

NOVEL APPROACHES FOR STUDYING CREATIVITY IN CREATIVE COGNITION, ARTISTIC PERFORMANCE AND ARTISTIC PRODUCTION

EDITED BY: Philip Fine, Amory H. Danek, Kathryn Friedlander, Ian Hocking
and William Forde Thompson
PUBLISHED IN: Frontiers in Psychology





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-217-6

DOI 10.3389/978-2-88963-217-6

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

NOVEL APPROACHES FOR STUDYING CREATIVITY IN CREATIVE COGNITION, ARTISTIC PERFORMANCE AND ARTISTIC PRODUCTION

Topic Editors:

Philip Fine, University of Buckingham, United Kingdom

Amory H. Danek, University of Heidelberg, Germany

Kathryn Friedlander, University of Buckingham, United Kingdom

Ian Hocking, Canterbury Christ Church University, United Kingdom

William Forde Thompson, Macquarie University, Australia



Inner Melody series. Abstract design made of colorful human and musical shapes on the subject of spirituality of music and performing arts

By agsandrew

Royalty-free stock illustration ID: 225930193

This eBook presents the current state of the art in creativity research, by showcasing novel and/or interdisciplinary methodological approaches for studying creativity in creative cognition, artistic performance and artistic production. Its aims are both to enhance our understanding of these domains of creativity, and to foster new research ideas and collaborations through the use of these novel approaches.

There is a long history of research into creative cognition and creative performance, addressing questions of the creative process, individual differences in creative ability, what constitutes a creative product, and finally environmental influences on creativity. However, as creativity is such a broad and multifaceted area, research has tended to focus on discrete areas of study, with little opportunity for cross-fertilization. It is thus important to integrate research ideas and empirical

methods and findings across a variety of disciplines. One way to achieve this is to share methodological approaches for investigating creativity, in particular novel ones.

We see four ways in which novel approaches or methodologies have emerged: 1) through innovative uses of new technologies; 2) through investigating hitherto neglected domains of creativity; 3) by accessing specific creative populations; and 4) by combining existing approaches and methods within and across disciplines.

This eBook contains 27 articles exploring all four of these novel approaches, together with an editorial. Whereas the editorial is organised by the various methodological themes found in the articles, this eBook as a whole is organised according to the main domain of creativity, whether creative cognition or creative art and artistic performance.

We anticipate that the articles in this eBook will foster interdisciplinary cross-fertilization by sharing and promoting novel methodological approaches for studying all aspects of creativity.

Citation: Fine P. A, Danek A. H, Friedlander K. J, Hocking I., Thompson W. F, eds. (2020). *Novel Approaches for Studying Creativity in Creative Cognition, Artistic Performance and Artistic Production*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-217-6

Table of Contents

07 Editorial: Novel Approaches for Studying Creativity in Problem-Solving and Artistic Performance

Philip A. Fine, Amory H. Danek, Kathryn J. Friedlander, Ian Hocking and William Forde Thompson

SECTION 1

CREATIVE COGNITION

CHAPTER 1

INSIGHT

11 “The Penny Drops”: Investigating Insight Through the Medium of Cryptic Crosswords

Kathryn J. Friedlander and Philip A. Fine

33 Connect 4: A Novel Paradigm to Elicit Positive and Negative Insight and Search Problem Solving

Gillian Hill and Shelly M. Kemp

44 Feelings-of-Warmth Increase More Abruptly for Verbal Riddles Solved With in Contrast to Without Aha! Experience

Jasmin M. Kizilirmak, Violetta Serger, Judith Kehl, Michael Öllinger, Kristian Folta-Schoofs and Alan Richardson-Klavehn

55 How Working Memory Provides Representational Change During Insight Problem Solving

Sergei Korovkin, Ilya Vladimirov, Alexandra Chistopolskaya and Anna Savinova

71 Paving the Way to Eureka—Introducing “Dira” as an Experimental Paradigm to Observe the Process of Creative Problem Solving

Frank Loesche, Jeremy Goslin and Guido Bugmann

89 The Role of Motor Activity in Insight Problem Solving (the Case of the Nine-Dot Problem)

Vladimir Spiridonov, Nikita Loginov, Ivan Ivanchei and Andrei V. Kurgansky

106 Normative Data for 84 UK English Rebus Puzzles

Emma Threadgold, John E. Marsh and Linden J. Ball

CHAPTER 2

OTHER ASPECTS OF CREATIVE COGNITION: DIVERGENT THINKING AND PROBLEM-SOLVING

121 How Artificial Intelligence Can Help us Understand Human Creativity

Fernand Gobet and Giovanni Sala

127 Use or Consequences: Probing the Cognitive Difference Between Two Measures of Divergent Thinking

Richard W. Hass and Roger E. Beaty

142 Exploring the Creative Process: Integrating Psychometric and Eye-Tracking Approaches

Dorota M. Jankowska, Marta Czerwotka, Izabela Lebuda and Maciej Karwowski

- 154** *Creativity in the Here and Now: A Generic, Micro-Developmental Measure of Creativity*
Elisa Kupers, Marijn Van Dijk and Andreas Lehmann-Wermser
- 168** *The Effect of Problem Construction on Team Process and Creativity*
Roni Reiter-Palmon and Vignesh Murugavel
- 176** *A Growth-Curve Analysis of the Effects of Future-Thought Priming on Insight and Analytical Problem-Solving*
Monica Truelove-Hill, Brian A. Erickson, Julia Anderson, Mary Kossoyan and John Kounios
- 185** *What Enables Novel Thoughts? The Temporal Structure of Associations and its Relationship to Divergent Thinking*
Peng Wang, Maarten L. Wijnants and Simone M. Ritter
- 199** *Engaging in Creativity Broadens Attentional Scope*
Marta K. Wronska, Alina Kolańczyk and Bernard A. Nijstad

SECTION 2

CREATIVE ARTISTIC PERFORMANCE AND PRODUCTION

CHAPTER 3

CREATIVE ARTISTIC PERFORMANCE: MUSIC, DANCE AND POETRY

- 213** *Investigating the Role of the Primary Motor Cortex in Musical Creativity: A Transcranial Direct Current Stimulation Study*
Aydin Anic, Kirk N. Olsen and William Forde Thompson
- 223** *Collaborative Musical Creativity: How Ensembles Coordinate Spontaneity*
Laura Bishop
- 240** *Learning Choreography: An Investigation of Motor Imagery, Attentional Effort, and Expertise in Modern Dance*
Katy Carey, Aidan Moran and Brendan Rooney
- 251** *Expertise in Evaluating Choreographic Creativity: An Online Variation of the Consensual Assessment Technique*
Lucie Clements, Emma Redding, Naomi Lefebvre Sell and Jon May
- 259** *The Improvisational State of Mind: A Multidisciplinary Study of an Improvisatory Approach to Classical Music Repertoire Performance*
David Dolan, Henrik J. Jensen, Pedro A. M. Mediano, Miguel Molina-Solana, Hardik Rajpal, Fernando Rosas and John A. Sloboda
- 280** *Let's Read a Poem! What Type of Poetry Boosts Creativity?*
Małgorzata Osowiecka and Alina Kolańczyk

CHAPTER 4

CREATIVE ARTISTIC PRODUCTION: ART, DESIGN AND FASHION

- 292** *What are the Stages of the Creative Process? What Visual Art Students are Saying.*
Marion Botella, Franck Zenasni and Todd Lubart
- 305** *Conceptualising and Understanding Artistic Creativity in the Dementias: Interdisciplinary Approaches to Research and Practise*
Paul M. Camic, Sebastian J. Crutch, Charlie Murphy, Nicholas C. Firth, Emma Harding, Charles R. Harrison, Susannah Howard, Sarah Strohmaier, Janneke Van Leewen, Julian West, Gill Windle, Selina Wray and Hannah Zeilig on behalf of the Created Out of Mind Team

- 317** *Portrait of an Artist as Collaborator: An Interpretative Phenomenological Analysis of an Artist*
Ian Hocking
- 327** *Looking at the Process: Examining Creative and Artistic Thinking in Fashion Designers on a Reality Television Show*
Jillian Hogan, Kara Murdock, Morgan Hamill, Anastasia Lanzara and Ellen Winner
- 340** *A Decision Tree Based Methodology for Evaluating Creativity in Engineering Design*
Trina C. Kershaw, Sankha Bhowmick, Carolyn Conner Seepersad and Katja Hölttä-Otto
- 359** *Spontaneous Visual Imagery During Meditation for Creating Visual Art: An EEG and Brain Stimulation Case Study*
Caroline Di Bernardi Luft, Ioanna Zioga, Michael J. Banissy and Joydeep Bhattacharya



Editorial: Novel Approaches for Studying Creativity in Problem-Solving and Artistic Performance

