the faces were always presented with a written word at the center that could match their gender or expression (congruent trials) or not (incongruent trials) (depending on the task; e.g., Etkin et al., 2006). Results showed an interesting trend: individuals with high trait cheerfulness showed a lower task-switching cost than those with low trait cheerfulness, especially in the conditions in which all the attributes were repeated between consecutive trials. These results were interpreted as showing that these individuals have higher cognitive flexibility in repetition conditions, precisely where cognitive flexibility is most necessary.

However, this interpretation should be taken cautiously due to several factors. First, the size of the observed effect was small (0.05) and the interaction between task change, group, and attribute repetition was only marginally significant, all of which suggests that the result should be further studied. Moreover, in that study we included the interference variable. Although this variable did not interact with trait cheerfulness, it might affect the analysis of task-switching costs as participants had to use more cognitive resources, especially on incongruent trials, which made the task especially harder.

In spite of the relevance of studying cheerfulness and cognitive flexibility, there are still no studies that have deepened on their possible relationships. In this study, we aimed at bridging this gap. As a first step and from a systematic point of view, we wanted to analyze whether trait cheerfulness had an impact on cognitive flexibility processes. We consider that this study is highly relevant because if cheerfulness indeed plays a role on executive functions such as cognitive flexibility, the assessment and training of cheerfulness could be considered as a relevant aspect in the improvement of skills focused in adaptation to the environment, which is a basic human function. In addition, with this study we could check whether previous relationships between cognitive processes and humor are expanded to a predisposition (as a trait) to sense of humor at the same time that its temperamental basis theoretical model is empirically tested (Ruch et al., 1996).

To achieve that aim, two studies were carried out. In both studies, two groups of participants scoring high versus low in trait cheerfulness performed a task-switching paradigm. In Study 1, we analyzed whether high trait cheerfulness people had better cognitive flexibility (as measured by a lower task-switching cost), whereas in Study 2 we extended and checked the consistency of the pattern of data observed in the first study.

STUDY 1

Taking previous research into account, we conducted this study to analyze whether trait cheerfulness (operationalized with the STCI-T) could be directly related to cognitive flexibility through a task-switching paradigm. To this end, as in our previous unpublished study (above described), participants carried out a task in which they had to correctly identify either the emotion or the gender of a face presented in the center of the screen; this task

was randomly repeated or alternated between consecutive trials. However, in this study, in order to simplify the experimental design, we removed the interference variable, that is, we did not present a word superimposed on the faces. In addition, half of the trials were preceded by a cue that anticipated the upcoming task, allowing participants to get ready for it. The inclusion of this variable is important, as it has been proven that the presentation of a cue that anticipates the demand reduces the cognitive effort required, which is likely to lead to a better performance in this type of task (see Kiesel et al., 2010). Based on the above-mentioned studies and taking into account that several studies have shown that positive affective states are associated with a lower task-switching cost (Yang and Yang, 2014), we predicted that, compared to individuals with low trait cheerfulness, individuals with high trait cheerfulness would have greater cognitive flexibility, thus showing a lower task-switching cost, particularly when performing trials that require greater cognitive flexibility (i.e., attribute repetition and no prior preparation).

Material and Methods

Participants

The sample was composed of 49 students from the University of Granada, who were selected from a total of 244 people according to their high versus low trait cheerfulness scores, obtained with the Spanish version of the STCI-T, cheerfulness dimension (Carretero-Dios et al., 2014). The average score ± 1 SD was used as a criterion to create the groups. Specifically, the high trait cheerfulness group comprised 24 participants (20 women, mean age 19.50 years, $SD = 5.82$, cut-off score ≥ 3.42), and the low trait cheerfulness group was made up of 25 participants (20 women, mean age 21.60 years, $SD = 7.65$, cut-off score ≤ 2.68). All participants had normal or corrected-to-normal vision, participated in the study voluntarily, and received course credit in exchange for participating. They signed an informed consent and had the possibility to stop the experimental session without any consequences. Data from one participant were not taken into account because the number of correct responses was low compared to the group (more than 2.5 SD below from the group mean). The study was part of a broader research project, approved by the Ethics Committee of the University of Granada, in accordance with the 1964 Declaration of Helsinki.

Stimuli

In order to conduct the study, eight photographs were selected from the database of the Karolinska Institute in Stockholm, Sweden (Lundqvist et al., 1998; Goeleven et al., 2008). The images showed two happy men (AM25HAS; AM10HAS), two angry men (AM09ANS; AM02ANS), two happy women (AF31HAS; AF14HAS), and two angry women (AF20ANS; AF25ANS). All the photographs were 141 mm × 191 mm in size. Additionally, a 100 ms sound was used to provide participants feedback on their performance during the practical part of the experiment.

Procedure

Participants went to the laboratory individually and were led to a soundproofed, dimly lit room. They were seated in a comfortable chair in front of a 15-inch computer monitor, at a distance of 60 cm. They gave their consent prior to the start of the experiment. Next, the researcher informed them that the goal of the study was to analyze their performance in a psychological task, to which they should respond as quickly as possible while trying to avoid any errors.

The researcher explained how they should respond to the task, and was present during some practice trials to ensure that they were performing them correctly. After that, the researcher left the room and the experimental trials were presented.

At the beginning of each trial, a fixation point appeared in the center of the screen for 1 s. Randomly, in half of the trials a green or purple mark (preparation condition) also appeared around the fixation point, anticipating the task participants had to perform next. After the second, one of the eight photographs previously described appeared on the screen, surrounded by a green or purple frame, which indicated the nature of the task to perform: to indicate either the emotion on the face (happiness vs. anger) or the gender (man vs. woman). In the half of the trials in which the colored frame did not appear along with the fixation point (no preparation condition), the frame was presented simultaneously with the photograph. The different trials were presented randomly and the specific sequences of consecutive trials were coded off-line in order to code the other variables of interest. Thus, in approximately half of the trials, the task was the same in two consecutive trials (same task), while in the rest of the trials it changed (different task). On the other hand, sometimes the attributes of the stimuli (gender and emotion) were repeated in two consecutive trials (complete repetition), whereas in other cases these characteristics were not repeated at all (complete alternation) or only one of them was repeated (partial repetition).

To prevent any biases, the color associated to each task was counterbalanced across participants as follows: for half of the sample the green color was associated with the gender task and the purple color was associated with the emotion task; the opposite was true for the other half of the sample. To respond, participants had to press the "Z," "M," "X," or "N" keys in a QUERTY keyboard. The correspondence between key and response was also counterbalanced across participants. Specifically, for half of the sample the "Z" key was associated with "male," "M" with "female," "N" with "happiness," and "X" with "anger," while for the other half of the sample "Z" was associated with "female," "M" with "male," "N" with "anger," and "X" with "happiness." The total duration of each trial was 4 s. **Figure 1** illustrates the sequence of events in two trials.

The experimental task was programmed using E-prime software (Schneider et al., 2002). It comprised 1 block of practice (32 trials) and 8 blocks of 64 trials each, with a total duration of 40–45 min.

Design

The data were analyzed using SPSS 21.0 statistical software, with a 2 (Group; High Trait Cheerfulness vs. Low Trait Cheerfulness) × 2 (Task; Emotion vs. Gender) × 3 (Repetition; Complete Alternation vs. Complete Repetition vs. Partial Repetition) × 2 (Task Change; Different vs. Same) × 2

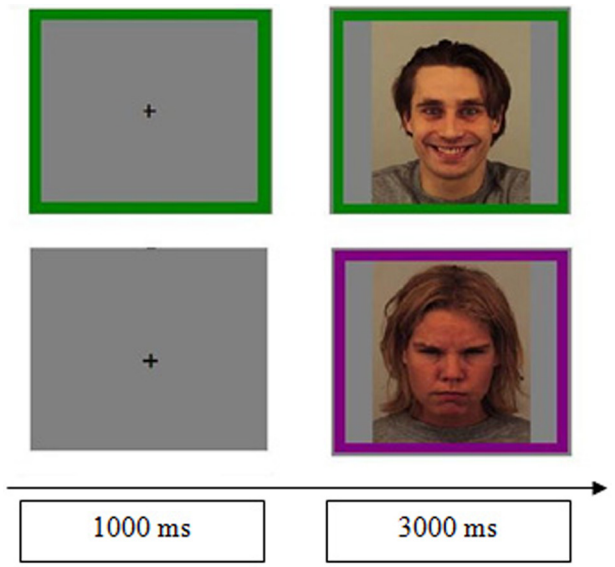


FIGURE 1 | Sequence of events in two trials. In the upper example, which illustrates the preparation condition, the fixation point is surrounded by a signal that anticipates the task. After 1000 ms, a picture of a happy man (AM25HAS) appears for 3 s, surrounded by a green frame. In the lower example (i.e., the no preparation condition), the target, an angry woman (AF25ANS) surrounded by a purple frame, is not preceded by any signal. The color of the frame indicates the demand to perform, which is to identify either the gender or the emotion of the face, depending on the counterbalancing.

(Preparation; Preparation vs. No Preparation) mixed factorial design. The first variable was manipulated between groups, and the rest were manipulated within participants. The dependent variables were reaction time (RT), which was calculated only for correct responses also preceded by correct responses, and error percentage (EP).

Results

Descriptive statistics are shown on **Table 1**. The analysis revealed a main effect of each of the within-participant variables: Task, $F(1,46) = 39.56$, $p < 0.001$, $\eta^2 = 0.46$, Repetition, $F(2,92) = 16.31$, $p < 0.001$, $\eta^2 = 0.26$, and Preparation, $F(1,46) = 339.00$, $p < 0.001$, $\eta^2 = 0.88$. Participants were faster to respond when the task was gender identification (898 ms vs. 966 ms), when all the attributes were repeated in two consecutive trials, compared to when none were repeated or only some of them were (912 ms vs. 945 ms vs. 939 ms, respectively), and when a cue was presented anticipating the task to perform (824 ms vs. 1040 ms). Moreover, our task replicated the expected task-switching cost results, $F(1,46) = 191.31$, $p < 0.001$, $\eta^2 = 0.81$, meaning that participants were faster when the task was repeated between two consecutive trials (134 ms task-switching cost). Additionally, as expected, this effect was modulated by attribute repetition, $F(2,92) = 21.66$, $p < 0.001$, $\eta^2 = 0.32$, preparation conditions, $F(1,46) = 46.82$, $p < 0.001$, $\eta^2 = 0.50$, and task type, $F(1,46) = 10.64$, $p = 0.002$, $\eta^2 = 0.19$. Specifically, participants showed a lower task-switching cost when none of the stimulus attributes (i.e., gender

TABLE 1 | Mean reaction time (in ms), error percentage and error standard deviation in each of the experimental conditions as a function of trait cheerfulness.

			Complete alternation				Complete repetition				Partial repetition			
			Different		Same		Different		Same		Different		Same	
			P	NP	P	NP	P	NP	P	NP	P	NP	P	NP
High Trait Cheerfulness	Emotion		889 (5.69, 6.53)	1225 (6.75, 7.31)	818 (2.54, 4.04)	1033 (5.89, 5.78)	946 (4.72, 5.17)	1184 (7.55, 7.42)	751 (2.29, 3.85)	952 (1.27, 2.56)	930 (5.66, 5.96)	1204 (5.93, 5.64)	824 (4.38, 4.42)	997 (5.62, 3.74)
	Gender		881 (6.13, 6.36)	1131 (8.59, 9.52)	786 (2.38, 3.71)	1009 (5.95, 6.39)	863 (1.88, 4.02)	1098 (6.17, 7.93)	729 (0.56, 1.90)	869 (2.70, 6.34)	831 (3.57, 5.23)	1108 (9.09, 6.82)	764 (3.37, 4.26)	968 (4.55, 4.85)
Low Trait Cheerfulness	Emotion		877 (3.14, 4.39)	1136 (6.10, 7.94)	820 (1.44, 3.50)	999 (2.44, 3.80)	950 (4.50, 4.99)	1118 (4.36, 4.87)	774 (1.34, 2.72)	914 (0.52, 1.76)	914 (4.24, 4.99)	1142 (4.63, 4.02)	824 (2.36, 2.93)	964 (2.95, 2.17)
	Gender		789 (3.19, 4.46)	1036 (5.29, 5.17)	730 (1.82, 3.78)	961 (4.77, 5.00)	817 (2.52, 3.55)	1050 (2.55, 3.86)	701 (1.40, 2.87)	878 (1.07, 2.50)	825 (3.03, 4.12)	1045 (3.28, 4.35)	758 (3.59, 4.49)	930 (4.29, 4.54)

P, preparation; NP, no preparation.