Philip A. Fine^{1*}, Amory H. Danek², Kathryn J. Friedlander¹, Ian Hocking³ and William Forde Thompson⁴

¹ School of Psychology and Wellbeing, University of Buckingham, Buckingham, United Kingdom, ² Department of Psychology, Universität Heidelberg, Heidelberg, Germany, ³ School of Psychology, Politics and Sociology, Christchurch Canterbury University, Canterbury, United Kingdom, ⁴ Department of Psychology, Macquarie University, Sydney, NSW, Australia

Keywords: creativity, problem solving, artistic performance, methodology, novel approach

Editorial on the Research Topic

Novel Approaches for Studying Creativity in Problem-Solving and Artistic Performance

INTRODUCTION

Creativity can be observed across multiple domains of human behavior including problem solving, artistic and athletic engagement, scientific reasoning, decision making, business and marketing, leadership styles, and social interactions. It has a long history of research in many disciplines, and involves a variety of conceptual and methodological approaches. However, given its multi-faceted character, and the multidisciplinary (though not necessarily interdisciplinary) nature of creativity research, it is perhaps unsurprising that such research has tended to examine discrete areas of study, thereby adopting a focused approach that lacks opportunity for cross-fertilization. It is therefore important to encourage interdisciplinary discourse and novel methodological approaches to investigating all aspects of creativity. This can best be achieved by sharing and integrating research ideas, methods, and findings across multiple domains and disciplines, including but not restricted to psychology, neuroscience, philosophy, linguistics, medicine, education, and performance science.

The aim of this Research Topic is to showcase recent creativity research involving new methodological approaches across a range of creativity domains and academic disciplines. Broadly speaking, we see three ways by which such novel methodological approaches can develop. Firstly, adopting technologies such as brain stimulation and EEG allow researchers to investigate creativity in new ways, and new digital research platforms allow researchers to more easily access domain-specific online populations. Secondly, traditional methodologies, already shown to be effective in one field of creativity research, can be employed to investigate hitherto neglected creativity domains. Thirdly, taking advantage of the interdisciplinary nature of creativity research, we can interrogate one domain of creative performance using research perspectives from another, such as viewing medicine as a performance science akin to music (Kneebone, 2016) or investigating insight moments with magic tricks (Danek et al., 2014). This novel juxtaposition of methods from multiple domains and disciplines allows new research questions to be addressed. These three ways of developing novel methodological approaches thus involve: the development of novel methods; the novel application of tried-and-tested methods; and the novel combination of previously separate methodologies.

OPEN ACCESS

Edited and reviewed by:

Aaron Williamon,
Royal College of Music,
United Kingdom

*Correspondence:

Philip A. Fine
philip.fine@buckingham.ac.uk

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 01 August 2019

Accepted: 23 August 2019

Published: 18 September 2019

Citation:

Fine PA, Danek AH, Friedlander KJ,
Hocking I and Thompson WF (2019)
Editorial: Novel Approaches for
Studying Creativity in Problem-Solving
and Artistic Performance.
Front. Psychol. 10:2059.
doi: 10.3389/fpsyg.2019.02059

The Research Topic contains 27 articles (20 Original Research articles, one Case Report, one Review, and five methodological or theoretical contributions). Twelve address questions of creative cognition, covering insight, divergent thinking, and problem solving. Eleven articles investigate creative arts and artistic performance, with a further four addressing other aspects of creativity. Given the focus of the Research Topic, we have decided to address the articles in terms of their methodological approaches, rather than the type of creativity under investigation. Indeed, we hope to encourage the development and ultimately the wider application of those methodological approaches described herein to any aspect or domain of creativity.

TRACKING THE PROCESS: PHYSIOLOGICAL APPROACHES

In line with the increasing pace of technological advancement, several articles utilize physiological techniques to measure and manipulate the creative process, including the electroencephalogram (EEG), and transcranial current stimulation, both direct (tDCS) and alternating (tACS). Dolan et al. employ EEG in both music performers and selected audience members during prepared and improvised renditions of the same piece of classical music, demonstrating what they call an “improvisatory state of mind.” Truelove-Hill et al. measure resting-state EEG in their investigation of the effects of near-future and far-future priming on insight and analytical problem-solving. Di Bernardi Luft et al. use both EEG and tACS in their case study of a professional visual artist with exceptionally vivid spontaneous visual imagery during meditation sessions. They demonstrate increased occipital gamma oscillations during visual imagery, and an effect of alpha tACS on the contents of the artist’s images. In another study of musical creativity, Anic et al. investigate the effects of both excitatory and inhibitory tDCS over the left hemisphere primary motor cortex (M1) of pianists who were improvising with their right hands: improvisations under excitatory tDCS were rated as significantly more creative, demonstrating the role of M1 in musical creativity.

Various other articles employ process-tracing methods to probe the creative process. Carey et al. investigate dance in a novel way, using pupillometry (a metric of mental effort) to demonstrate greater pupil dilation in novice, rather than intermediate, dancers as they performed or imagined dance movements. Jankowska et al. use both eye-tracking and think-aloud (verbal protocol) analyses whilst adults completed a creative drawing task, demonstrating methodological synergy between both types of process-tracing and various psychometric measures of drawing creativity. Spiridonov et al., Loesche et al., and Dolan et al. all track physical movement during various creative acts. Spiridonov et al. examine the classic 9-dot problem by tracking the position and movement of the solver’s index finger on a tablet, and demonstrate specific patterns of motor behavior characterizing the differences between unsuccessful and successful solvers. Similarly, Loesche et al. investigate the chronology of insight moments in a novel insight eliciting task,

“Dira,” by tracking the position of the mouse cursor, allowing them to better pinpoint the moment when solutions emerge. Finally, Dolan et al. investigate musical creativity in ensemble playing in various ways, including continuous 3D tracking of the musicians’ movement. This enables them to explore movement pattern differences between improvised and prepared renditions, as well as demonstrate, for instance, that the flutist and pianist correlated their fast movements significantly more in an improvised rendition than a classically prepared one.

THE TIME-COURSE OF CREATIVITY

One common theme, found in 10 articles, is the study of temporal or chronometric aspects of the creative and associated processes. Three articles involving process-tracing, focusing particularly on moment-to-moment aspects of the creative process, have already been mentioned (Loesche et al., Spiridonov et al., and Dolan et al.). Hass and Beatty directly compare performance on the Alternative Uses Task (AUT) and Consequences Task, showing that both approximate well to an exponential cumulative response time model; they also provide an explanation for why later responses are generally rated as more creative than earlier ones, known as the serial order effect. Kizilirmak et al. measure feelings of warmth (FoW) ratings for Compound Remote Associate Tasks as a function of task difficulty, whether it was successfully solved, and whether the solution (if it occurred) was an example of insight; they demonstrate that FoW ratings increase more abruptly for trials solved with compared to without an insight experience. Kupers et al. measure moment-to-moment ratings of novelty and appropriateness in their study of children’s creativity using a novel coding framework. Botella et al. explore the stages of the creative artistic process, which they propose differs from both the creative process and the artistic process, by interviewing visual graphic arts students, integrating their findings into Creative process Report Diaries.

Rather than focusing on the creative process itself, three articles measure the time-course of associated processes. Wang et al. explore the temporal structure of semantic associations in an association chain task and its relationship to divergent thinking. Korovkin et al. use a dual-task procedure to track the temporal dynamics of working memory involvement throughout both insight and non-insight problem-solving experiences. Truelove-Hill et al. investigate the effects of a priming procedure on creative problem-solving by asking problem-solvers to think about the near vs. distant future in order to differentially impact their cognitive style, in accordance with construal level theory. They then apply growth-curve analysis in a novel way to uncover the time-course of these transient priming effects.

PROMOTING AND MEASURING CREATIVITY: PSYCHOMETRIC APPROACHES

Several articles describe novel approaches to promote, track or measure creativity. Three articles propose novel methods for inducing insight. Friedlander and Fine posit a new protocol

for eliciting insight moments, that of cryptic crossword solving, drawing parallels between certain cryptic clue mechanisms and problem types already found in the insight literature, such as rebus puzzles, remote associate problems, anagrams, and jokes. Such an approach could be instrumental in exploring individual differences in insight ability, and identifying insight experts. In order to investigate multiple instances of both positive (Aha!) and negative (Uh-oh!) insight experiences, Hill and Kemp use the well-known adversarial game of Connect 4, asking participants to label each move as insight or search (either positive or negative) and collecting concomitant phenomenological ratings. Loesche et al. have developed a new game, “Dira,” based on the existing game “Dixit,” in which participants must find a connection between a short sentence and one of six visual images. However, only the image (or text) over which the mouse is hovering is clearly visible: this allows real-time process-tracing via mouse movements, and provides information about relevant metacognitive and behavioral mechanisms, such as the intensity of the insight moment.