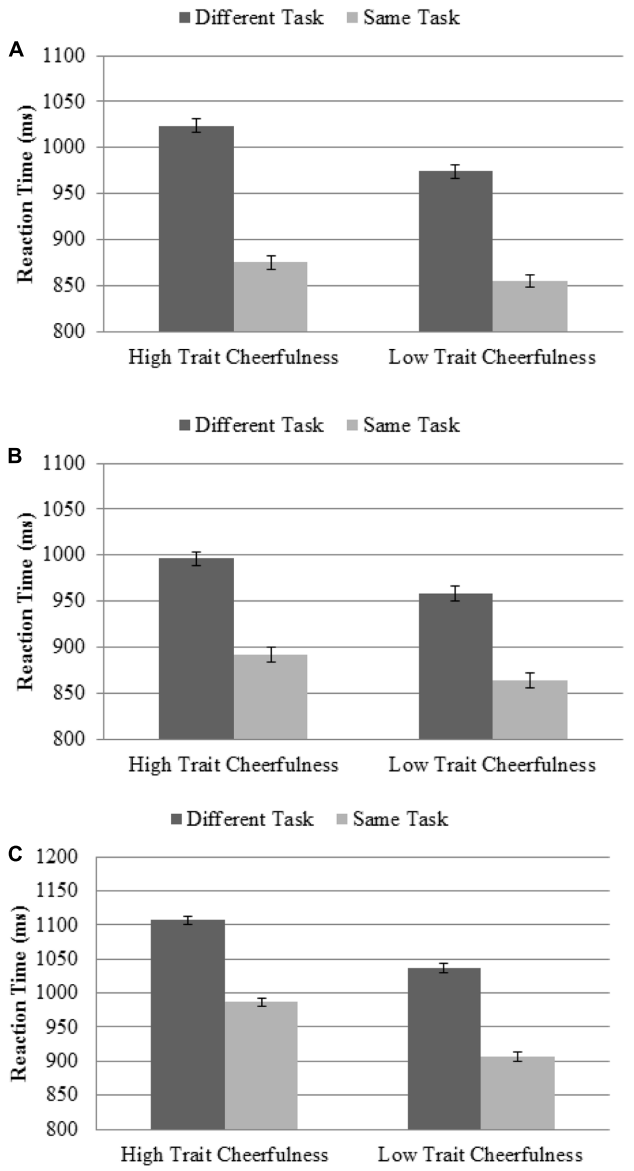


FIGURE 2 | Effect of the task-switching cost as a function of trait cheerfulness group for: **(A)** Study 1; **(B)** Study 2, preparation part; and **(C)** Study 2, no preparation part. Note the lack of differences between both groups. If anything, the general trend is in the opposite direction, as individuals with high trait cheerfulness showed a higher task-switching cost (measured as the difference between a task being repeated or not) than individuals with low trait cheerfulness. The error bars represent the standard error of the mean, with variability between participants removed by means of Coussineau's method.

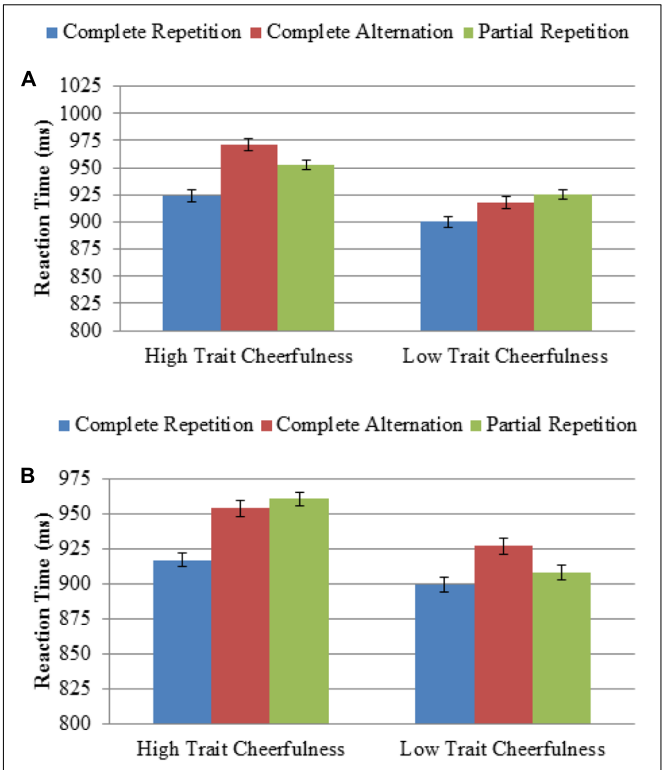


FIGURE 3 | Effect of attribute repetition as a function of the trait cheerfulness group for: **(A)** Study 1; and **(B)** Study 2, preparation part. Both graphs reveal that participants with high trait cheerfulness showed a higher repetition effect than those with low trait cheerfulness, measured as an increased difference when all attributes were repeated compared to no repetition or partial repetition. The error bars represent the standard error of the mean, with variability between participants removed by means of Coussineau's method.

or emotion) were repeated in consecutive trials, compared to when they were repeated, which generated the highest task-switching cost (101 ms vs. 182 ms). In addition, the task-switching cost was lower when the task involved recognizing the gender than when it required recognizing the emotion (116 ms vs. 154 ms), and in the preparation conditions compared to those in which there was no preparation cue (103 ms vs. 167 ms).

More directly related to our main goal, and perhaps most importantly, we did not find any evidence of a lower task-switching cost in the high trait cheerfulness group (see **Figure 2**). In fact, we observed a non-significant trend in RT, $F(1,46) = 2.23$, $p = 0.14$, $\eta^2 = 0.05$, in the opposite direction (149 ms task-switching cost in the high trait cheerfulness group, compared to 120 ms cost in the low trait cheerfulness group).

Interestingly, however, group was found to modulate other relevant variables. For example, the Group × Repetition interaction was significant, $F(2,92) = 3.30$, $p = 0.041$, $\eta^2 = 0.07$. Specifically, the previously described effect of repetition (i.e., faster responses when all attributes were repeated than when none were repeated) was present to a greater extent in the high trait cheerfulness group compared to the low trait cheerfulness group (47 ms vs. 18 ms; see **Figure 3**). The Group × Task × Preparation interaction was also significant, $F(1,46) = 7.54$, $p = 0.009$, $\eta^2 = 0.14$, showing a higher preparation effect in the high versus low trait cheerfulness group, although this was only observed in the emotion recognition task [$F(1,46) = 5.31$, $p = 0.026$, $\eta^2 = 0.10$, 239 ms vs. 185 ms] and not in the gender recognition task ($F < 1$).

The analysis of EP showed significant main effects in the variables Repetition, $F(2,92) = 14.32$, $p < 0.001$, $\eta^2 = 0.24$, and

Preparation, $F(1,46) = 17.68$, $p < 0.001$, $\eta^2 = 0.28$. Overall, the pattern was very similar to that observed in RT: participants made fewer errors when the stimuli attributes were repeated than when they were not repeated or were only partially repeated (2.8% vs. 4.5% vs. 4.4%), and also when a cue was provided anticipating the demand to perform (3.2% vs. 4.7%). Again, our task replicated the predicted effects of task-switching cost, $F(1,46) = 42.23$, $p < 0.001$, $\eta^2 = 0.48$: participants made fewer errors when the task was repeated in two consecutive trials (2% task-switching cost). Furthermore, as expected, this effect was significantly modulated by attribute repetition, $F(2,92) = 4.35$, $p = 0.016$, $\eta^2 = 0.09$, and marginally modulated by task type, $F(1,46) = 3.36$, $p = 0.073$, $\eta^2 = 0.07$. Specifically, we observed a higher task-switching cost when all attributes were repeated than when no attributes were repeated or when only some were repeated (2.9% vs. 2.2% vs. 1%). We also observed a trend toward a higher cost when the task to perform was expressed emotion recognition (2.5% vs. 1.6%).

Regarding our main goal, the analysis revealed a main effect of Group $(1,46) = 6.80$, $p = 0.012$, $\eta^2 = 0.13$, which reflected that individuals with high trait cheerfulness had a higher EP than those with low trait cheerfulness (4.7% vs. 3.1%). We also observed a significant interaction between Group × Task × Task Change, $F(1,46) = 5.52$, $p = 0.023$, $\eta^2 = 0.11$. The interaction revealed that individuals with high trait cheerfulness showed a higher task-switching cost than those with low trait cheerfulness, although this only applied to the gender recognition task (2.6% vs. 0.5%), not to the emotion recognition task (2.4% vs. 2.7%).

Additionally, a higher effect of preparation was observed in individuals with high versus low trait cheerfulness (2.2% vs. 0.8%) regardless of the task, as reflected by the marginally significant Group × Preparation interaction, $F(1,46) = 3.90$, $p = 0.054$, $\eta^2 = 0.08$ (see **Figure 4**).

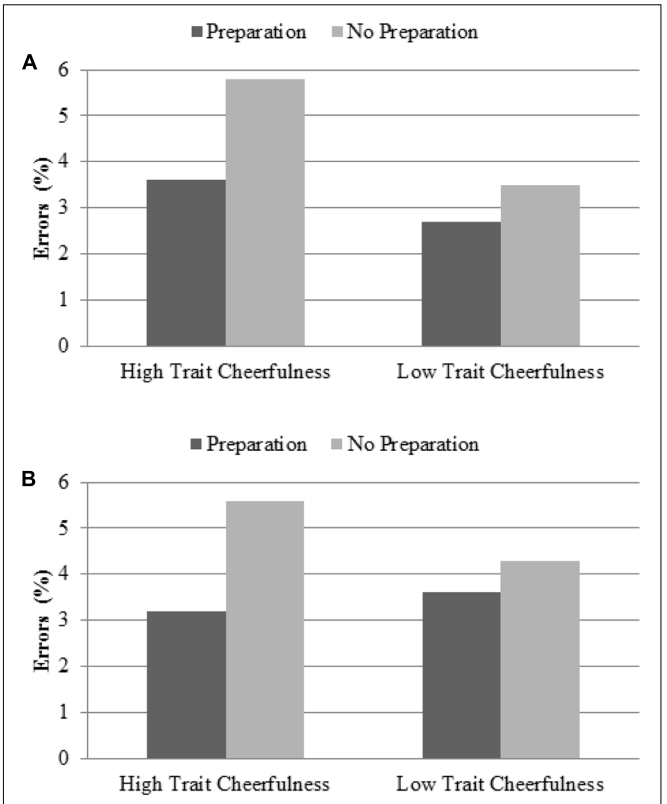


FIGURE 4 | Effect of preparation as a function of the trait cheerfulness group for: **(A)** Study 1; and **(B)** Study 2, preparation part. Both graphs reveal that participants with high trait cheerfulness showed a higher effect of preparation than those with low trait cheerfulness, measured as a greater reduction in error percentage when the target was preceded by a signal that anticipated the upcoming demand than when it was not present.

Discussion

In this study, our aim was to replicate the modulation of cognitive flexibility by trait cheerfulness observed in a previous study and further analyze these relationships. Results proved that the task-switching paradigm we used was an efficient instrument to study this process, since the usual task-switching cost pattern was observed (e.g., modulation by task type, attribute repetition, and preparation cue). However, it is important to note that, overall, our data reflected that individuals with high trait cheerfulness do not seem to show a lower task-switching cost than those with low trait cheerfulness. If anything, the little evidence collected indicated the opposite, as the EP results revealed a higher, not lower, task-switching cost in individuals with high trait cheerfulness in the gender recognition task. The pattern observed in RT followed the same trend, although differences were not significant. Hence, our result pattern did not support the idea of a link between trait cheerfulness and a lower task-switching cost and thus increased cognitive flexibility.

However, we did observe significant effects of group with regard to the repetition of the stimuli attributes and the prior preparation to them. Specifically, individuals with higher trait cheerfulness showed a larger effect of stimuli repetition and a larger effect of task preparation, particularly in the expressed emotion recognition task. We consequently decided to carry out a second study with the goal of verifying if, indeed, trait cheerfulness did not modulate the task-switching cost, and also of exploring whether the effects of repetition and preparation were consistent.

STUDY 2

Considering the findings of Study 1, we conducted Study 2 to further explore whether trait cheerfulness modulated the task-switching cost, and studying whether it was possible to replicate the modulation by trait cheerfulness of the repetition of the stimuli attributes and the preparation to the stimuli. A previous study had produced some evidence suggesting that individuals with high trait cheerfulness show a lower task-switching cost compared to individuals with low trait cheerfulness (López-Benítez et al., unpublished). Yet, this effect was not replicated in Study 1. This could be due to the presence of a demand anticipating cue in half of the trials, given that, if the participant has sufficient preparation, the effect of task-switching cost as a function of trait cheerfulness may diminish or even disappear.

Note that, in the previous study, no preparation cue was presented.

Therefore, the present study had two parts (of four blocks each) that were counterbalanced. Half of the blocks followed the same structure as in Study 1, but in the other half the demand anticipating cue was eliminated (as in López-Benítez et al., unpublished). If the determining factor in the differential effect of task-switching cost as a function of trait cheerfulness is anticipation of the demand, we hypothesized that participants with high versus low trait cheerfulness will show a lower task-switching cost (i.e., higher cognitive flexibility) in an experiment in which the demand is not anticipated. Furthermore, in line with Study 1, we expected to find a higher effect of both attribute repetition and preparation to the task in individuals with high trait cheerfulness than in those with low trait cheerfulness.