Other cognitive methods applied to creativity research in the current articles include: the use of verbal protocol analysis to probe metacognitive and self-regulation mechanisms together with eye-movement measures during a creative drawing task (Jankovska et al.); the measurement of feelings of warmth during insight and non-insight puzzle solving (Kizilirmak et al.); and the application of the classic dual-task paradigm to investigate the effect of working memory load on solving insight and non-insight problems (Korovkin et al.). Camic et al. also describe the potential utility, for those with dementia, of Visual Thinking Strategies (VTS), an arts-based facilitated learning methodology involving moderated group discussions, permitting individuals to create meaning through viewing visual art.

Two articles probe novel and interesting causal relationships between creativity and other cognitive activities or processes. Having a broad attentional scope has previously been shown to enhance creativity, but Wronska et al. demonstrate the reverse relationship, that divergent thinking can broaden visual attention on a subsequent visual scanning task and enhance peripheral target recognition. Osowiecka and Kolanczyk show that silently reading poetry can both increase and decrease divergent thinking performance, depending on the type of poetic metaphors, the poetic narration style, and individual differences in long-term exposure to poetry.

Several articles explore novel psychometric methods for measuring and otherwise quantifying aspects of creativity. Threadgold et al. present a newly validated normative pool of 84 rebus puzzles freely available for future use in problem-solving and insight studies. Kupers et al. propose a micro-level domain-general systematic coding framework for measuring novelty and appropriateness of creative products on a continual basis. Kershaw et al. apply a novel originality scoring method, the Decision Tree for Originality Assessment in Design (DTOAD), to creative ideation within engineering design. Clements et al. adapt Amabile’s Consensual Assessment Technique (CAT; Amabile, 1982; Cseh and Jeffries, 2019) for online use so as to have a broader reach, by which they investigate the effects of varying levels of dance expertise and experience on ratings

of choreographic creativity. Loesche et al.’s exploration of the chronometry of insight moments and Threadgold et al.’s construction of a normative database of rebus puzzles both treat the strength of the Eureka experience as a continuum rather than a dichotomous all-or-none phenomenon, which has generally been a more common approach; similarly, some articles, including Hill and Kemp, and Loesche et al., consider phenomenological correlates of the insight moment as continua.

TECHNOLOGICAL AND METHODOLOGICAL ADVANCES

In addition to the studies using tDCS, tACS, and EEG already mentioned, two articles in particular employ methods novel to creativity research to increase the reach of their studies. For their direct comparison of the AUT and the Consequences Task, Hass and Beatty’s participants were recruited from Amazon Mechanical Turk (MTurk) using psiTurk, an open-access web-app which interfaces with MTurk, allowing online experimental control and response collection. In their study of choreographic creativity, Clements et al. use an online version of the CAT together with a snowball sampling technique in which participants could rate as few or as many as they wished out of 23 randomly ordered short videos: this yielded 2153 individual ratings from 850 raters.

Camic et al. advocate the use of wearable technology for measuring psychophysiological changes on a continuous basis during creative behaviors, particularly where it is important that such data collection is unobtrusive, for instance in persons with dementia. Wearable technology such as wristbands can record 3D position using accelerometers, as well as physiological indices of arousal and stress including heart rate, heart rate variability, skin conductance, and skin temperature. Finally, in their Perspective article, Gobet and Sala advocate the use of methods in Artificial Intelligence (AI), which they argue are less susceptible to mental set issues, in both the design of new experiments and the generation of new theory in relation to the study of creativity.

INVESTIGATING CREATIVE PEOPLE AND POPULATIONS

Several articles focus more on the creative person, by studying either specific (and sometimes less-studied) populations, or interpersonal aspects of teamwork, ensemble, and co-creativity. Hogan et al. investigate budding fashion designers on a reality television programme in which they are tasked with designing garments. The authors analyze the designers’ thinking dispositions using qualitative analysis of the programme transcripts in terms of the 8 Studio Habits of Mind. In a multi-institutional wide-ranging Conceptual Analysis article, Camic et al. explore how we can conceptualize and understand artistic creativity in the dementias, a population easily and undeservedly overlooked in creativity studies. An interesting aspect of the article is their discussion of co-creativity,

which focuses on shared processes. Hocking, too, addresses co-creativity, in his dyadic case study of the subjective experience of a professional artist as seen through the eyes of a psychological researcher and thus artistic collaborator, using Interpretative Phenomenological Analysis (IPA). Another case study of an artist (Di Bernardi Luft et al.) employs neuroimaging to investigate spontaneous vivid visual imagery, central to this artist's creativity. Though still focusing on the creative process, Kupers et al. present two case studies specifically investigating children's creativity, exemplified by two empirical examples, a music composition task and the solving of a physics problem: their coding framework will no doubt also be applicable to adults (and to other domains of creativity).

Other articles addressed questions of interpersonal interaction with reference to teamwork and ensemble. Reiter-Palmon and Murugavel demonstrate the utility of problem construction in teams by studying the social and cognitive processes involved. Both Bishop and Dolan et al. investigate aspects of ensemble playing and collaborative processes in music performance. Bishop reviews recent literature on collaborative musical creativity, in terms of how ensembles achieve creative spontaneity, through the lenses of embodied music cognition, emergence, and group flow. Dolan et al. explore synchrony of movement in ensemble music performers as a function of the level of improvisation.

MULTIDISCIPLINARY, INTERDISCIPLINARY, AND BLENDED METHODOLOGICAL APPROACHES

As noted in the introduction to this editorial, one of the main drivers of this Research Topic is that of fostering interdisciplinary cross-fertilization. Two articles explicitly use such a multidisciplinary approach. Wang et al. combine approaches from computational linguistics, complex systems, and creativity research in their investigation of the relationship between semantic association and divergent thinking tasks. Camic et al.'s article about artistic creativity in the dementias is the culmination of a 2-year interdisciplinary study

involving research psychologists and neurologists, artists, and media professionals.

Certain articles, although focusing more on a single discipline (often psychology), use a blended approach of multiple methods, some comparing different methodologies directly, such as Hass and Beatty's comparison of the AUT and the Consequences Task. Dolan et al., in their study of an improvisatory approach to performing classical music, measure various performance-related parameters, post-performance ratings from both performers and audience members, EEG signals again from both performers and selected audience, and 3D motion tracking of the performers' movements. This broad range of measures enables them to demonstrate convergent evidence for differences between improvised and prepared musical performances. Jankowska et al. integrate psychometric, eye-tracking, and verbal protocol analysis in their study of creative drawing. Finally, Carey et al. combine measures of motor imagery, dance performance, and pupillometry to investigate dancers' learning of dance moves.

THE FUTURE OF CREATIVITY RESEARCH

Given the breadth of creativity research, investigating as it does at least the creator, the creative process, the creative product, and environmental influences on creativity (Rhodes, 1961; Abdulla and Cramond, 2017), it is important to integrate research ideas, methods, and findings across diverse disciplines. The 27 articles in this Research Topic present a broad picture of contemporary creativity research across multiple disciplines and domains. Separately and together they present a range of novel approaches for studying all aspects of creativity which we hope will encourage further interdisciplinary cross-fertilization. Creativity research is clearly thriving, and through the methodological creativity of developing innovative research methods and approaches, we are in a strong position to advance our understanding of creativity in all its forms.

AUTHOR CONTRIBUTIONS

PF wrote the first draft of this editorial, and all authors equally contributed to the revisions.

REFERENCES

- Abdulla, A. M., and Cramond, B. (2017). After six decades of systematic study of creativity: what do teachers need to know about what it is and how it is measured? *Roeper. Rev.* 39, 9–23. doi: 10.1080/02783193.2016.1247398
- Amabile, T. M. (1982). Social psychology of creativity: a consensual assessment technique. *J. Personal. Soc. Psychol.* 43, 997–1013. doi: 10.1037/0022-3514.43.5.997
- Cseh, G. M., and Jeffries, K. K. (2019). A scattered CAT: a critical evaluation of the consensual assessment technique for creativity research. *Psychol. Aesthet. Creat. Arts* 13, 159–166. doi: 10.1037/aca0000220
- Danek, A. H., Fraps, T., Von Mueller, A., Grothe, B., and Öllinger, M. (2014). Working wonders? Investigating insight with magic tricks. *Cognition* 130, 174–185. doi: 10.1016/j.cognition.2013.11.003
- Kneebone, R. L. (2016). Performing surgery: commonalities with performers outside medicine. *Front. Psychol.* 7:1233. doi: 10.3389/fpsyg.2016.01233
- Rhodes, M. (1961). An analysis of creativity. *Phi Delta Kappan* 42, 305–310.

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 Fine, Danek, Friedlander, Hocking and Thompson. 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.