Material and Methods

Participants

Following the same method as in Study 1, 48 new students from the University of Granada were selected out of 569 people[2]. In this case, the high trait cheerfulness group was made up of 25 participants (19 women, mean age 22.36 years, $SD = 4.37$, cut-off score ≥ 3.50), while the low trait cheerfulness group comprised 23 participants (19 women, mean age 21.83 years, $SD = 3.42$, cut-off score ≤ 2.63). All the participants had normal or corrected-to-normal vision, performed the task voluntarily, and received course credit in exchange for participating. They signed an informed consent and had the possibility to stop the experimental session without any consequences. Data from one participant were not taken into account because the number of correct responses was low compared to the group (more than 2.5 SD below the group mean). Again, the study was part of a broader research project, approved by the Ethics Committee of the University of Granada, in accordance with the 1964 Declaration of Helsinki.

Stimuli and Procedure

The stimuli and procedure were the same as in Study 1, with two exceptions. First, instead of being composed of eight similar blocks, the study was divided into two distinct parts, each of which comprised four blocks. The first part was the same as in Study 1, but in the second part no pre-target cue was given to indicate the upcoming task. Both parts were counterbalanced between groups. Second, in order to maintain the alertness level of participants, an audio feedback signal was used every time a wrong response or no response was given.

Design

The data were analyzed using SPSS 21.0 statistical software. We decided to analyze this study separately depending on whether the trials with a previous preparation condition were mixed with those that did not have any (preparation part), or there was rather no mix between trials (no preparation part). We used the same design as in Study 1 in the blocks in which there was a possibility of preparing for the demand: 2 (Group; High Trait Cheerfulness vs. Low Trait Cheerfulness) × 2 (Task; Emotion vs. Gender) × 3 (Repetition; Complete Alternation vs. Complete Repetition vs. Partial Repetition) × 2 (Task Change; Different vs. Same) × 2 (Preparation; Preparation vs. No Preparation). The same design was used for the analysis of the blocks of trials in which there was no possibility of preparing for the demand, with the sole exclusion of the preparation variable: 2 (Group; High Trait Cheerfulness vs. Low Trait Cheerfulness) × 2 (Task; Emotion vs. Gender) × 3 (Repetition; Complete Alternation vs. Complete Repetition vs. Partial Repetition) × 2 (Task Change; Different vs. Same). Again, RT, which was calculated only for correct responses that were also preceded by correct responses, and EP were analyzed as dependent variables.

[2]The total sample of 569 people was composed by the 244 people from Study 1, and 325 new participants. From the total sample, only participants with extreme scores and who did not perform the experimental task in Study 1 were selected.

Results

Analysis of the Preparation Part

Descriptive statistics are shown on **Table 2**. The analysis revealed a main effect of each of the within-participant variables: Task, $F(1,45) = 52.53$, $p < 0.001$, $\eta^2 = 0.54$, Repetition, $F(2,90) = 13.51$, $p < 0.001$, $\eta^2 = 0.23$, and Preparation, $F(1,45) = 261.45$, $p < 0.001$, $\eta^2 = 0.85$. As in Study 1, participants were faster to respond when the task was gender recognition (877 ms vs. 978 ms), when all the attributes were repeated between two consecutive trials, compared to no or partial attribute repetition (908 ms vs. 940 ms vs. 934 ms), and when a cue was used to anticipate the demand (810 ms vs. 1044 ms). Once again, our procedure additionally showed the expected task-switching cost effects, $F(1,45) = 74.24$, $p < 0.001$, $\eta^2 = 0.62$, meaning that participants' responses were faster when the task was repeated in two consecutive trials (99 ms task-switching cost). This effect was modulated by attribute repetition, $F(2,90) = 24.66$, $p < 0.001$, $\eta^2 = 0.35$, preparation conditions, $F(1,45) = 30.17$, $p < 0.001$, $\eta^2 = 0.40$, and task type, $F(1,45) = 9.71$, $p = 0.003$, $\eta^2 = 0.18$. Thus, the task-switching cost was lower when none of the stimuli attributes (i.e., gender or emotion) were repeated in consecutive trials than when they were repeated; the latter condition generated the highest task-switching cost (46 ms vs. 166 ms). The task-switching cost was also lower in the preparation conditions (63 ms vs. 135 ms) and when the task was gender recognition (75 ms vs. 127 ms). In addition, the lower task-switching cost in preparation conditions was modulated by attribute repetition, $F(2,90) = 8.34$, $p < 0.001$, $\eta^2 = 0.16$, as this effect was lower when only some or none of the stimuli attributes were repeated between two consecutive trials than when all the attributes were repeated (27 ms vs. 49 ms vs. 145 ms).

Regarding our goal, and as shown in **Figure 2**, no evidence was found of a lower task-switching cost in individuals with high versus low trait cheerfulness ($F < 1$). However, we replicated the modulation of attribute repetition by trait cheerfulness, as reflected in the Group × Repetition interaction, $F(2,92) = 3.30$, $p = 0.041$, $\eta^2 = 0.07$. This confirmed that, compared to individuals with low trait cheerfulness, those with high trait cheerfulness showed a higher effect of repetition when all the attributes were repeated between two consecutive trials than

TABLE 2 | Mean reaction time (in ms), error percentage, and error standard deviation in each of the experimental conditions as a function of trait cheerfulness.

		Complete alternation						Complete repetition						Partial repetition					
		Different			Same			Different			Same			Different			Same		
		P	NP	NPP	P	NP	NPP	P	NP	NPP	P	NP	NPP	P	NP	NPP	P	NP	PNP
High Trait Cheerfulness	Emotion	915 (5.00, 7.43)	1219 (8.78, 10.08)	1123 (5.80, 7.50)	804 (1.64, 4.63)	1063 (4.21, 7.86)	1059 (4.11, 7.75)	891 (8.82, 10.10)	1174 (6.97, 16.63)	1178 (5.12, 6.83)	804 (1.90, 5.28)	937 (0.97, 3.42)	931 (1.99, 3.70)	936 (5.24, 5.77)	1187 (8.20, 11.13)	1148 (5.07, 6.67)	847 (4.36, 5.75)	1081 (6.92, 6.33)	1034 (4.23, 5.11)
	Gender	756 (3.13, 6.02)	1056 (4.15, 9.97)	1057 (3.53, 4.50)	785 (2.00, 4.20)	1032 (7.30, 11.03)	1008 (4.94, 6.31)	819 (3.13, 6.02)	1089 (4.79, 9.91)	1062 (4.80, 5.82)	752 (1.44, 4.00)	869 (1.74, 5.02)	909 (2.09, 4.48)	811 (2.90, 5.46)	1099 (4.86, 5.97)	1073 (5.76, 7.71)	736 (3.30, 4.94)	989 (3.69, 5.88)	978 (5.21, 6.03)
Low Trait Cheerfulness	Emotion	886 (4.17, 8.74)	1142 (6.39, 8.74)	1098 (8.51, 9.00)	860 (3.92, 9.16)	1049 (4.14, 8.80)	968 (4.44, 6.24)	915 (5.48, 6.74)	1180 (7.27, 9.72)	1078 (5.62, 7.17)	790 (2.13, 6.49)	926 (1.28, 3.30)	876 (1.06, 3.47)	904 (5.57, 6.13)	1129 (5.94, 6.66)	1109 (6.14, 7.25)	831 (4.46, 6.00)	1009 (7.50, 9.36)	984 (3.51, 4.73)
	Gender	732 (2.62, 7.39)	1002 (2.18, 4.94)	964 (4.33, 5.47)	750 (2.35, 5.29)	991 (3.56, 8.07)	898 (4.01, 4.42)	790 (3.89, 7.98)	1072 (3.16, 6.59)	988 (4.98, 6.96)	691 (3.31, 8.35)	827 (1.82, 5.88)	824 (2.91, 4.69)	746 (3.00, 6.28)	996 (4.73, 6.05)	978 (4.70, 5.13)	701 (2.30, 4.12)	947 (3.28, 5.06)	892 (4.10, 6.47)

P, preparation; NP, no preparation; NPP, no preparation part.

when only some of them were repeated (44 ms vs. 10 ms; see **Figure 3**).

Error percentage analysis revealed significant main effects in the following variables: Task, $F(1,45) = 10.86$, $p = 0.002$, $\eta^2 = 0.19$, Repetition, $F(2,90) = 3.13$, $p = 0.049$, $\eta^2 = 0.07$, and Preparation, $F(1,45) = 11.57$, $p = 0.001$, $\eta^2 = 0.20$. In general, the pattern was very similar to that observed in RT and with that observed in Study 1. In fact, participants made fewer errors when the task was gender recognition (3.3% vs. 5%), when the stimuli attributes were repeated, compared to no repetition or partial repetition (3.5% vs. 4.3% vs. 4.8%), and when a cue was given anticipating the demand (3.4% vs. 4.9%). Once more, we observed the expected effects of task-switching cost, $F(1,45) = 14.90$, $p < 0.001$, $\eta^2 = 0.25$, reflected in a higher accuracy when the task was repeated in two consecutive trials (1.5% task-switching cost). Additionally, and as expected, this effect was significantly modulated by attribute repetition, $F(2,90) = 4.91$, $p = 0.010$, $\eta^2 = 0.10$, and by task type, $F(1,45) = 5.83$, $p = 0.020$, $\eta^2 = 0.12$. In this regard, we found that the task-switching cost was higher when all the attributes were repeated, compared to no repetition or partial repetition (3.3% vs. 0.6% vs. 0.6%), and when the task was expressed emotion recognition (2.4% vs. 0.5%).

Regarding our main goal, no evidence was found that trait cheerfulness modulated the effect of task-switching cost ($F < 1$). However, as observed in Study 1, the Group × Preparation interaction was found to be marginally significant, $F(1,45) = 3.70$, $p = 0.061$, $\eta^2 = 0.08$, replicating the trend toward a higher overall effect of preparation in participants with high versus low trait cheerfulness (2.4% vs. 0.7%; see **Figure 4**).

Analysis of the No Preparation Part

Descriptive statistics are shown on **Table 2**. As in the previous studies, the analysis revealed a main effect of each of the within-participant variables: Task, $F(1,45) = 22.95$, $p < 0.001$, $\eta^2 = 0.34$, and Repetition, $F(2,90) = 17.59$, $p < 0.001$, $\eta^2 = 0.28$. Specifically, participants were faster when the task was gender recognition (969 ms vs. 1049 ms) and when all attributes between two consecutive trials were repeated, as opposed to no repetition or partial repetition of attributes (981 ms vs. 1022 ms vs. 1025 ms). Once again, our study showed that participants were faster when the task was repeated between two consecutive trials (124 ms task-switching cost), $F(1,45) = 185.69$, $p < 0.001$, $\eta^2 = 0.81$. As expected, this effect was again modulated by task type, $F(1,45) = 12.60$, $p = 0.001$, $\eta^2 = 0.22$, and attribute repetition, $F(2,90) = 28.72$, $p < 0.001$, $\eta^2 = 0.39$. In this regard, the task-switching cost was lower when the task was gender recognition (101 ms vs. 148 ms) and also when none of the stimuli attributes (i.e., gender or emotion) were repeated in consecutive trials, compared to when they were repeated, which generated the highest task-switching cost (78 ms vs. 192 ms).

With regard to our main goal, as can be seen in **Figure 2**, individuals with high trait cheerfulness did not show a lower task-switching cost than those with low trait cheerfulness ($F < 1$). In fact, cheerfulness did not modulate any other variable, such as repetition ($F < 1$).

The accuracy analysis revealed a main effect of the Repetition variable, $F(2,90) = 5.13$, $p = 0.008$, $\eta^2 = 0.10$, that is, participants made fewer errors when all the stimuli attributes were repeated between two trials than when none were repeated (3.6% vs. 5%). As expected, accuracy increased when the task was repeated in two consecutive trials, $F(1,45) = 23.85$, $p < 0.001$, $\eta^2 = 0.35$, showing a 1.8% task-switching cost. This effect was also modulated by task type, $F(1,45) = 9.11$, $p = 0.004$, $\eta^2 = 0.17$, and marginally modulated by attribute repetition, $F(2,90) = 2.68$, $p = 0.074$, $\eta^2 = 0.06$. In other words, the task-switching cost was lower when the task was gender recognition (3.6% vs. 5%) and also when no (or only some) attributes were repeated, compared to complete attribute repetition (1.1% vs. 1.1% vs. 3.1%).

As happened with RT, individuals with high trait cheerfulness did not show a lower task-switching cost than individuals with low trait cheerfulness ($F < 1$). We did not find any relationship with other relevant variables either ($F < 1$).

Discussion

The goal of this study was to study whether individuals with high trait cheerfulness showed a lower task-switching cost by exploring whether this modulation could be caused by the presentation of a cue anticipating the demand and hence the response. We also intended to verify whether the higher effect of attribute repetition and task preparation in participants with high trait cheerfulness found in Study 1 was replicated.

As in Study 1, Study 2 confirmed the suitability of the task for the study of task-switching cost. Again, our data did not provide evidence that individuals with higher trait cheerfulness showed higher cognitive flexibility, measured as a lower task-switching cost, than those with low trait cheerfulness.