“The Penny Drops”: Investigating Insight Through the Medium of Cryptic Crosswords

Kathryn J. Friedlander* and Philip A. Fine

Department of Psychology, University of Buckingham, Buckingham, United Kingdom

OPEN ACCESS

Edited by:

George Kachergis,
Radboud University Nijmegen,
Netherlands

Reviewed by:

John Kounios,
Drexel University, United States
Carola Salvi,
Northwestern University, United States

*Correspondence:

Kathryn J. Friedlander
kathryn.friedlander@buckingham.ac.uk

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 31 October 2017

Accepted: 17 May 2018

Published: 03 July 2018

Citation:

Friedlander KJ and Fine PA (2018)
“The Penny Drops”: Investigating
Insight Through the Medium of Cryptic
Crosswords. *Front. Psychol.* 9:904.
doi: 10.3389/fpsyg.2018.00904

A new protocol for eliciting insight (“Aha!”/Eureka) moments is proposed, involving the solving of British-style cryptic crosswords. The mechanics of cryptic crossword clues are briefly explained, and the process is set into the insight literature, with parallels being drawn between several different types of cryptic crossword clues and other insight-triggering problems such as magic, jokes, anagrams, rebus, and remote association puzzles (RAT), as well as “classic” thematic or spatial challenges. We have evidence from a previous survey of cryptic crossword solvers that the “Aha!” moment is the most important driver of continued participation in this hobby, suggesting that the positive emotional “payback” has an energizing effect on a participant’s motivation to continue solving. Given the success with which a good quality cryptic crossword elicits “Aha!” moments, cryptics should prove highly valuable in exploring insight under lab conditions. We argue that the crossword paradigm overcomes many of the issues which beset other insight problems: for example, solution rates of cryptic crossword clues are high; new material can easily be commissioned, leading to a limitless pool of test items; and each puzzle contains clues resembling a wide variety of insight problem types, permitting a comparison of heterogeneous solving mechanisms within the same medium. Uniquely among insight problems, considerations of expertise also come into play, allowing us to explore how crossword solving experts handle the deliberate misdirection of the cryptic clue more effectively than non-expert, but equally experienced, peers. Many have debated whether there is such a thing as an “insight problem” *per se*: typically, problems can be solved with or without insight, depending on the context. We argue that the same is true for cryptic crosswords, and that the key to the successful triggering of insight may lie in both the difficulty of the challenge and the degree to which misdirection has been used. Future research is outlined which explores the specific mechanisms of clue difficulty. This opens the way to an exploration of potential links between solving constraints and the experiencing of the “Aha!” moment, which may shed light on the cognitive processes involved in insight solution.

Keywords: cryptic crossword expertise, Aha! insight problem-solving, representational change, chunk decomposition, opportunistic assimilation, rebus and remote association puzzles, jokes, anagrams

INTRODUCTION: INSIGHT AND “INSIGHT PROBLEMS”

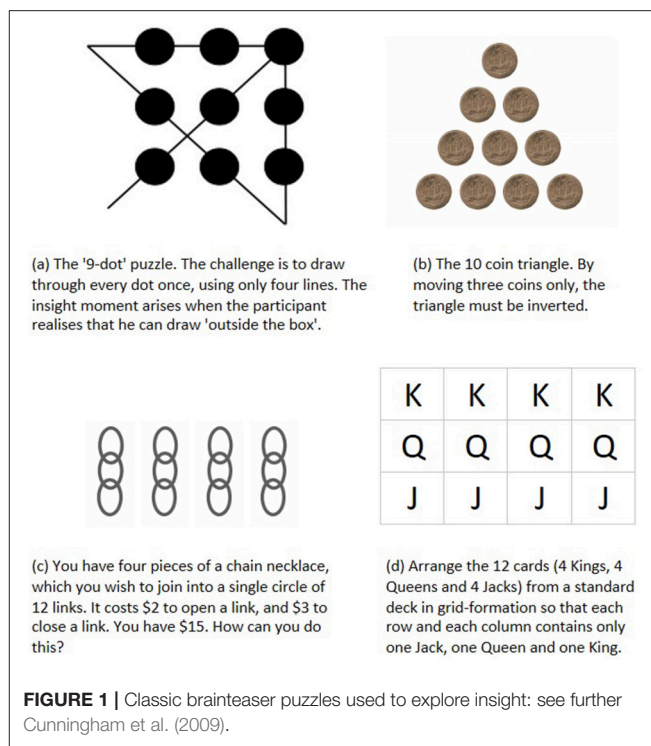
The feeling of insight—a sudden, euphoric “cognitive snap” (Weisberg, 2015) signaling a breakthrough in the solution of a problem—is well-known to most of us. In terms of its phenomenological experience, four key elements of the insight, or “Aha!” moment have been identified: first, the suddenness and unexpectedness of the resolution, which arrives unheralded by conscious awareness of the solution path or “feelings of warmth” at the approaching dénouement; secondly that—however difficult it had proved before (perhaps involving a state of impasse)—the problem can be rapidly processed once the solution has been identified; thirdly that there is a strong, typically positive, emotional response at the point of resolution; and finally that the solver is fully convinced that the correct solution has been identified (Topolinski and Reber, 2010a; see also Metcalfe, 1986; Davidson, 1995; Gick and Lockhart, 1995; Danek et al., 2014a,b; Kounios and Beeman, 2014; Shen et al., 2015; on negative insight (“Uh-oh”) see also Hill and Kemp, 2016). The phenomenological experience of the “Aha!” moment is thus complex, with at least four contributory components: suddenness, surprise, happiness and certainty (Gick and Lockhart, 1995; Danek et al., 2014a, 2016).

One of the key problems in studying insight is the unpredictability of this moment in everyday life. Although “everyday insight moments” can be experienced (such as the sudden realization of where a bunch of keys has been left), the sudden and fleeting nature of this moment has led most studies to attempt to elicit responses artificially under laboratory conditions, using a bank of so-called “insight problems” intended to trigger the identical phenomenological response (Hill and Kemp, 2016). Nonetheless, even this approach is not without issues, primarily centered upon the difficulty of finding an effective, convenient, and reliable insight-triggering task for the participant to solve.

Current Obstacles in Exploring Insight in the Laboratory

Lab studies of insight in problem solving have met with a number of obstacles, which have been well rehearsed in the literature. These include the historic paucity of standardized problem material (MacGregor and Cunningham, 2008; Batchelder and Alexander, 2012; Danek et al., 2014b); the difficulty and complexity of the tasks, leading to low solution rates and low numbers of problem trials within the practical limitations of investigative time-frames (Bowden and Jung-Beeman, 2003b; MacGregor and Cunningham, 2008; Batchelder and Alexander, 2012; Danek et al., 2016); and the memory advantage obtained for solutions arrived at by insight (Dominowski and Buyer, 2000; Danek et al., 2013) which rules out test-retest options (MacGregor and Cunningham, 2008).

This last issue poses a particular problem for controlled, lab-based research, given that the solutions to so many of the classic riddle-style “insight problems” (e.g., the 9-dot problem,



the reversed triangle of coins, the broken necklace challenge—Cunningham et al., 2009—see Figure 1) are now freely available on-line and in puzzle collections; this commonly leads to the need to discard trials due to familiarity with the puzzles (Öllinger et al., 2014; see also Danek et al., 2016).

Following attempts to increase the pool of test material in recent years, larger collections of calibrated problems do now exist (Chu and MacGregor, 2011); these have moved away from the classic “riddle-style” puzzles (Webb et al., 2016) and might include matchstick arithmetic problems (Knoblich et al., 1999), compound remote association problems (“CRA”—a variation of “Remote Association Test” (RAT) problems—Bowden and Jung-Beeman, 2003b), the “Car Park Game” (Jones, 2003), rebus puzzles (MacGregor and Cunningham, 2008), Bongard problems and “tricky series completion” problems (Batchelder and Alexander, 2012). Recently, magic tricks have been added to the list of available paradigms (Danek et al., 2014b).

When is Insight “Insight”?

The use of a canonical set of “insight problems” to explore “Aha!” moments in the laboratory has led to a long-standing debate concerning the underlying cognitive mechanisms involved in their solution: specifically, whether an “Aha!” feeling is the result of “special” thought processes, or is merely an epiphenomenon arising from cognitive processes which are “business as usual” (for a review of this debate see Davidson, 1995; Bowden et al., 2005; Ohlsson, 2011; Gilhooly et al., 2015; Weisberg, 2015). One confounding issue which has hampered investigation of this question is the common assumption in many historical studies that “insight problems” are, *per se*, always solved with insight

by every successful solver; in other words, that triggering insight is an inherent and objective property of the “insight problem” which unfailingly comes into play (Bowden and Jung-Beeman, 2007; Ohlsson, 2011; Öllinger et al., 2014). Crucially, as a result of this *a priori* assumption, no check was typically made as to whether the “Aha!” moment had actually been experienced in these trials, leading to a highly problematic circularity: “Insight problems are problems that require insight, and insight occurs when insight problems are solved” (Öllinger and Knoblich, 2009, p. 277; see also Danek et al., 2016; Webb et al., 2016). An early attempt (Weisberg, 2015; see Ash et al., 2009) to circumvent this problem by categorizing “insight problems” into “pure” problems (those that could only be solved with insight), “hybrid” problems (those that could be solved through insight and other methods) and “non-insight” problems (those which are always resolved through an analytical approach) nonetheless still requires that a subset of problems exists which infallibly trigger insight.

A critical flaw in this approach is that it overlooks the interactive nature of problem solving: successful solving arises from the interplay of problem and person, with each individual bringing a unique blend of knowledge, experience and cognitive approaches to bear upon it (Ash et al., 2009; Ohlsson, 2011). It is therefore entirely possible for a so-called “insight puzzle” to be solved through controlled, deliberate, systematic and evaluative means by some solvers—analytic “Type 2” thinking according to dual process theory (Evans and Stanovich, 2013; Sowden et al., 2015; Weisberg, 2015)—which is not thought to give rise to a characteristically strong emotional response, other than satisfaction at the job completed (Kounios and Beeman, 2014).

Others, however, may solve the same puzzle with a flash of inspiration that they could not predict, through processes operating below the threshold of their awareness, and will experience the impact of the “Aha!” moment. Much will depend on what each solver brings to the solving process: “each problem can be solved without insight if the initial problem representation is adequate and the appropriate heuristics are available” (Öllinger et al., 2014, p. 267), and this will vary from solver to solver according to their skill-set and experience. The presence or absence of insight thus resides in the solver’s approach to solving the puzzle, not simply in the problem itself (Bowden and Jung-Beeman, 2007; Cunningham et al., 2009; Webb et al., 2016), and the categorization of “insight problem” stimuli as “pure” or “hybrid”, or “insight/non-insight” on the grounds of a hypothetical cognitive task analysis appears to be fundamentally flawed (Ash et al., 2009; Webb et al., 2016).

The purpose of insight research should not therefore be to develop a single theory which accounts for all solutions to “insight problems” arrived at by any manner under experimental conditions (Ohlsson, 2011), but to isolate those solutions which have evoked the phenomenological events specifically characteristic of an “Aha!” event, and to use these to explore the cognitive mechanisms underlying this experience (Webb et al., 2016). More contemporary studies have typically achieved this by collecting subjective feedback from trial participants as to whether they have actually experienced an “Aha!” moment at the point of solution (Bowden and Jung-Beeman, 2007; Kounios et al., 2008; Cranford and Moss, 2011; Jarosz et al., 2012; Danek

et al., 2014b; Salvi et al., 2016b; Webb et al., 2016). This technique has been validated by a number of neuroimaging studies, which have empirically demonstrated meaningful differences between problems identified by participants as being solved with insight, or in a step-wise fashion (Zhao et al., 2013; Kounios and Beeman, 2014).

Representational Change Theory

Notwithstanding this, it would be unhelpful to reject the term “insight problem” altogether, given that it is clear that some cognitive puzzles are more likely to trigger insight moments than others (Danek et al., 2014a), and indeed “insight problems” may operate along a continuum of efficacy (Webb et al., 2016). In particular, Representational Change Theory (“RCT”—Ohlsson et al., 1992; Knoblich et al., 1999; Ohlsson, 2011; Öllinger et al., 2014) suggests that especially effective insight-triggering puzzles use the solver’s prior knowledge and expectations to deliberately induce a false conceptualization of the problem (Ovington et al., 2016), leading to self-imposed constraints which impede a solution. This can result in a feeling of “impasse”: the situation where the solver feels that they have explored all possible approaches to resolving the problem, and is now at a loss as to what to try next (Knoblich et al., 2001).

The moment of insight is argued to be the point at which the hindering constraint is suddenly removed, leading to a relaxation of the impasse and the rapid redefining of the problem space, followed by a swift solution. The initially incorrect reading of the problem—termed mental set by the Gestalt school (Wiley, 1998; Öllinger et al., 2008)—is argued to arise unavoidably and unconsciously from implicit assumptions or well-practiced procedures which are activated highly automatically (Ohlsson et al., 1992; Knoblich et al., 1999; DeYoung et al., 2008; Öllinger et al., 2008; Danek et al., 2014b; Patrick et al., 2015), making the less obvious, but correct, interpretation of the problem very unlikely to come to mind. It is the dropping of the incorrect assumptions, and disengagement from the outdated hypothesis, which is argued to allow progress to be made.

Heterogeneous Nature of Insight Puzzles and Their Mechanisms

It is thus widely acknowledged that “insight problem” solving involves some form of reconstructive change of the initial representation of the problem (Chronicle et al., 2004; Cunningham et al., 2009; Danek et al., 2014a); however, the precise mechanisms to achieve this reconstruction—and whether they are in any way “special”—remain unclear.

A number of theoretical models to explain this restructuring in classic insight puzzles, such as the 9-dot or the 8-coin puzzles, have been put forward: for example “elaboration, re-encoding or constraint relaxation” (Ohlsson et al., 1992); “opportunistic assimilation” (Seifert et al., 1995); “constraint relaxation and chunk decomposition” (Knoblich et al., 1999); “solution-recoding” (Chronicle et al., 2004); see further the reviews by Ash et al. (2009) and Batchelder and Alexander (2012). Nonetheless, since the formulation of these theories, a wider range of insight-triggering paradigms has been developed which on at least superficial grounds differ greatly in their appearance

and the demands they make upon the solver (Bowden et al., 2005). It is therefore at least possible that the cognitive processes leading up to the moment of restructuring differ according to the specific puzzle parameters at play (Bowden and Jung-Beeman, 2007), making a single-process theory of restructuring difficult (Cunningham et al., 2009).

In a study comparing the relationships among a small range of diverse insight puzzles (classic “spatial” puzzles, RAT puzzles and rebus problems), Cunningham and colleagues identified the following characteristics of restructuring which they believed were displayed, to a greater or lesser extent, by each of their puzzle formats of interest (Cunningham et al., 2009). As predicted by RCT, some puzzles involved the need to overcome misdirection or the relaxation of automatically elicited constraints concerning the existing components of the puzzle or its spatial layout (Cunningham et al., 2009). However, in others, the primary difficulty appeared to lie in identifying what the eventual solution would look like, perhaps requiring the assimilation of extra incidental information, a sudden “figure-ground” reversal of perspective, or additional steps in order to hit upon the solution (Cunningham et al., 2009).

One methodological issue thus lies in how “well-defined” a problem type is (DeYoung et al., 2008; see also Simon, 1973; Davidson, 2003; Pretz et al., 2003; Hélie and Sun, 2010; Danek et al., 2016; Ovington et al., 2016; Webb et al., 2016). An ill-defined problem has no clear representation of the problem space in terms of key features such as the initial conceptualization of the challenge, the final goal state, and the mechanizable steps which need to be taken to achieve this goal. By contrast, “well-defined” problems may be tackled by controlled and systematic paradigmatic processes leading to steady progress toward a known target state (Smith, 2003; DeYoung et al., 2008), and better defined problems of this kind therefore lead less often to solution through insight (Webb et al., 2016).

Despite early attempts to categorize insight puzzles (e.g., as pure/hybrid) according to solving process (Ohlsson et al., 1992; Weisberg, 1995; Ansburg and Dominowski, 2000), the heterogeneous nature of the various problem collections therefore makes equivalence studies difficult (Weisberg, 1995; Cunningham et al., 2009), and this limits our understanding of the core components of problem solving with insight (Bowden and Jung-Beeman, 2003b; MacGregor and Cunningham, 2008). Attempts to find one single explanation of the cognitive processes leading to insight solution by pitting alternative theories against each other on a single puzzle type (e.g., Jones, 2003) may on this account be doomed: it is entirely possible that insight could arise from different interacting sets of preceding processes depending upon the context and the challenge inherent in the problem and that these processes may only imperfectly map onto these traditional problem type categories (Bowden and Jung-Beeman, 2007; Shen et al., 2016). A theoretical or computational model of “insight problem” solving which satisfactorily explains all facets and styles of insight challenge is therefore proving elusive (Ash et al., 2009; Batchelder and Alexander, 2012).

Rapid Solving and Incubated Problems

Equally vexed is the question of whether a period of impasse is always involved in insight problem-solving (as argued e.g., by Ohlsson et al., 1992), with some studies reporting that—even within puzzle type—solvers did not uniformly experience a period of impasse (Ash et al., 2012; Cranford and Moss, 2012; Danek et al., 2014a).