However, although only in the preparation part, individuals with high trait cheerfulness again displayed both a larger effect of attribute repetition between two consecutive trials, and a larger effect of task preparation, thus replicating the findings of Study 1.

GENERAL DISCUSSION

The main aim of this research was to study the modulation of cognitive flexibility processes by trait cheerfulness, as a temperamental basis of sense of humor (Ruch et al., 1996, 1997), by using a task-switching paradigm. Although the procedure used showed the typical effects of task-switching cost, the results reflected that high trait cheerfulness people did not show a lower task-switching cost, that is, a better cognitive flexibility compared to low trait cheerfulness individuals.

Some authors have pointed out the potential benefits of positive emotions in areas such as cognition (see, for example, Lyubomirsky et al., 2005, for a review). Specifically, it has been observed that positive affect reduces the task-switching cost in a paradigm with no emotional implications (i.e., task-switching between color and shape; Yang and Yang, 2014). It has also shown that some personality characteristics may benefit (DeYoung et al., 2005) or impair (Compton, 2000; Campbell et al., 2011) performance on cognitive flexibility tasks. Additionally, previous research has suggested that cognitive flexibility processes could be involved in contexts where people have to detect or enjoy humor (Polimeni et al., 2010; Weiss et al., 2013) as well as in situations in which emotion regulation strategies are applied to alter an event's affective impact on people's affective state (Malooly et al., 2013; Gul and Khan, 2014). Therefore, considering that trait cheerfulness is a positive predisposition to detect, produce, enjoy, and maintain humoristic stimuli as well as positive emotions (Ruch and Hofmann, 2012), it has been positively associated with personality variables related to a better cognitive flexibility (Carretero-Dios et al., 2014), and with a better coping with negative emotions (Papousek and Schulter, 2010). On the other hand, considering the results of our previous study, it could then be inferred that individuals with high trait cheerfulness should have a lower task-switching cost, that is, a higher cognitive flexibility, compared to individuals with low trait cheerfulness. Our findings, however, did not confirm this hypothesis.

From a personality perspective, our results could be partially explained. On the one hand, it is true that trait cheerfulness is closely related to extraversion (Ruch and Köhler, 2007; Carretero-Dios et al., 2014), which is negatively associated with the performance in tasks that involve cognitive flexibility (Campbell et al., 2011). This fact could justify that individuals characterized by high trait cheerfulness did not show higher cognitive flexibility in our study (as measured by a lower task-switching cost). If anything, our results indicated the opposite trend, i.e., a higher task-switching cost for high trait cheerfulness people. Moreover, trait cheerfulness is also positively linked to openness and agreeableness, and negatively related to neuroticism (Carretero-Dios et al., 2014), which promote (Jensen-Campbell et al., 2002; DeYoung et al., 2005) and impair (Compton, 2000), respectively, cognitive flexibility. In this sense, high trait cheerfulness people should have a greater ability to shift their mental set when they are working on different tasks. However, this is not the case.

In addition, it has been sometimes reported that positive states do not have benefits on cognitive processes (Mitchell and Phillips, 2007). For example, some studies have failed to find a clear pattern of task-switching cost reduction when a motivational intensity induction is carried out (high interest) compared to negative emotional states or a control condition (Zhou and Siu, 2015). Others studies have not found a clear pattern of benefits from positive affective induction in multitasking conditions either (Morgan and D'Mello, 2016). Contradictory results were also observed by Phillips et al. (2002), who revealed a poorer performance after a positive affective state induction, compared with a neutral induction, in task-switching conditions between naming the color versus the word in Stroop tasks. Yet, they found a smaller difference between alternation and non-alternation conditions in a verbal fluency task (i.e., alternating or not between saying words starting with a specific letter and words from a specific semantic category).

Furthermore, it is important to note the nature of the task and how cognitive flexibility is measured. In our studies, flexibility is assessed as the ability to change between mental sets for adapting to new demands in a cognitive task. In this sense, it might be possible that the cognitive nature of this task involves cognitive flexibility processes different from those that are relevant to recognize humor (Weiss et al., 2013), which would

be more associated to cheerfulness. In addition, trait cheerfulness is related to a greater coping with and recovery from negative emotions (see Papousek and Schulter, 2010; Ruch and Hofmann, 2012), which has also been associated with cognitive flexibility processes (Malooly et al., 2013). In a recent study (López-Benítez et al., under review), it has been found that people with high versus low trait cheerfulness frequently use reappraisal strategies in their daily lives. However, they did not have a better ability to apply reappraisal strategies for down- regulating negative emotions. In this sense, if our task is testing cognitive flexibility as ability rather than a general use of it, it could be thought that the frequent use of reappraisal strategies is not enough to also have a greater cognitive flexibility. In any case, future studies are needed to test these hypotheses.

The present pattern of results could be influenced by the sample size. To solve this limitation, we carried out an omnibus analysis with data from Study 1 ($N = 49$), Study 2 ($N = 48$), and our unpublished study ($N = 46$), with a total of 72 high trait cheerfulness participants and 71 low trait cheerfulness participants. Only trials that appeared in all studies were selected, that is, trials where there was not a prior preparation to the task, given that the preparation condition was presented in some studies but not in others. The mixed ANOVA with Task Change (Different vs. Same) as within participants variable and Group (High Trait Cheerfulness vs. Low Trait Cheerfulness) and Study as between participants variables showed a complete absence of Group × Task Change interaction, $F(1,137) = 0.07$, $p = 0.791$, $\eta^2 = 0.00$.

In order to see whether this absence of evidence could be taken as evidence for absence of modulation of group over the task-switching costs, a Bayesian approach was used. This procedure assesses how much support we could obtain for the null hypothesis through the Bayes Factor (BF_{10}), which represents how strongly a result supports our hypothesis (i.e., lower task-switching cost for high trait cheerfulness people, or Group × Task Change interaction, H_1) over the null hypothesis (i.e., no Group × Task Change interaction). Three ranges of values for BF_{10} are commonly accepted to interpret the output: (a) evidence of the absence of an effect (from 0 up to 0.33); (b) inconclusive evidence (from 0.33 up to 3); (c) evidence of an effect (from 3 and up). The Bayesian analysis was carried out with JASP Team (2017). Our results indicated, again reflecting that task-switching cost did not depend on trait cheerfulness, substantial evidence for a null effect ($BF_{10} = 0.186$, for the Group × Task Change interaction or a *t*-test comparing the two groups on the task-switching cost).

A tentative explanation of the present results is related to the subject of this research and the demands required by the task itself. Cheerfulness is a positive affective predisposition associated with sense of humor (Ruch and Köhler, 2007). It is therefore related to the manifestation, enhancement, and maintenance of positive emotions, along with a lower manifestation of negative emotions and a higher resilience to them (Zweyer et al., 2004; Papousek and Schulter, 2010). This endows it with qualities that are very closely linked to processes of an emotional nature, such as induction processes, regulation, and emotional intelligence (e.g., Ruch, 1997; Yip and Martin, 2006), and processes more related to social interaction and empathy (e.g., Ruch and Köhler, 2007; Beermann and Ruch, 2009). From this viewpoint, given the affective, humoristic, communicative, expressive, and social characteristics that compound trait cheerfulness, it might be possible that trait cheerfulness has a higher predictive power and play a relevant role in tasks that involve processes of this nature, compared to cognitive tasks which do not include elements typical of humoristic, emotional, or social stimulation. Further research needs to be carried out in this field to clarify these ideas.

In addition, and although this was not our main goal, in Study 1 and in the preparation part of Study 2 we observed that, compared to individuals with low trait cheerfulness, those with high trait cheerfulness showed a higher effect of attribute repetition between two consecutive trials (e.g., Hommel, 2004). They also showed a tendency toward a higher effect of preparation when presented with a cue anticipating the demand in a trial that immediately followed (e.g., Kiesel et al., 2010) that was even higher in the expressed emotion recognition task (Study 1).

To our knowledge, no studies have explored the modulation of the effects of attribute repetition by predisposition to affective states (or affective states themselves). However, if our findings are confirmed, it may be possible to explain them in terms of the broaden-and-build theory (Fredrickson, 2001). According to this approach, positive states often lead to a more holistic processing of the context, thus expanding the attention focus (see, for example, Johnson et al., 2010). Taking into account that trait cheerfulness is a predisposition toward positive affective states, it could be inferred that individuals with high trait cheerfulness are defined by a more global processing style. In this sense, even if all participants were to benefit from attribute repetition between consecutive trials and from a cue anticipating the next demand, it would be possible to theorize that, due to their more global mindset configuration, individuals with high trait cheerfulness benefit more from these facilitation effects, having the information on the demand to carry out more active in their short-term memory, which would improve their immediate response, particularly in the expressed emotion recognition task (Study 1), which is considered more complicated (e.g., Egner et al., 2008; Ochsner et al., 2009).

Another possible explanation is derived from affective induction contexts. In a recent study, López-Benítez et al. (in press) have found that, compared to low trait cheerfulness people, individuals characterized by high trait cheerfulness experienced a larger affective state change as a consequence of watching amusing and sad stimuli. The authors interpreted this finding as larger affective sensitivity or permeability to the environment, thus promoting some psychological, social, and physical benefits in high trait cheerfulness individuals (e.g., Yip and Martin, 2006; Carretero-Dios et al., 2014; Delgado-Domínguez et al., 2016). In this sense, it is possible that the presentation of a cue to anticipate the task might be a powerful element to capture and focus high cheerfulness people's attention, promoting a better permeability (larger preparation effects) to it. Moreover, this fact also would explain, at least partially, the larger attribute repetition effect for high trait cheerfulness individuals that was only significant when a cue that prepares to a subsequent demand was displayed.

Therefore, from this point of view, it might be possible that high trait cheerfulness individuals have a higher receptivity to process useful and relevant nuances and contextual cues, which could help them to a better adaptation to the environment. In any case, future studies should replicate and extend these findings in order to understand the role of trait cheerfulness on these phenomena.

Notwithstanding the importance of the results, our study had some limitations. First, as pointed out above, participants in our studies were selected according to their trait cheerfulness scores. Ruch et al. (1996, 1997) suggest that the temperamental basis of sense of humor have two manifestations, as traits and as states, which are closely related to one another. Clear dissociations have been observed between traits and states, which have differential modulation effects on attentional processes in other areas such as anxiety (Pacheco-Unguetti et al., 2010). In addition, and following Fredrickson's (2001) theoretical proposal, a positive affective induction rather than a positive trait might have a greater impact in aspects such as cognitive resources. Therefore, it would be interesting to verify whether the induction of state cheerfulness, as opposed to the selection of participants with high trait cheerfulness, would have the same effects as those caused by trait cheerfulness or if, on the other hand, participants' state at the time of the task is a more powerful predictive factor to explain cognitive flexibility. Moreover, further research is needed to assess whether other elements of sense of humor are relevant for making predictions on this type of processes. For example, taking into account that the task used here might put participants in a telic state of mind, that is, more goal oriented (Apter and Smith, 1977) and assuming that seriousness is described from a cognitive, attitudinal, and reflexive perspective, trait or, even more importantly, state seriousness may modulate to a greater extent the effect of these processes, which have a more cognitive nature. Additionally, based on studies that have found a relationship between negative affective states and a poorer performance in multitasking conditions, which require high cognitive flexibility (Morgan and D'Mello, 2016), it could be inferred that bad mood, through its affective properties, may also modulate cognitive flexibility, leading to a lower task-switching cost.

Second, taking into account that trait cheerfulness is linked to personality characteristics that may affect the performance on tasks that require cognitive flexibility (e.g., Compton, 2000; Jensen-Campbell et al., 2002), they should be incorporated in future studies together with related variables such as, for example, optimism, to observe their differential weight in cognitive tasks compared to trait cheerfulness.

Finally, assuming the conceptualization of the cheerfulness construct (for a review, see Ruch and Hofmann, 2012), it might be more interesting to analyze the modulation of emotional induction processes by cheerfulness, in its trait and state manifestation, in the presence not only of positive but also of negative emotions. It would also be interesting to explore its possible relationship with emotion regulation strategies, which are involved in these processes with the goal of modifying the affective response experienced by an individual.

In short, two studies were conducted in this research to verify whether individuals with high trait cheerfulness, compared to those with low trait cheerfulness, showed higher cognitive flexibility, manifested as a lower task-switching cost. The results did not confirm this scenario. This is important taking into account that a relation between this cheerfulness and cognitive flexibility can be predicted from the literature on both humor and cheerfulness. Nevertheless, individuals with high versus low trait cheerfulness showed higher effects of attribute repetition and task preparation. Therefore, although replication of this finding seems necessary, it suggests a new path of exploration. The higher permeability to contextual cues of high cheerfulness individuals shown in the current and previous studies (López-Benítez et al., in press) could underlay a better adaptation to the environment that calls for future research. In addition, new studies should analyze whether these effects can be generalized to other cognitive processes such as creativity while exploring the modulation of affective processes by cheerfulness.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: RL-B, JL, AA, HC-D. Performed the experiments: RL-B. Analyzed the data: RL-B, JL, AA, HC-D. Interpreted the data and drafted the manuscript: RL-B, JL, AA, HC-D. All authors read and accepted the final manuscript submitted for publication.