Indeed, studies have suggested that solvers can experience an instantaneous “Aha!” moment within seconds of the presentation of the puzzle. In a study of anagram solving, Novick and Sherman noted that “pop-out” solutions tended to be the first solution offered and to occur within 2 s of the presentation of the letters (Novick and Sherman, 2003). In trials of highly skilled anagram solvers, 47% of the solutions were reported to be immediate “pop-out” solutions, where the solver agreed that, “The solution came to mind suddenly, seemingly out of nowhere. I have no awareness of having done anything to try to get the answer.” By contrast 27% of solutions occurred with insight after a period of trying fruitless combinations; and 26% were generated incrementally by the recursive testing of morphemically probable combinations (non-insight search solutions).

Similarly, a study of RAT problems (Cranford and Moss, 2012), found that 171 out of 218 solutions arrived at with self-reported insight, under think-aloud conditions, were solved almost immediately, in a mean time of 7.1 s. These were categorized as “Immediate Insight” (II) moments; however, the authors also raised the possibility that the solution might simply have occurred so fast that it appeared sudden and surprising, without evoking the full phenomenological experience (Cranford and Moss, 2012; see also Topolinski and Reber, 2010b). Indeed, an fMRI study comparing II with Delayed Insight (DI) RAT solutions showed large differences in activation patterns for the two types of insight, suggesting that they may represent distinct solution processes (Cranford and Moss, 2011). For this reason, some later studies have excluded II solutions from their discussion, on the grounds that they may not reflect the full “Aha!” experience (e.g., Salvi et al., 2016a).

Conversely, the benefits of a period of incubation (non-conscious solving activity, or a period of respite away from the problem) in resolving problems which have reached impasse have been well-documented (see the meta-analytic review by Sio and Ormerod, 2009; also Ohlsson, 2011; Baird et al., 2012; Sio and Ormerod, 2015; Gilhooly, 2016), although the mechanisms which account for the facilitation of the solution (e.g., “unconscious work,” “intermittent work,” “beneficial forgetting”—Gilhooly, 2016) are as yet unclear. Incubation is clearly not always involved in insight problem resolution—though it was present as the second of Wallas’ (1926) four stages of insight problem-solving (Sio and Ormerod, 2009)—and is rather seen as an ancillary feature, to be utilized where necessary (Gilhooly, 2016). Engaging in a diversionary activity with a low cognitive load appears to be most helpful (Sio and Ormerod, 2009), and many people report that the problem solution occurs to them when engaged in everyday activities such as walking, driving, or showering (Hill and Kemp, 2016; Ovington et al., 2018); a substantial number also report facilitation overnight, during their dreams or immediately upon waking (Ovington et al., 2018).

BOX 1 | Illustration of cryptic clue mechanisms: misleading surface readings.**Clue 1(a) Active women iron some skirts and shirts (9)**—(Schulman, 1996, p. 309)

The definition is “Active women” = an obliquely phrased straight definition for FEMINISTS

The wordplay comprises: FE (iron, chemical symbol) + MINIS (plural form of a type of skirt, hence the word “some”) + TS (= plural of “T”, an abbreviation for “T-Shirt”)

The surface meaning is highly misleading; additionally, the interpretation of IRON relies on a linguistic ambiguity (homonym employing different part of speech - noun, not verb).

Clue 1(b) Grown-up kid starts to gossip on aunt's Twitter (4)

The definition is “Grown-up kid” = a misleading circumlocution for GOAT

The wordplay plays on the word “starts” (in the nounal sense of “leading letters,” not verbal sense of “begins”) as an acrostic indicator: “Gossip On Aunt's Twitter.”

Clue 1(c) Scrub the cooker top and clean out (6) - (Cleary, 1996, from the Guardian, No. 20248, 26 Jan 1995)

The definition is “Scrub” = CANCEL, a non-prototypical interpretation.

The wordplay is a complex anagram of “C” (= “the cooker top” i.e. its initial letter) + CLEAN. The anagram indicator is the word “OUT.”

An important secondary function of the wordplay is to guide the solver away from the required definition of the target word, and to strongly promote the more prototypical sense “Scrub = Clean” by contextual means (Cleary, 1996).

Wordplay elements (Friedlander and Fine, 2016)

The algebraic/programming nature of the cryptic clue means that wordplay components may be flexibly recombined or anagrammed to form new units, e.g.:

- $A+B = C$ (FAT+HER = FATHER)
- $rev(A) = B$ (TRAMS -> SMART)
- $anag(A+B) = C$ (CAT+HAT = ATTACH)
- $trunc(A) = B$ (CUTTER -> UTTER)

Clues usually contain an “indicator” identifying what type of transformation is required (Biddlecombe, 2009), but equally might be of a punning/novelty type (usually indicated by a question mark at the end of the clue).

CRYPTIC CROSSWORDS AS POTENTIAL TRIGGERS OF INSIGHT

Cryptic (British-style) crosswords afford a unique opportunity to explore the mechanisms of insight and the issues highlighted above within an existing, readily available puzzle format. Devised in the mid 1920's (Connor, 2014), cryptic crosswords employ an extensive variety of highly ingenious puzzle mechanisms, many of which also draw on shared characteristics with a range of other types of “insight problem” (see review below). One puzzle may thus encapsulate a wide range of these mechanisms, presenting a compendium of heterogeneous insight challenges unrivaled by any other insight puzzle format. Studying cryptic crosswords may therefore enable us to understand better the antecedents, solving processes and key triggers of the insight moment.

What Are “Cryptic Crosswords”?

The nature of the cryptic crossword has been described in some detail in an earlier paper (Friedlander and Fine, 2016), but key aspects are highlighted again below. Example cryptic crossword clues, together with an explanation of the cryptic instructions for achieving the required solution, are set out in **Boxes 1, 2, 4–6**.

Unlike their “straight definition” American cousins, the challenge of the British-style cryptic crossword lies not in the obscurity of the vocabulary to be retrieved, but in the

quasi-algebraic coded instructions which must be executed precisely in order to achieve the correct answer to the clue (Friedlander and Fine, 2016): see **Box 1**. Cryptic crossword clues usually comprise two elements: a straight definition, plus the cryptic instructions for assembling the required solution—the “wordplay” (Friedlander and Fine, 2016; Pham, 2016). It is not always obvious which part of the clue is fulfilling what role, and there is often no clear division between the two parts (Friedlander and Fine, 2016). Even the “definitional” element of the clue might be obliquely or whimsically referenced, consciously exploiting ambiguities such as grammatical form, phrasal semantics, homophones, synonyms, and roundabout expressions (Cleary, 1996; Aarons, 2015; Friedlander and Fine, 2016). The clue type also has to be identified and interpreted. All these factors mean that that cryptic crosswords are typically ill-defined in both problem conceptualization and solution methodology (Johnstone, 2001).

Each cryptic crossword clue is thus a tricky linguistic puzzle using non-literal interpretations of deconstructed clue components in a “truly slippery and fundamentally ambiguous” fashion (Aarons, 2012, p. 224), stretching the conventions of everyday speech at all levels of structure and context (Aarons, 2015). The misdirection is deliberate: the surface reading of the clue evokes our tacit knowledge of language to suggest a plausible, yet unhelpful, interpretation of the clue (the “red herring”), setting up a constraint which must be resolved for progress

BOX 2 | Illustration of cryptic clue mechanisms: jokes and puns.**Clue 2(a) Frightened to death? (6,5)** - (Cleary, 1996)

Answer = SCARED STIFF, with a punning reference to “STIFF” = “corpse,” confirming the correctness of the solution.

Clue 2(b) Discovered why electrical equipment was dangerous? (9) - (Collingridge, 2010)

Answer = UNEARTHED (the latent secondary sense relates to electrical wiring)

Clue 2(c) Yorkshire beauty queen, we hear, pulls the wool over one's eyes (8) (“Orlando,” in Connor, 2011b)

Answer = MISLEADS. The pun (“Miss Leeds”) is indicated by a homophone indicator “we hear,” common in joke-style clues.

Clue 2(d) A wicked thing? (6) - (Aarons, 2015)

Answer = CANDLE. The clue relies on the two different homographic senses of the word “wicked.” Difficulty is heightened by the distinctly different pronunciation (/wɪk'ɪd/; /wɪkt/) and by the non-prototypical sense of “wicked” which is required (= “possessing a wick”). As in most punning or riddle-style clues, the quirky or nonsensical nature of the answer is flagged by the use of a question mark, which serves as a clue-type indicator.

to be made (Aarons, 2015; Friedlander and Fine, 2016). Once accomplished, the “Aha!” experience is triggered: this is termed the “Penny Dropping Moment” or “PDM” by crossword solvers (Friedlander and Fine, 2016).

In this use of misdirection, cryptic crosswords are similar to magic tricks: in both areas, the practitioner exploits implicit assumptions of the audience which are activated highly automatically, either (in magic) because of long-term exposure to the natural laws governing everyday life, such as gravity (Danek et al., 2014b) or (in crosswords) because of a lifetime's parsing habits as a reader and interpreter of standard text (Schulman, 1996). The task of the setter, as for the magician, is to conceal the clue mechanism so subtly that the pathway is not readily detectable (Friedlander and Fine, 2016).