FUNDING

This research is part of the doctoral dissertation by RL-B, and it was supported by the Spanish Ministerio de Educación, Cultura, y Deporte with a predoctoral grant (FPU-AP2012-1806) and with the Spanish grants of PSI2014-52764-P, from Ministerio de Economía, Industria, y Competitividad (MINECO), and PSI2013-45567P from Dirección General de Investigación Científica y Técnica-Ministerio de Educación y Ciencia (DGICYT-MEC).

REFERENCES

Apter, M. J., and Smith, K. C. P. (1977). "Humour and the theory of psychological reversals," in *It's a Funny Thing, Humour*, eds A. J. Chapman and H. C. Foot (Oxford: Pergamon Press), 95–100.

Baumann, N., and Kuhl, J. (2005). Positive affect and flexibility: overcoming the precedence of global over local processing of visual information. *Motiv. Emot.* 29, 123–134. doi: 10.1007/s11031-005-7957-1

Beermann, U., and Ruch, W. (2009). How virtuous is humor? What we can learn from current instruments. *J. Posit. Psychol.* 22, 395–417. doi: 10.1080/17439760903262859

Campbell, A. M., Davalos, D. B., McCabe, D. P., and Troup, L. J. (2011). Executive functions and extraversion. *Pers. Individ. Dif.* 51, 720–725. doi: 10.1016/j.paid.2011.06.018

Carretero-Dios, H., Benítez, I., Delgado-Rico, E., Ruch, W., and López-Benítez, R. (2014). Temperamental basis of sense of humor: the Spanish long form of the

trait version of the State-Trait-Cheerfulness-Inventory. *Pers. Individ. Dif.* 68, 77–82. doi: 10.1016/j.paid.2014.03.045

Carretero-Dios, H., Eid, M., and Ruch, W. (2011). Analyzing multitrait-multimethod data with multilevel con?rmatory factor analysis: an application to the validation of the State–Trait Cheerfulness Inventory. *J. Res. Pers.* 45, 153–164. doi: 10.1016/j.jrp.2010.12.007

Compton, R. J. (2000). Ability to disengage attention predicts negative affect. *Cogn. Emot.* 14, 401–415. doi: 10.1080/026999300378897

Crone, E. A., Wendelken, C., Donohue, S. E., and Bunge, S. A. (2006). Neural evidence for dissociable components of task switching. *Cereb. Cortex* 16, 475–486. doi: 10.1093/cercor/bhi127

Davidson, M. C., Amso, D., Anderson, L. C., and Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* 44, 2037–2078. doi: 10.1016/j.neuropsychologia.2006.02.006

Delgado-Domínguez, C. J., Escudero-Contreras, A., Font-Ugalde, P., Ruiz-Vílchez, D., Collantes-Estévez, E., and Carretero-Dios, H. (2016). Upswings in cheerful mood and disease activity in patients with rheumatoid arthritis. *Int. J. Behav. Med.* 23, 606–610. doi: 10.1007/s12529-016-9550-2

DeYoung, C. G., Peterson, J. B., and Higgins, D. M. (2005). Sources of openness/intellect: cognitive and neuropsychological correlates of the fifth factor of personality. *J. Pers.* 73, 825–858. doi: 10.1111/j.1467-6494.2005.00330.x

Egner, T., Etkin, A., Gale, S., and Hirsch, J. (2008). Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. *Cereb. Cortex* 18, 1475–1484. doi: 10.1093/cercor/bhm179

Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attent. Percept. Psychophys.* 16, 143–149. doi: 10.3758/BF03203267

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., and Hirsch, J. (2006). Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* 51, 871–882. doi: 10.1016/j.neuron.2006.07.029

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: the broaden-and-build theory of positive emotions. *Am. Psychol.* 56, 218–226. doi: 10.1037//0003-066X.56.3.218

Goeleven, E., De Raedt, R., Leyman, L., and Verschuere, B. (2008). The Karolinska directed emotional faces: a validation study. *Cogn. Emot.* 22, 1094–1118. doi: 10.1080/02699930701626582

Gul, A., and Khan, K. (2014). Emotion regulation strategies can predict task-switching abilities in euthymic bipolar patients. *Front. Hum. Neurosci.* 8:847. doi: 10.3389/fnhum.2014.00847

Hommel, B. (1998). Event files: evidence for automatic integration of stimulus-response episodes. *Vis. Cogn.* 5, 183–216. doi: 10.1080/713756773

Hommel, B. (2004). Event files: feature binding in and across perception and action. *Trends Cogn. Sci.* 8, 494–500. doi: 10.1016/j.tics.2004.08.007

JASP Team (2017). *JASP (Version 0.8.1.1) [Computer Software]*. Available at: Jasp-Stats. Org

Jensen-Campbell, L. A., Rosselli, M., Workman, K. A., Santisi, M., Rios, J. D., and Bojan, D. (2002). Agreeableness, conscientiousness, and effortful control processes. *J. Res. Pers.* 36, 476–489. doi: 10.1016/S0092-6566(02)00004-1

Johnson, K. J., Waugh, C. E., and Fredrickson, B. L. (2010). Smile to see the forest: facially expressed positive emotions broaden cognition. *Cogn. Emot.* 24, 299–321. doi: 10.1080/02699930903384667

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24, 175–219. doi: 10.1016/0010-0285(92)90007-O

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., et al. (2010). Control and interference in task switching—A review. *Psychol. Bull.* 136, 849–874. doi: 10.1037/a0019842

López-Benítez, R., Acosta, A., Lupiáñez, J., and Carretero-Dios, H. (in press). High trait cheerfulness individuals are more sensitive to the emotional environment. *J. Happiness Stud.* doi: 10.1007/s10902-017-9871-0

Lundqvist, D., Flykt, A., and Öhman, A. (1998). *The Karolinska Directed Emotional Faces – KDEF (CD ROM).* Stockholm: Karolinska Institute, Department of Clinical Neuroscience, Psychology Section.

Lyubomirsky, S., King, L., and Diener, E. (2005). The benefits of frequent positive affect: does happiness lead to success? *Psychol. Bull.* 131, 803–855. doi: 10.1037/0033-2909.131.6.803

Malooly, A. M., Genet, J. J., and Siemer, M. (2013). Individual differences in reappraisal effectiveness: the role of affective flexibility. *Emotion* 13, 302–313. doi: 10.1037/a0029980

Marzecová, A., Bukowski, M., Correa, Á., Boros, M., Lupiáñez, J., and Wodniecka, Z. (2013). Tracing the bilingual advantage in cognitive control: the role of flexibility in temporal preparation and category switching. *J. Cogn. Psychol.* 25, 586–604. doi: 10.1080/20445911.2013.809348

Mitchell, R. L., and Phillips, L. H. (2007). The psychological, neurochemical and functional neuroanatomical mediators of the effects of positive and negative mood on executive functions. *Neuropsychologia* 45, 617–629. doi: 10.1016/j.neuropsychologia.2006.06.030

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Monsell, S. (2003). Task switching. *Trends Cogn. Sci.* 7, 134–140. doi: 10.1016/S1364-6613(03)00028-7

Morgan, B., and D'Mello, S. K. (2016). The influence of positive vs. negative affect on multitasking. *Acta Psychol.* 170, 10–18. doi: 10.1016/j.actpsy.2016.06.006

Murdock, K. W., Oddi, K. B., and Bridgett, D. J. (2013). Cognitive correlates of personality. *J. Individ. Differ.* 34, 97–104. doi: 10.1027/1614-0001/a000104

Ochsner, K. N., Hughes, B., Robertson, E. R., Cooper, J. C., and Gabrieli, J. D. (2009). Neural systems supporting the control of affective and cognitive conflicts. *J. Cogn. Neurosci.* 21, 1841–1854. doi: 10.1162/jocn.2009.21129

Pacheco-Unguetti, A. P., Acosta, A., Callejas, A., and Lupiáñez, J. (2010). Attention and anxiety different attentional functioning under state and trait anxiety. *Psychol. Sci.* 21, 298–304. doi: 10.1177/0956797609359624

Papousek, I., and Schulter, G. (2010). "Don't take an X for a U. why laughter is not the best medicine, but being more cheerful has many benefits," in *Psychological Well-being*, ed. I. Wells (Hauppauge, NY: Nova Science Publishers), 1–75.

Phillips, L. H., Bull, R., Adams, E., and Fraser, L. (2002). Positive mood and executive function: evidence from stroop and fluency tasks. *Emotion* 2, 12–22. doi: 10.1037/1528-3542.2.1.12

Polimeni, J. O., Campbell, D. W., Gill, D., Sawatzky, B. L., and Reiss, J. P. (2010). Diminished humour perception in schizophrenia: relationship to social and cognitive functioning. *J. Psychiatr. Res.* 44, 434–440. doi: 10.1016/j.jpsychires.2009.10.003

Posner, M. I., and Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annu. Rev. Psychol.* 58, 1–23. doi: 10.1146/annurev.psych.58.110405.085516

Ruch, W. (1997). State and trait cheerfulness and the induction of exhilaration: a FACS study. *Eur. Psychol.* 2, 328–341. doi: 10.1027/1016-9040.2.4.328

Ruch, W., and Hofmann, J. (2012). "A temperament approach to humor," in *Humor and Health Promotion*, ed. P. Gremigni (Hauppauge, NY: Nova Science Publishers), 79–113.

Ruch, W., and Köhler, G. (2007). "A temperament approach to humor," in *The Sense of Humor: Explorations of a Personality Characteristic*, ed. W. Ruch (Berlin: Mouton de Gruyter), 203–230.

Ruch, W., Köhler, G., and van Thriel, C. (1996). Assessing the "humorous temperament": construction of the facet and standard trait forms of the State-Trait-Cheerfulness-Inventory—STCI. *Humor* 9, 303–339. doi: 10.1016/j.neulet.2008.09.017

Ruch, W., Köhler, G., and van Thriel, C. (1997). To be in good or bad humor: construction of the state form of the State-Trait-Cheerfulness-Inventory—STCI. *Pers. Individ. Dif.* 22, 477–491. doi: 10.1016/S0191-8869(96)00231-0

Schneider, W., Escaman, A., and Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburg, PA: Psychology Software Tools, Inc.

Schuch, S., Werheid, K., and Koch, I. (2012). Flexible and inflexible task sets: asymmetric interference when switching between emotional expression, sex, and age classification of perceived faces. *Q. J. Exp. Psychol.* 65, 994–1005. doi: 10.1080/17470218.2011.638721

Simon, J. R. (1969). Reactions toward the source of stimulation. *J. Exp. Psychol.* 81, 174–176. doi: 10.1037/h0027448

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651

Suls, J. A. (1972). "A two-stage model for the appreciation of jokes and cartoons: an information-processing analysis," in *Psychology of Humor: Theoretical Perspectives and Empirical Issues*, eds J. H. Goldstein and P. E. McGhee (New York, NY: Academic Press), 81–100.

Tsoi, D. Y., Lee, K. H., Gee, K. A., Holden, K. L., Parks, R. W., and Woodruff, P. W. (2008). Humour experience in schizophrenia: relationship with executive dysfunction and psychosocial impairment. *Psychol. Med.* 38, 801–810. doi: 10.1017/S0033291707002528

Wadlinger, H. A., and Isaacowitz, D. M. (2006). Positive mood broadens visual attention to positive stimuli. *Motiv. Emot.* 30, 87–99. doi: 10.1007/s11031-006-9021-1

Weiss, E. M., Gschaidbauer, B. C., Samson, A. C., Steinbäcker, K., Fink, A., and Papousek, I. (2013). From ice age to Madagascar: appreciation of slapstick humor in children with Asperger's syndrome. *Humor* 26, 423–440. doi: 10.1515/humor-2013-0029

Yang, H., and Yang, S. (2014). Positive affect facilitates task switching in the dimensional change card sort task: implications for the shifting aspect of executive function. *Cogn. Emot.* 28, 1242–1254. doi: 10.1080/02699931.2013.879053

Yip, J. A., and Martin, R. A. (2006). Sense of humor, emotional intelligence, and social competence. *J. Res. Pers.* 40, 1202–1208. doi: 10.2466/PR0.66.1.51-58

Zhou, Y., and Siu, A. F. (2015). Motivational intensity modulates the effects of positive emotions on set shifting after controlling physiological arousal. *Scand. J. Psychol.* 56, 613–621. doi: 10.1111/sjop.12247

Zweyer, K., Velker, B., and Ruch, W. (2004). Do cheerfulness, exhilaration, and humor production moderate pain tolerance? A FACS study. *Humor* 17, 85–119. doi: 10.1515/humr.2004.009

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 López-Benítez, Carretero-Dios, Acosta and Lupiáñez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 November 2017
doi: 10.3389/fpsyg.2017.01973

Are Clowns Good for Everyone? The Influence of Trait Cheerfulness on Emotional Reactions to a Hospital Clown Intervention

*Sarah Auerbach**

Personality and Assessment, Department of Psychology, University of Zurich, Zürich, Switzerland

Trait cheerfulness predicts individual differences in experiences and behavioral responses in various humor experiments and settings. The present study is the first to investigate whether trait cheerfulness also influences the impact of a hospital clown intervention on the emotional state of patients. Forty-two adults received a clown visit in a rehabilitation center and rated their emotional state and trait cheerfulness afterward. Facial expressions of patients during the clown visit were coded with the *Facial Action Coding System*. Looking at the total sample, the hospital clown intervention elicited more frequent facial expressions of genuine enjoyment (Duchenne smiles) than other smiles (Non-Duchenne smiles), and more Duchenne smiles went along with more perceived funniness, a higher level of global positive feelings and transcendence. This supports the notion that overall, hospital clown interventions are beneficial for patients. However, when considering individual differences in the receptiveness to humor, results confirmed that high trait cheerful patients showed more Duchenne smiles than low trait cheerful patients (with no difference in Non-Duchenne smiles), and reported a higher level of positive emotions than low trait cheerful individuals. In summary, although hospital clown interventions on average successfully raise the patients' level of positive emotions, not all patients in hospitals are equally susceptible to respond to humor with amusement, and thus do not equally benefit from a hospital clown intervention. Implications for research and practitioners are discussed.