Once deconstructed in this manner, there is no requirement for the cryptic components to make further sense as a coherent whole: the beguilingly smooth surface reading of the clue is typically abandoned in favor of a *potpourri* of dissociated cryptic fragments, each serving a quite different purpose entirely ungoverned by word-order, grammatical or orthographic considerations (Pham, 2016). In this way cryptic crosswords can be seen as a type of “non-*bona fide* communication” (Aarons, 2015, p. 357): the solver understands that the normal rules of communication must be temporarily suspended (just as they are required to suspend disbelief at a magic show), and that the clue itself is simply a vehicle for the intellectual challenge of solving the clue.

Range of Cryptic Clue Challenges and Parallels With Other Insight Problems

Although there is general agreement that the clues have to be fairly constructed (i.e., unambiguously solvable), there are no hard-and-fast guidelines as to what the rules of engagement are (Aarons, 2015; Friedlander and Fine, 2016), leading to an almost infinite number of innovative ways to exploit the “versatile and quirky English language” (Connor, 2013). Nevertheless, there is some consensus over a number of basic mechanism types, and a range of “Teach-Yourself” primers exist (Friedlander and Fine, 2016: see also now the on-line solving channel - Anthony and

Goodliffe vlog, n.d.). A brief review of the most striking parallels between a variety of insight puzzles and the mechanics of solving cryptic crosswords follows.

Jokes and Cryptic Crosswords: Deliberate Misdirection

Individual differences in the ability to appreciate humor have been previously identified (Cunningham and Derks, 2005; Kozbelt and Nishioka, 2010; Dunbar et al., 2016) and cryptic crossword solvers appear to be particularly attuned to and to enjoy verbal ambiguity and wordplay. In a study involving solvers and non-solvers (Underwood et al., 1988) the strongest correlation associated with cryptic puzzle-solving was the frequency of incidentally elicited laughter during an experiment involving associative priming (e.g., “strawberry” priming “traffic” through the unrepresented word “jam”).

Linguistic jokes share many characteristics with cryptic crosswords, including deliberate misdirection (Aarons, 2015), and—although only rarely used as such in the lab—jokes have been identified as a type of insight puzzle (Gick and Lockhart, 1995; Ramachandran, 1998; Robertson, 2001; Kounios and Jung-Beeman, 2009; Kozbelt and Nishioka, 2010; Amir et al., 2015) on the basis of the suddenness and rapidity of the solution, the lack of “feeling-of-warmth,” the pleasant feelings evoked at the moment of understanding, and the feeling of certainty in the correctness of the solution. A punning joke is typically based on two alternative interpretations of a scripted feed-line, which are both plausible in some sense, however absurd, “until the punchline, which highlights the initially less obvious one, and reveals the other to be a dummy, designed intentionally to mislead the listener” (Aarons, 2015, p. 352).

Working in a parallel tradition to that of psychological insight studies, linguistic humor studies have long explored the operation of jokes in the context of a two-stage process of “Incongruity-Resolution” (for a review see Forabosco, 2008), which shares many points of similarity with RCT. “Incongruity-Resolution” proposes that the expectations of the joke's audience

BOX 3 | Rebus puzzles.**3(a) poPPd** (MacGregor and Cunningham, 2008)

Solution: “Two peas in a pod”: auditory pun on “P” = “pea,” together with spatial location of the letters inside the word “pod.”

3(b) TIMING TIM ING (Smith and Blankenship, 1989)

Solution: “Split second timing”: the second instance of “timing” is split into two parts.

3(c) M CE /M CE /M CE (Salvi et al., 2016b)

Solution: “Three Blind Mice”: the mice have no “I”s (eyes)

3(d) R. P. I. (MacGregor and Cunningham, 2009)

Solution: “A grave error” (it should have been written as R.I.P.)

are deliberately manipulated to predict a sensible, but incorrect outcome, making the actual punchline initially unexpected or incongruous (the “surprise” phase). In the second phase (termed “coherence”), the listener then engages in a rapid form of problem-solving in order to revisit and resolve the incongruity, enabling the punchline to make plausible sense once it has been reconciled with an amusing and perhaps off-beat alternative interpretation of the original joke setting (Suls, 1972; Bartolo et al., 2006; Forabosco, 2008; Hurley et al., 2011; Canestrari and Bianchi, 2012). In other words, they must backtrack to search for an implicit constraint in their interpretation of the joke wording, which can be relaxed sufficiently to accommodate both the joke setting and its punchline within a revised interpretative structure (Suls, 1972; Navon, 1988). This process takes only a short time: there is an inverted relationship between speed of appreciation and funniness ratings (Cunningham and Derks, 2005; Kozbelt and Nishioka, 2010), and a joke falls flat if the explanation is too labored (Kozbelt and Nishioka, 2010).

If interpreted literally, the initially less dominant meaning (“latent content”—Kozbelt and Nishioka, 2010; Erdelyi, 2014) underpinning the correct interpretation of the punchline is often inappropriate, impossible or surreal: an “as if” resolution (Navon, 1988; Amir et al., 2015) which is “seemingly appropriate but virtually inappropriate” (Navon, 1988, p. 210) and—as for cryptic crosswords and magic tricks—functions “only on account of a willing suspension of disbelief” (Attardo et al., 2002, p. 5). It is at this point that we experience the emotional payback, as we “get” the joke, with the sudden, absurd resolution eliciting laughter; recent studies have begun to explore the neural correlates of these humorous insight moments (Amir et al., 2015; Chan, 2016).

The workings of this mechanism are exemplified in the following joke:

*‘So, I bought some animal crackers, and the box said:
“Do not consume if the seal is broken”...’ (attrib. Brian Kiley)*

Here, the listener is primed to interpret the term “seal” in terms of the intact packaging containing the foodstuff. The punchline seems incongruously out of place given that a joke is ostensibly being recounted: it appears to be a banal repetition

of standard wording commonly found on packaged goods, and is not inherently amusing. The feeling of “missing something”—that “nagging” sort of anxiety when you sense that something is funny-huh” (Hurley et al., 2011, p. 79) evokes an uncomfortable state of incongruity akin to cognitive dissonance (Festinger, 1957; Forabosco, 2008; Yim, 2016), and this discomfort will provide the motivational drive to reconcile or reduce the perceived inconsistency by reassessing the initial interpretation of the joke setting. It is only upon reinterpreting the word “seal” (in the context of “animal crackers”) that the alternative and nonsensical latent content of the joke emerges: that the crackers should not be eaten if the seal biscuit is broken.

Similarly, the cryptic crossword clue at **Box 2a** leads initially to a deceptively straightforward solution (“Scared stiff”), which perhaps only subsequently reveals the underlying pun “Stiff—> Corpse—> Frightened to death,” confirming the accuracy of the solution.

Fundamental to punning humor of this nature is the concept of “bisociation”—the perceiving of a situation in two incompatible frames of reference (Koestler, 1964; Dienhart, 1999; Canestrari and Bianchi, 2012). Following this account, ambiguous phonetic forms such as homophones, homonyms, and polysemes can act as triggers which abruptly switch the listener from one semantic script (e.g., “seal = box packaging”) to another (e.g., “seal = biscuit shape”). Koestler sees this as a sudden “Gestalt” reversal (Koestler, 1964).

Key to the workings of the joke or crossword clue is the initial concealment of the alternative meaning; and indeed it is a general feature of insight puzzles that the solution typically involves a statistically infrequent response, such as an unusual use for an object, or a less familiar, less dominant meaning for a word or phrase (Dominowski, 1995). So, for example, the cryptic crossword clue at **Box 2b** requires the solver to recognize that a potential solution word (“unearthed”), in its prototypical sense of “discovered,” has a second, non-intuitive but highly appropriate role to play in the clue (“without an earth wire”).

The cryptic crossword solver is thus often gulled into a readily available, but false interpretation of the clue setting (the “surface reading”) based on a *prima facie* interpretation of everyday linguistic rules, ambiguous phonetic forms, learned phraseological conventions, and context. This approach leads initially to nagging puzzlement, impasse and cognitive

BOX 4 | Illustration of cryptic clue mechanisms: rebus-like components.

Clue 4(a): Player with only one leg? (4) (Guardian Crossword No. 25351, by Tramp; 17 June 2011)

Answer = IPOD, a type of music player.

The clue works by comic analogy to “TRIPOD,” with the letter “I” standing in for the numeral “one.” This is very similar to the rebus puzzle at **Box 3a**.

Clue 4(b): Must’ve? (5,7,2,3,3) (Guardian Crossword No. 25351, by Tramp; 17 June 2011)

Answer = THINK OUTSIDE OF THE BOX.