Keywords: trait cheerfulness, Facial Action Coding System, Duchenne smile, hospital clown, amusement, transcendence

OPEN ACCESS

Edited by:
René T. Proyer, Martin Luther University of Halle-Wittenberg, Germany and University of Zurich, Switzerland

Reviewed by:
William Larry Ventis, College of William & Mary, United States
Alberto Dionigi, Federazione Nazionale Clown Dottori (FNC), Italy

***Correspondence:**
Sarah Auerbach
sarahsina.auerbach@uzh.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *24 August 2017*
Accepted: *27 October 2017*
Published: *13 November 2017*

Citation:
Auerbach S (2017) Are Clowns Good for Everyone? The Influence of Trait Cheerfulness on Emotional Reactions to a Hospital Clown Intervention. Front. Psychol. 8:1973. doi: 10.3389/fpsyg.2017.01973

INTRODUCTION

Whereas in some situations all people behave more or less in the same way, in other situations, individual differences co-determine people's actions and reactions. Research in the field of positive psychology has shown that in situations designed to promote happiness and well-being, the fit between a person's personality and the type of activity is in part responsible for its success (Schueller, 2012; Senf and Liau, 2013). The present study focuses on hospital clown interventions[1], which aim at bringing positive experiences to ailing patients. Hospital clown interventions have

[1] Also referred to as clown therapy, clown visit, medical clowning, or clown care.

been described to "represent a particular way of using humor in order to promote people's well-being" (Dionigi et al., 2012, p. 1). The idea goes back to Freud (1960), who described humor as a tool that allows the individual to face adversity. In a situation normally associated with negative emotions (such as a hospital stay), humor can help the individual to cope with the situation by providing an alternative perspective on the situation. Although the art of clowning does not solely consist of humor (e.g., Peacock, 2009), humor has frequently been characterized as the main component of hospital clowning (Dionigi et al., 2012), and thus hospital clown interventions have been defined as humorous interventions[2] (Ruch and Hofmann, 2017).

Although to date, hospital clown interventions are widely used in hospitals, nursing homes and other care facilities, and research has shown some positive effects for patients (see Effects of Hospital Clown Interventions on Individuals), no study has investigated whether these humorous interventions are beneficial for all recipients, or whether some groups of individuals benefit more than others (only age or gender differences were tested so far; e.g., Fernandes and Arriaga, 2010; Vagnoli et al., 2010). Hence, no research is available on whether individual differences influence the effects of a hospital clown intervention on the emotional reactions of patients.

Individual Differences in Emotional Reactions to Humor

Research on personality and humor demonstrates that people habitually differ in the way they cognitively evaluate humorous stimuli (Ruch and Hehl, 2007), use and communicate humor in everyday life (Craik et al., 1996; Fox et al., 2016), and emotionally respond to humor (Ruch, 2007; Platt et al., 2013; Ruch et al., 2015). The predominant emotional reaction to humor was labeled *exhilaration*[3] (or amusement), which in classifications of emotions is defined as a facet of joy (Ruch, 1993). One personality trait in particular, *trait cheerfulness*, has been studied in a variety of humor experiments and settings as a stable disposition for cheerful mood states and the easiness with which amusement is induced. Trait cheerfulness is characterized by a prevalence of cheerful mood, a low threshold for smiling and laughter, a composed view of adverse life circumstances, a broad range of active elicitors of cheerfulness and smiling and laughter, and a generally cheerful interaction style (Ruch et al., 1996). Together with trait seriousness and trait bad mood, it forms the temperamental basis for the sense of humor (Ruch and Carrell, 1998). Trait cheerfulness can be classified into the higher-order dimension of extraversion (Carretero-Dios et al., 2014), but has a higher specificity in predicting the intensity of amusement in response to humor than extraversion (Ruch, 1997). Ruch et al. (2011) postulated five relationships between trait cheerfulness and a cheerful state: high trait cheerful individuals have a lower threshold, a higher intensity, a longer duration, a higher robustness of cheerful mood (even when facing adversity), and a faster mood recovery (after a mood alteration to the negative) than low trait cheerful individuals. These postulates were tested in various experiments and contexts using subjective as well as objective methods, such as the observation of facial signs, to infer on the emotional state (for an overview see Ruch and Hofmann, 2012).

The universal facial expression of enjoyable emotions is smiling (Ekman, 2003). Research has repeatedly shown that there are different types of smiles, but especially one type (*Duchenne smile*) is a valid indicator of genuine enjoyment (Ekman et al., 1990; Sauter et al., 2013). It is characterized by the joint and timely corresponding contraction of the zygomatic major muscle (pulling the lip corners up) and the lateral part of the orbicularis oculi muscle (contracting the region around the eye producing crow's feet). Other types of smiles occur in situations without genuinely felt enjoyment (*Non-Duchenne smiles*). These types of smiles are present, for example, when individuals mask a negative emotional state (masking smile) or smile when nothing much is felt (phony smile) but individuals attempt to appear as if positive emotions are felt (Ekman and Friesen, 1982; Harris and Alvarado, 2005). The different types of smiles can be assessed with an objective and reliable technique for coding observable facial actions, the *Facial Action Coding System* (FACS; Ekman et al., 2002a), which enables coding the frequency, intensity, timing, duration, laterality and symmetry of 44 different action units. In a series of studies the FACS was used as an objective measure of amusement to demonstrate the influence of trait cheerfulness on the emotional reaction to humorous stimuli. For example, during an interaction with a clowning experimenter, individuals high in trait cheerfulness showed more frequent, more intense and longer lasting signs of facial amusement (Duchenne smiling and laughter[4]) than individuals low in trait cheerfulness (Ruch, 1997). When high trait cheerful individuals saw their own distorted photographs as a surprise, they showed more frequent Duchenne smiling and laughter than low trait cheerful individuals (Beermann and Ruch, 2011). When a virtual companion was present during a funny film, high trait cheerful individuals had higher frequencies of Duchenne laughter than low trait cheerful individuals (Hofmann et al., 2015).

Effects of Hospital Clown Interventions on Individuals

To date, a few studies have evaluated hospital clown interventions and have consequently shown that hospital clown interventions can have a beneficial effect on patients (mostly children). For example, studies found a reduction of preoperative anxiety and worries in children undergoing medical procedures when interacting with a clown pair compared to a control group without a clown visit (e.g., Golan et al., 2009; Fernandes and Arriaga, 2010; Vagnoli et al., 2010). Regarding changes in positive states after interacting with hospital clowns, one

[2]Ruch and Hofmann (2017) argue that there are different types of humor interventions (such as self-administered individual humor interventions stemming from a positive psychology tradition, humor training programs to strengthen an individuals' level of the sense of humor, or hospital clown interventions).

[3]Exhilaration has been defined as the main emotional response to humor and denotes either the process of making cheerful or the temporary rise and fall of a cheerful state. To exhilarate in this sense means to make cheerful, or to amuse (Ruch, 1993).

[4]Duchenne laughter typically occurs at higher levels of reported amusement, while Duchenne smiling occurs at lower levels of reported amusement (Ruch, 1993).

study found an increase in self-rated positive affect in children (Fernandes and Arriaga, 2010), and another study found an increase in self- and parent reported well-being (Pinquart et al., 2011).

Only two studies have examined the positive emotions elicited by hospital clowns in individuals in more detail. Auerbach et al. (2014) developed and tested the *29 Clown Emotion List* (CLEM-29), which is a collection of single adjectives and short phrases, but can be reduced to four factors: amusement, transcendence, unease and arousal. The factor of amusement merges a variety of positive humor-related states including a calmer cheerfulness and a more aroused hilarity. Transcendence was defined according to its non-religious connotation as the feeling of being uplifted and surpassing the ordinary. It includes positive feelings induced by clowns such as feeling privileged, appreciated, connected to the clown and elevated. The negative factor of unease consisted of negative feelings induced by clowns (e.g., threatened, fearful, confused). Ratings of the factor of arousal relate to different states of arousal, in which positive loadings (touched and speechless) refer to a calm state, that is, low arousal, whereas negative loadings (overexcited and *schadenfreude*) refer to a more heightened arousal. Studies that used the CLEM-29 showed that individuals watching videos of (Auerbach et al., 2014) and patients interacting with (Auerbach et al., 2016) a hospital clown reported a higher level of amusement compared to individuals who watched or experienced a nurse intervention. Furthermore, in both samples a combination of amusement and transcendence best predicted the total amount of positive affect after a hospital clown intervention. The authors concluded that a hospital clown intervention induces not only the typical humor reaction in recipients (amusement), but also adds a unique quality to the clown-patient interaction (transcendence).

In summary, previous research has provided evidence that hospital clown interventions are a suitable method to enhance the emotional state of individuals. The studies used subjective assessment tools, either self-reports or external reports of the key variables. So, the next step in a comprehensive evaluation of hospital clown interventions is to validate the subjectively assessed state of patients during the clown intervention by including observable signs of non-verbal behavior. Research (e.g., Ruch and Hofmann, 2012) has shown that humorous stimuli successfully generate facial amusement in various experiments if the subjects experience amusement (objective and subjective markers of amusement are typically moderately related; Ruch, 1995). Platt et al. (2013) showed that of the 16 enjoyable emotions proposed by Ekman (2003), amusement was one facet of joy that went along with both Duchenne smiles and Duchenne laughter. It is assumed that amusement will be the enjoyable emotion most elicited by hospital clown interventions, going along with Duchenne smiles and laughter.

Another still unnoted issue in evaluations of hospital clown interventions is the personality influence. Taking into account the trait cheerfulness model and its empirical evidence (Ruch and Hofmann, 2012), it can be assumed that high trait cheerful individuals benefit more from the intervention (i.e., more positive emotions) than low trait cheerful individuals.

Aims and Hypotheses

The present study aims to contribute to a better understanding of hospital clown interventions in three ways: the investigation of patient's facial signs of enjoyment during an interaction with a hospital clown, its relationship to their subjective states, and the replication of the theory of trait cheerfulness as predictor of the emotional reaction of patients to humorous stimuli. The first hypothesis is that the hospital clown intervention on average elicits Duchenne smiles more often than Non-Duchenne smiles. The second hypothesis is that higher frequencies of Duchenne smiles are associated with higher levels of a positive experience, and lower levels of a negative one, whereas higher frequencies of Non-Duchenne smiles are associated with lower levels of positive experiences, and higher levels of negative ones. The third hypothesis is that high trait cheerful individuals show more Duchenne smiles and less Non-Duchenne smiles, and simultaneously report higher levels of positive emotions than low trait cheerful individuals.

MATERIALS AND METHODS

Sample

The sample consisted of $N = 42$ adult German speaking patients from a physical rehabilitation center (81% male), and was a convenient sample. Patients suffered from paraplegia, amputations, or other multiple injuries. The age of patients ranged from 19 to 75 years ($M = 45.36$, $SD = 16.56$). Inclusion criteria were age 18 or older, voluntary participation, not bedridden, and being cognitively and physically able to participate in the study. Patients were filmed during the study, and videos of a subsample of 26 patients could be used for coding facial actions.

Instruments

The standard trait version of the *State-Trait-Cheerfulness Inventory* (STCI-T < 60 >; Ruch et al., 1996) consists of 60 items to reliably and validly assess trait cheerfulness, trait seriousness and trait bad mood. To compose the *trait cheerfulness* scale in the current study, eight items were selected representing the facets of a low threshold for smiling and laughter and a generally cheerful interaction style (hilarity[5]; e.g., "I am a merry person"). The answer format is a four-point Likert-scale ranging from 1 (strongly disagree) to 4 (strongly agree) and Cronbachs alpha was 0.87.

The *29 Clown Emotion List* (CLEM-29; Auerbach et al., 2014) is a list of 29 adjectives and short phrases assessing emotional states in the context of clowning. Participants rate their current state on a 7-point Likert scale ranging from 1 (=not at all) to 7 (=very strongly). As the sample size in the present study is too small to test the hypotheses with all single ratings, factor scores were used instead (transcendence, uneasiness, amusement, and arousal; the procedure is described in detail in Auerbach

[5]The present study addresses emotional reactions to clowns, which are a trigger of hilarity. Hence, only items representing facets of hilarity were selected to compose trait cheerfulness.

et al., 2016), which are sensitive enough to capture changes in clown-induced emotional states (Auerbach et al., 2014).