Wordplay: MUSE [think] outside of TV [“the box”] - a rebus-like construction, also telling the solver what he must literally do to solve the clue. The punctuation is a highly distracting feature.

Clue 4(c): Part of it ‘it an iceberg (7) - (Moorey, 2009)

Answer = TITANIC.

Wordplay: substring(A+B+C+D) leading to a hidden word, indicated by the instruction “Part of.” The Titanic did indeed hit an iceberg, making this an “&Lit” (or “all-in-one”) clue: the clue as a whole functions as both the definition and the wordplay (Manley, 2014; Aarons, 2015).

Clue 4(d): GEGS (9,4) - (A well-known but unattributed clue, see Aarons, 2015).

Answer = SCRAMBLED EGGS. There is no guidance in the clue: the solver must literally “say what they see.” Compare the rebus examples **3(b)** and **3(c)** in **Box 3** above.

Clue 4(e): H,I,J,K,L,M,N,O (5) - (Another old chestnut of uncertain provenance, see Aarons, 2015)

Answer = WATER. Wordplay: “H to O”, if spoken aloud, sounds like H₂O.

Clue 4(f): Somewhat swollen condition of female diving bird? (9) - Times 24451, Feb 3rd 2010

Answer = PUFFINESS = “Somewhat swollen condition”

Wordplay = a quirky charade of PUFFIN + “-ESS” suffix, often indicative of a female in an animal species (e.g., “lioness”).

dissonance, since the original interpretation cannot be made to yield the desired answer (the solver is “missing something”). This provides the motivation to detect and explore alternative interpretations (some perhaps fruitlessly) in order to arrive at the moment of insight. As with jokes, the cryptic crossword’s “pay-off” (the final understanding of the clue) arrives when the original constraints are abruptly overturned in favor of a switch to an alternative, non-intuitive reading of the cryptic elements—often leading to surprise, laughter and the delight of the PDM (Aarons, 2015). No matter how lengthy and difficult this problem-solving phase has been, the clue is typically processed rapidly once the constraint is cracked (Topolinski and Reber, 2010a).

Rebus Puzzles and Cryptic Crosswords: Reinterpretation of Visual/Spatial Elements

Although many cryptic crossword clues rely heavily on punning misdirection, many also employ clue mechanisms which indicate that letters or letter blocks must be transposed, reversed, removed, substituted, extracted from a sequence or read as an acrostic (Aarons, 2015). In these clues, the elements providing the wordplay fodder must be decontextualized from the natural surface reading, either abandoning meaning altogether, or taking on new meaning of their own. Once these problem-irrelevant “chunks” have been decomposed (Knoblich et al., 1999) the components are redeployed in quasi-algebraic fashion to form new units answering to the clue definition (Friedlander and Fine, 2016): see further **Box 1**.

One clue type of this nature is the “charade”: a type of riddle in which the whole word is hinted at enigmatically by reference to its component syllables (Chambers, 2014). In this process,

cryptic crosswords may not observe morphological rules: for example, the word “discourage” would be segmented linguistically as “dis-courage,” but in a cryptic crossword might be clued, as “*Di (girl’s name) + scour + age*” (Aarons, 2015). See further clues 1(a) and 4(f) in **Boxes 1, 4**.

Similarly, rebus puzzles rely on the manipulation of words and word fragments to suggest common phrases which fit the clues displayed in a “word-picture.” Common rebus types involve charades, the interpretation of the spatial locations of words in relation to each other, typographical trends (letter size growing, decreasing), font size or color (capitalization etc.), numbers, and letters as words (MacGregor and Cunningham, 2008; Salvi et al., 2016b): see examples in **Box 3**. Rebus puzzles are also examples of ill-defined problems (Salvi et al., 2016b): the mechanisms for achieving the problem solution are unclear to the solver, who may have to try multiple strategies before hitting upon a productive approach. As with cryptic crosswords, the solver has to relax the ingrained rules of reading in order to overcome their tacit understanding of word-form and contextual interpretation and to achieve a restructuring of the problem space (Salvi et al., 2016b). For this reason, they are likely to trigger the insight experience (MacGregor and Cunningham, 2008; Salvi et al., 2016b).

Rebus puzzles typically rely on the literal and quirky interpretation of encrypted elements and their spatial arrangement, which are interpreted as part of the solution (MacGregor and Cunningham, 2008). In the British TV programme “*Catchphrase*,” which was based upon the solving of pictorially displayed rebus-type puzzles, the host, Roy Walker, used the tag line “Say what you see” in order to prompt contestants to find the solution (Wikipedia, 2017b). This is

BOX 5 | Illustration of cryptic clue mechanisms: Anagram clues.**5(a) Tube taken to theatre for three-act play (8)** (Aarons, 2015, p. 371)

ANSWER = CATHETER (=“Tube taken to theatre”).

Letter fodder = THREE-ACT; anagram indicator = “PLAY”.

There is heavy misdirection drawing the solver away from the required medical context and into theatrical performance and the “London Underground” (the “Tube”).

5(b) Doctor Watson's kit - or bits of modern office furniture (12) (Biddlecombe, 2009)

ANSWER = WORKSTATIONS (“bits of modern office furniture”)

Letter fodder = WATSON'S KIT OR; anagram-indicator = “Doctor”

Misleading disguise of anagram indicator in the name “Doctor Watson”, making the parsing of the clue unclear.

5(c) Find rare new frequencies beyond the visible range (8) (Johnstone, 2001, p. 70)

ANSWER = INFRARED (“frequencies beyond the visible range”)

Letter fodder = FIND RARE; anagram indicator = NEW

Johnstone points out that solvers often write out candidate letters as shown below, in order to facilitate the solving process:

I	F
N	R
A	R
E	D

precisely the approach needed by a number of the rebus-style cryptic crossword clues in **Box 4** which use highly inventive gimmicks to cryptically represent the solution word (clues 4 b-e).

Anagrams and Cryptic Crosswords: Dechunking, Pattern Detection, and Misdirection

Anagrams have been routinely used in investigations of insight (for a review, see Ellis et al., 2011)—both for anagram solving (e.g., Novick and Sherman, 2003; Kounios et al., 2008; Salvi et al., 2016a) and through the use of a paradigm requiring a simple judgment as to whether the anagram was solvable or not, in order to explore “feelings of warmth” and solution speed (e.g., Novick and Sherman, 2003; Topolinski and Reber, 2010b).

Studies of anagram solution have consistently reported that solvers approach anagram problems using two different strategies (e.g., Novick and Sherman, 2003; Kounios et al., 2008; Ellis et al., 2011; Salvi et al., 2016a): a search methodology, using a process of serially testing out and rejecting solutions based on morphemically probable letter combinations; and “pop-out” solutions (Novick and Sherman, 2003) whereby the solution bursts suddenly into consciousness without apparent work, often almost instantaneously. EEG research has demonstrated that self-reports distinguishing between “pop-out” and search anagram solving are reliably accurate (Kounios et al., 2008); this study also provides evidence that individual differences determine the solver's preferred strategy, and that different patterns of brain activity are associated with the two approaches.

It is well-established that structural features of the letter stimuli which are to be anagrammed (such as whether they are pronounceable, or form a real word in their own right) affect the difficulty and solution times of the puzzle. Thus, ZELBA or OARLY should be more difficult to resolve than HNWEI or

AOSLR; and HEART should be more difficult to unscramble than THREA (Dominowski, 1969; Novick and Sherman, 2008; Ellis and Reingold, 2014; for a review see Topolinski et al., 2016). Dominowski suggests that the pronounceability of the letters leads solvers to deal with them as a unit rather than as a letter-sequence (Dominowski, 1969): in other words, that familiarity with the letter patterns sets up an obstacle to solution by accessing automatically stored “chunks” of data which will be inappropriate to the solution (cf. Knoblich et al., 1999). It is the decomposing of these chunks into component letters which paves the way to the solution.

Anagram clues are a staple of cryptic crosswords (Upadhyay, 2008b; Aarons, 2015, p. 371), being formed of the letters to be anagrammed (the “fodder”), an anagram indicator and the definition of the resulting word (see **Box 5**). The letter fodder is typically concealed in misleading word units, which will be unhelpful to the anagram solution as indicated above; for this reason, many solvers will write out the letter-fodder in a random arrangement (such as a circle), in order to try to break up the prior associations and allow new patterns to form (Johnstone, 2001—see **Box 5**). However, difficulty can also be heightened by misdirection in the surface reading and by heavy disguise of the anagram indicator.

Remote Association Puzzles and Cryptic Crosswords: Spreading Activation

The Remote Associates Test (RAT), originally developed as a test of creativity (Mednick, 1962), has been refined and updated on a number of occasions, resulting in several sets of test materials [Functional Remote Associates Test (FRAT) (Worthen and Clark, 1971); Compound Remote Associates (CRA