The *Hospital Study Evaluation Form* (HSEF; Auerbach et al., 2016) contains 22 single ratings, of which seven ratings concern the stay in the care facility (HSEF-General; e.g., quality of meals, care) and the evaluation of the hospital clown intervention (HSEF-Current; e.g., global positive and negative feelings during the situation). The answer format is a 7-point Likert scale. A second set of 15 single ratings (HSEF-Preferences), which are related to patients' general preferences for clowns, was given to patients at the end of the study (e.g., general liking of clowns; 5-point Likert scale).

Procedure

Prior to the study, the local ethics committee approved the study. Consent forms were handed out before and after the experiment. The core of the current study was a surprise visit from a hospital clown pair. The study took place in a separate room in the rehabilitation center, and the procedure was highly standardized. Patients were recruited with the cover story that they were going to participate in an evaluation of patient satisfaction in hospitals. They were also told that a staff member would conduct a routine assessment, which they were to evaluate afterward. Two patients participated in each trial. Patients first filled out the HSEF-General, followed by a baseline assessment of emotional states (CLEM-29, HSEF-Current). Afterward, the clown intervention of a predetermined length took place (Min = 4.00, Max = 8.85, $M = 6.65$, $SD = 1.17$). It consisted of a semi-standardized performance of a hospital clown pair (one male clown with 17 years of experience, and one female clown with 16 years of experience), aiming at the induction of a positive emotional state in the patients. The same clowns performed in all trials, and used the same roles, clothes and make-up. They worked according to a script and did the same performance (same punchlines) in every trial. They were instructed to limit the length of the interaction to about 5–8 min. Both wore a red nose. The male clown carried a ukulele, wore a Doctor-like jacket. The female clown wore a yellow dirndl dress with yellow socks, a pink blouse. She had an abnormally large handbag in one hand filled with requisites; e.g., a pig nose that makes a farting sound when squeezed, and a thimble, used to demonstrate a magic trick together. The clown pair behaved like Auguste and Whiteface: the female clown was more dominant, slightly aggressive, bossy and pompous, while the male clown was the foolish, clumsy and more sensitive partner. After the clowns left the room, patients filled out the state measures (CLEM-29, HSEF-Current). Patients subsequently were debriefed about the real aim of the study (to investigate emotional reactions to hospital clowns) and asked not to disclose the use of clowns to other patients until the study was completed. For the last step of the study, they filled out the trait measures (STCI-T < 60 >, HSEF-Preferences).

Full color, digital format films with a close-up view of the patients' face were recorded. To be able to code the same clown-patient interactions for all subjects, ten standardized scenes occurring in all trials (about 10–20 s long) were extracted, each containing a studied punch line produced by the clowns followed by the reaction of patients. A certified FACS coder[6] coded the resulting 260 observations (26 patients with 10 scenes each) with the help of the FACS (Ekman et al., 2002a). A Duchenne smile was defined as a symmetric and timely coincidental movement of the orbicularis oculi muscle around the eye (AU6) and zygomatic major muscle at the corners of the mouth (AU12). It could be accompanied by a tightening of the eyelids (AU7) and mouth opening (AU25, AU26, AU27), but no other action unit[7] (Ekman and Friesen, 1982). The Non-Duchenne smile was defined as AU12 alone, or AU12 plus further action units that are associated with negative feelings (Ekman et al., 2002b). Laughter vocalizations were coded using one of four codes: "single unvoiced (ch)," "single voiced (ha)," "multiple unvoiced (ch ch ch)," or "multiple voiced (ha, ha, ha)."

RESULTS

Three scores were built for use in the analyses. As they were sum scores over ten different standardized scenes during the interaction between the clowns and a patient, Cronbach's alpha was calculated for each score as a measure of the homogeneity of behaviors during the ten scenes. A frequency score for enjoyment smiles was built by summing up all Duchenne smiles in ten scenes, which showed high internal consistency ($\alpha = 0.80$). A frequency score for Non-Duchenne smiles was built by summing up all Non-Duchenne smiles in the same ten scenes ($\alpha = 0.58$). The Non-Duchenne smile category was more heterogeneous than the Duchenne smile one, as it comprised different types of Non-Duchenne smiles. A laughter score was built by summing up all four types of laughter vocalizations in ten standardized scenes ($\alpha = 0.75$). All variables used in the analyses were normally distributed.

Frequency of Different Types of Smiles

Patients on average smiled 8.92 times ($SD = 3.76$) during the 10 scenes. The percentage of Duchenne smiles among all smiles was 76.29%. The minimum was zero Duchenne smiles; the maximum was 16 ($M = 6.81$, $SD = 3.74$). The minimum of Non-Duchenne smiles was zero; the maximum was eight ($M = 2.12$, $SD = 2.01$). Patients on average laughed 3.58 times during the ten scenes ($SD = 4.14$) with a maximum of 15 laughter vocalizations. Thirty percent of patients did not produce any laughter vocalizations during the selected scenes.

Relationship between Subjective and Objective Assessment

Negative affect after the clown visit was very low ($M = 1.73$, $SD = 1.32$), and positive affect was high ($M = 5.12$, $SD = 1.5$;

[6]The coder proved high reliability (agreement index = 0.78) scoring video material from real interactions in the FACS final test.

[7]In few cases, high intensity combinations of AU12 and AU6 (intensity scores of D-E on a scale from A – E) during apex were accompanied by nose wrinkling (AU9) and/or eyebrow-lowering frowning (AU4). Following Hofmann (2014), these smiles were classified as high intensity Duchenne smiles. In the offset of the AU12, few patients pulled down their lip corners (AU15) or slightly pressed their lips together (AU24). There was no time overlap with the apexes of the AU12 and thus they were seen as regulatory mechanism (Ekman et al., 2002b).

scale from 1 to 7). Patients enjoyed participating in the study to a high extent ($M = 4.24$, $SD = 0.77$; scale from 1 to 5), and 81.5% stated that they felt better after the clown visit. The frequency of Duchenne smiles was positively correlated with funniness of the clown visit ($r = 0.57$, $p < 0.01$), global positive feelings ($r = 0.46$, $p < 0.01$), transcendence ($r = 0.40$, $p < 0.05$) and the joy of participating in the study ($r = 0.43$, $p < 0.05$), and negatively correlated with global negative feelings after the clown visit ($r = -0.38$, $p < 0.05$). The frequency of Non-Duchenne smiles was negatively correlated with the joy of participating in the study ($r = -0.62$, $p < 0.01$), transcendence ($r = -0.59$, $p < 0.01$), feeling better after the clown visit ($r = -0.33$), and positively correlated to unease ($r = 0.34$, both marginally not significant, $p = 0.06$). Laughter vocalizations were positively correlated with Duchenne smiles ($r = 0.37$, $p < 0.05$), transcendence ($r = 0.46$, $p < 0.05$), amusement ($r = 0.44$, $p < 0.05$), funniness of the clown visit ($r = 0.46$, $p < 0.01$), and feeling better after the clown visit ($r = 0.41$, $p < 0.05$).

The Influence of Trait Cheerfulness

Next, it was tested whether high trait cheerful individuals had higher levels of positive emotions during the clown intervention than low trait cheerful individuals. To build two groups of equal sizes, ten patients with the lowest scores were allocated to group 1 (low trait cheerful), and ten patients with the highest scores to group 2 (high trait cheerful). A 2 × 2 repeated measures ANOVA with trait cheerfulness (high vs. low) and type of smile (Duchenne smile vs. Non-Duchenne smile) was computed for the frequency of smiling. Results are displayed in **Figure 1**.

Patients showed more Duchenne smiles than Non-Duchenne smiles, $F(1,18) = 31.58$, $p < 0.001$, $\eta_p^2 = 0.64$, and high trait cheerful individuals smiled more frequently than low trait cheerful individuals, $F(1,18) = 6.90$, $p < 0.05$, $\eta_p^2 = 0.28$. The interaction just failed to be significant, $F(1,18) = 2.84$, $p = 0.11$. However, there was a numerical trend toward higher levels of Duchenne smiles in the high trait cheerful group ($M = 8.60$, $SD = 2.32$) than in the low trait cheerful group ($M = 5.30$, $SD = 3.13$). An independent samples t-test confirmed that the two groups significantly differed in their frequency of Duchenne smiles, $t(18) = -2.68$, $p < 0.05$. No difference was found for Non-Duchenne smiles, $t(18) = -0.30$, $p = 0.77$.

Individuals high in trait cheerfulness reported higher positive feelings, $t(31) = -2.35$, $p < 0.05$, higher funniness ratings of the clowns, $t(31) = -2.82$, $p < 0.01$, higher levels of transcendence (marginally not significant), $t(28) = -1.75$, $p = 0.09$, and a lower level of unease, $t(28) = 3.09$, $p < 0.01$, than individuals low in trait cheerfulness. The two groups did not differ in their general preference for clown performances, $t(31) = -1.54$, $p = 0.14$, and laughter, $t(18) = -0.70$, $p = 0.49$.

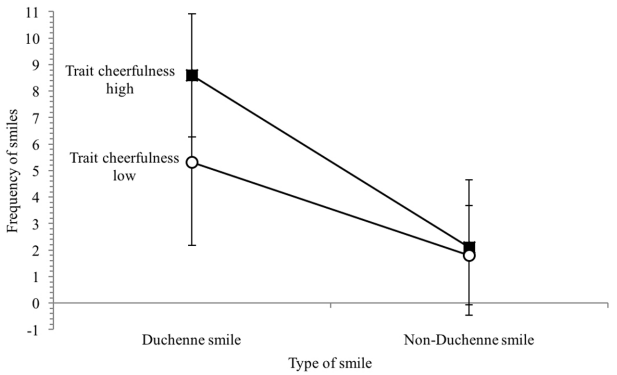


FIGURE 1 | Frequencies of smiles during the hospital clown intervention for individuals high and low in trait cheerfulness.

DISCUSSION

Humor interventions have been used frequently in research to increase happiness and lower depression in various settings and samples, including hospital clown interventions (for an overview see Ruch and Hofmann, 2017). One consistent finding stemming from humor research is that individuals habitually differ in their readiness to react with amusement to humorous stimuli (Ruch and Hofmann, 2012). However, this has never been tested in patients receiving a hospital clown visit. Hence, the present study was the first to investigate individual differences in the emotional state of patients in response to a hospital clown intervention, and to use the FACS as a comprehensive, reliable technique for the objective assessment of the patient's emotions. This made it possible to distinguish between Duchenne smiles (genuine expressions of enjoyment) and other smiles in patients during clown-patient interactions.

First, the results confirmed that both types of smiles can occur during a humorous intervention (Harris and Alvarado, 2005), but eight out of ten smiles were Duchenne smiles, which is associated (and was positively correlated) with a positive emotional state (Ekman, 2003). In the present study, the facial expression of enjoyment was not only highly related to funniness ratings of the hospital clown performance indicating amusement (which replicated findings from humor research; Ruch, 1995, 1997), but also positively related to the felt level of transcendence in patients (extending humor research). Hence, the present study complements the work of other researchers and practitioners who stress that hospital clown interventions are not eliciting amusement, but contribute to the elicitation of other positive experiences. Patch Adams, one of the pioneers of hospital clowning, described the work of hospital clowns as a combination of humor and *love* (Adams, 2002). Kontos et al. (2017) found that clowns working with elderly care residents apply a mixture of humor and empathy. Linge (2012) interviewed children after a hospital clown intervention and concluded that a close connection between the clown and the recipients (*magical attachment*) is a core component of a (successful) clown-patient interaction. Hospital clown interventions apparently elicit feelings that go beyond the typical humor response, such as feelings of connection, liberation, appreciation or playfulness. In this sense, the present research validates studies

using self-report measures (Auerbach et al., 2014, 2016), and strengthens the widespread assumption of practitioners and clown organizations (see Dionigi et al., 2012) that on average hospital clown interventions successfully create positive experiences and emotions for patients in need of care.

Second, another important, yet unanswered question was whether a hospital clown intervention is successful in eliciting a positive emotional state in all patients, or whether some groups of patients benefit more from the intervention than other groups. Derived from the theory of the temperamental basis of the sense of humor (Ruch et al., 1996), a trait could be identified that has been shown to be an important predictor for the emotional reaction to humorous stimuli repeatedly – trait cheerfulness (e.g., Ruch, 1997; Hofmann et al., 2015). The present study gives further validation to trait cheerfulness as predictor of positive emotions by demonstrating that a hospital clown intervention does not lead to high levels of amusement in all cases. Hence, not all patients benefit equally from the clown intervention. Clowns working in the field should always bear in mind that some patients do not want to be involved in a humorous and playful interaction, look for signs of refusal, and act accordingly. At the same time, the results can also be a justification for practitioners on a 'bad day' (e.g., in case their performance does not lead to the intended success, i.e., the patients do not smile or laugh). In fact, in many clown organizations hospital clown training includes interpersonal skills, the sensitization of the clowns to the current state of patients, and the appropriate handling of uncertainty and refusal (Dionigi et al., 2012), which seems even more important given the results presented here.

The present study has some limitations. First, only one clown pair was used. A next step could be to study possible interactions between high and low trait cheerful individuals and different kinds of clowns with different techniques (clown-person fit). The clown pair used in this study had a rather playful, interactive, hilarity-based style, while other clowns work in a more sensitive, insightful and composed way (Hofmann et al., 2014). Also, cultural differences in humor and clowning have not been studied here. Second, the sample was rather small and very heterogeneous with a wide age range and few females, which was due to the convenience sampling method. Also, only one physical rehabilitation center was included. Future research should collect larger samples more representative of hospitalized adult patients in different settings. Third, the situation was somewhat artificial – as patients were overtly filmed during the intervention – and the intervention was highly standardized, and other than in real life the subjects were committed to take part in the study. Results presented here might underestimate the true relationships between the behavior, subjective experience and personality of individuals. A next study should also aim to differentiate the different types of Non-Duchenne smiles and have a look at their correlations with different emotional states, while subjects are unobtrusively filmed. A recent study suggests that in spontaneous and unobserved situations, the emotional state of *Schadenfreude* goes along with the Duchenne smile, whereas in social situations (such as the openly filmed hospital clown intervention), subjects try to mask or suppress the expression of *Schadenfreude* (Hofmann and Ruch, 2015). It would be interesting to study the different types of smiles during a natural unobserved clown-patient interaction and during a social, observed situation, such as the one used in the present study. Fourth, although the main aim was to standardize the interactions between the clown pair and the patients, it is safe to say that not all trials were executed in exact the same manner. The clowns were instructed to perform as standardized as possible, but also as realistic as possible, meaning that in case the subject tried to interrupt the clown pair, they should not ignore him or her but react in a natural way before continuing with the scripted performance. After all, it was a real interaction between the clowns and the patients in a natural setting, and therefore not perfectly standardized. However, for the analyses only those scenes were chosen that occurred in every interaction in the same manner (same punchline), and thus the biasing effect on the results is expected to be rather small.

Despite the limitations, the results promote the use of hospital clown interventions for the enhancement of a positive emotional state in patients in need of care, but also point out the relevance of accounting for individual differences in recipients of the interventions. It is much to be hoped that this will stimulate future studies in that other researchers also combine objective and subjective assessment methods to get a clearer picture of the variety and uniqueness of emotional responses of patients during a hospital clown intervention. Furthermore, this knowledge can be used by organizations that train clowns to raise the awareness of signs that help explaining the success and failure of hospital clown interventions in their work in hospitals to prevent unwanted side effects such as the induction of negative emotions and rejection. Research demonstrates that emotional expressivity may be a reliable sign of cooperative tendency in humans (Schug et al., 2010), indicating that clowns should watch out for facial signs of emotions in patients to find out whether they want to cooperate (and thus join the game). These days, many clown organizations already include a sensitive, careful and responsible approach in the interaction with patients in their curriculum, emphasizing to always pay attention to the emotional impact of their visit to patients (Dionigi et al., 2012). Clown organizations could go one step further and specifically include the recognition and interpretation of facial expressions of individuals into their training programs.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of "Swiss Psychological Association"; and the Ethics Committee of the Department of Psychology, University of Zurich, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

SA conceived and designed the work and was also involved in data collection, data analysis and interpretation, drafting the article, critical revision of the article, and final approval of the published version.

FUNDING

This publication benefited from the support of the Red Noses Clowndoctors International Organization. The author is grateful to the Red Noses Clowndoctors International for their financial assistance.

ACKNOWLEDGMENTS

The author would like to thank Prof. Dr. Willibald Ruch, Dr. Tracey Platt, Dr. Jennifer Hofmann for helpful comments on a prior version of the manuscript. She would also like to thank the clowns of the Red Noses Clowndoctors International Organization for participating in the study, and Annette Fehling for assisting in the data collection.

REFERENCES

Adams, P. (2002). Humour and love: the origination of clown therapy. *Postgrad. Med. J.* 78, 447–448. doi: 10.1136/pmj.78.922.447

Auerbach, S., Hofmann, J., Platt, T., and Ruch, W. (2014). An investigation of the emotions elicited by hospital clowns in comparison to circus clowns and nursing staff. *Eur. J. Humour Res.* 1, 26–53. doi: 10.7592/EJHR2013.1.3.auerbach

Auerbach, S., Ruch, W., and Fehling, A. (2016). Positive emotions elicited by clowns and nurses: an experimental study in a hospital setting. *Transl. Issues Psychol. Sci.* 2, 14–24. doi: 10.1037/tps0000055

Beermann, U., and Ruch, W. (2011). Can people really "laugh at themselves?" – Experimental and correlational evidence. *Emotion* 11, 492–501. doi: 10.1037/a0023444

Carretero-Dios, H., Benítez, I., Delgado-Rico, E., Ruch, W., and López-Benítez, R. (2014). Temperamental basis of sense of humor: the Spanish long form of the trait version of the State-Trait-Cheerfulness-Inventory. *Pers. Individ. Dif.* 68, 77–82. doi: 10.1016/j.paid.2014.03.045

Craik, K. H., Lampert, M. D., and Nelson, A. J. (1996). Sense of humor and styles of everyday humorous conduct. *Humor Int. J. Humor Res.* 9, 273–302. doi: 10.1515/humr.1996.9.3-4.273

Dionigi, A., Flangini, R., and Gremigni, P. (2012). "Clowns in hospitals," in *Humor and Heath Promotion,* ed. P. Gremigni (New York, NY: Nova Science Publishers), 213–227.

Ekman, P. (2003). *Emotions Revealed. Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York, NY: St. Martin's Press.

Ekman, P., Davidson, R. J., and Friesen, W. V. (1990). The Duchenne smile: emotional expression and brain physiology II. *J. Pers. Soc. Psychol.* 58, 343–353. doi: 10.1037/0022-3514.58.2.342

Ekman, P., and Friesen, W. V. (1982). Felt, false and miserable smiles. *J. Nonverbal Behav.* 6, 238–252. doi: 10.1007/BF00987191

Ekman, P., Friesen, W. V., and Hager, J. C. (2002a). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., and Hager, J. C. (2002b). *Facial Action Coding System: Investigator's Guide*. Salt Lake City, UT: Research Nexus.

Fernandes, S. C., and Arriaga, P. (2010). The effects of clown intervention on worries and emotional responses in children undergoing surgery. *J. Health Psychol.* 15, 405–415. doi: 10.1177/1359105309350231

Fox, C. L., Hunter, S. C., and Jones, S. E. (2016). Children's humor types and psychosocial adjustment. *Pers. Individ. Dif.* 89, 86–91. doi: 10.1016/j.paid.2015.09.047

Freud, S. (1960). *Jokes and their Relation to the Unconscious*. New York, NY: Norton.

Golan, G., Tighe, P., Dobija N., Perel, A., and Keidan, I. (2009). Clowns for the prevention of preoperative anxiety in children. A randomized control trial. *J. Pediatr. Anesth.* 19, 262–266. doi: 10.1111/j.1460-9592.2008.02903.x

Harris, C. R., and Alvarado, N. (2005). Facial expressions, smile types, and self-report during humour, tickle, and pain. *Cogn. Emot.* 19, 655–669. doi: 10.1080/02699930441000472

Hofmann, J. (2014). Intense or malicious? The decoding of eyebrow-lowering frowning in laughter animations depends on the presentation mode. *Front. Psychol.* 5:1306. doi: 10.3389/fpsyg.2014.01306

Hofmann, J., Platt, T., Ruch, W., Niewiadomski, R., and Urbain, J. (2015). The influence of a virtual companion on amusement when watching funny films. *Motiv. Emot.* 39, 434–447. doi: 10.1007/s11031-014-9461-y

Hofmann, J., and Ruch, W. (2015). Gibt es ein lachen der schadenfreude? *Z. Semiotik* 37, 55–79.

Hofmann, J., Ruch, W., Platt, T., and Auerbach, S. (2014). Humor als ressource im alter: zur fähigkeit, über sich selber lachen zu können. [Humor as a resource in old age: about the ability to laugh at oneself]. *ProAlter* 4, 49–51.

Kontos, P., Miller, K. L., Mitchell, G. J., and Stirling-Twist, J. (2017). Presence redefined: the reciprocal nature of engagement between elder-clowns and persons with dementia. *Dementia* 16, 46–66. doi: 10.1177/1471301215580895

Linge, L. (2012). Magical attachment: children in magical relations with hospital clowns. *Int. J. Qual. Stud. Health Well Being* 7, 11862–11874. doi: 10.3402/qhw.v7i0.11862

Peacock, L. (2009). *Serious Play. Modern Clown Performances*. Bristol: Intellect Books.

Pinquart, M., Skolaude, D., Zaplinski, K., and Maier, R. F. (2011). Do clown visits improve psychological and sense of physical well-being of hospitalized pediatric patients? *Klin. Pädiatr.* 223, 74–78. doi: 10.1055/s-0030-1267932

Platt, T., Hofmann, J., Ruch, W., and Proyer, R. T. (2013). Duchenne display responses towards sixteen enjoyable emotions: individual differences between no and fear of being laughed at. *Motiv. Emot.* 37, 776–786. doi: 10.1007/s11031-013-9342-9

Ruch, W. (1993). "Exhilaration and humor," in *The Handbook of Emotions*, eds M. Lewis and J. M. Haviland (New York, NY: Guilford Publications), 605–626.

Ruch, W. (1995). Will the real relationship between facial expression and affective experience please stand up: the case of exhilaration. *Cogn. Emot.* 9, 33–58. doi: 10.1080/02699939508408964

Ruch, W. (1997). State and trait cheerfulness and the induction of exhilaration: a FACS study. *Eur. Psychol.* 2, 328–341. doi: 10.1027/1016-9040.2.4.328

Ruch, W. (ed.). (2007). *The Sense of Humor: Explorations of a Personality Characteristic*. New York, NY: Mouton de Gruyter. doi: 10.1515/9783110804607

Ruch, W., and Carrell, A. (1998). Trait cheerfulness and the sense of humor. *Pers. Individ. Dif.* 24, 551–558. doi: 10.1016/S0191-8869(97)00221-3

Ruch, W., and Hehl, F.-H. (2007). "A two-mode model of humor appreciation: its relation to aesthetic appreciation and simplicity-complexity of personality," in *The Sense of Humor*, ed. W. Ruch (New York, NY: Mouton de Gruyter), 109–142. doi: 10.1515/9783110804607-007

Ruch, W., and Hofmann, J. (2012). "A temperament approach to humor," in *Humor and Health Promotion*, ed. P. Gremigni (Hauppauge, NY: Nova Science Publishers), 79–113.

Ruch, W., and Hofmann, J. (2017). "Fostering humour," in *Positive Psychology Interventions in Practice*, ed. C. Proctor (New York, NY: Springer), 65–80. doi: 10.1007/978-3-319-51787-2_5

Ruch, W., Hofmann, J., and Platt, T. (2015). Individual differences in gelotophobia and responses to laughter-eliciting emotions. *Pers. Individ. Dif.* 72, 117–121. doi: 10.1016/j.paid.2014.08.034

Ruch, W., Köhler, G., and van Thriel, C. (1996). Assessing the "humorous temperament": construction of the facet and standard trait forms of the State-Trait-Cheerfulness-Inventory–STCI. *Humor Int. J. Humor Res.* 9, 303–339. doi: 10.1515/humr.1996.9.3-4.303

Ruch, W., Proyer, R. T., Esser, C., and Mitrache, O. (2011). "Cheerfulness and everyday humorous conduct," in *Yearbook of the 'George Baritiu' Institute of History in Cluj-Napoca: Series Humanistica*, Vol. IX, ed. Romanian Academy (Bucharest: Romanian Academy Publishing House), 67–87.

Sauter, D., McDonald, N. M., Gangi, D., and Messinger, D. S. (2013). "Nonverbal expressions of positive emotions," in *Handbook of Positive Emotions*, eds M. N. Tugade, M. N. Shiota, and L. D. Kirby (New York, NY: Guilford).

Schueller, S. M. (2012). Personality fit and positive interventions: extraverted and introverted individuals benefit from different happiness increasing strategies. *Psychology* 3, 1166–1173. doi: 10.4236/psych.2012.312A172

Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T., and Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evol. Hum. Behav.* 31, 87–94. doi: 10.1016/j.evolhumbehav.2009.09.006

Senf, K., and Liau, A. K. (2013). The effects of positive interventions on happiness and depressive symptoms, with an examination of personality as a moderator. *J. Happiness Stud.* 14, 591–612. doi: 10.1007/s10902-012-9344-4

Vagnoli, L., Caprilli, S., and Messeri, A. (2010). Parental presence, clowns or sedative premedication to treat preoperative anxiety in children: what could be the most promising option? *Pediatr. Anesth.* 20, 937–943. doi: 10.1111/j.1460-9592.2010.03403.x

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Auerbach. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership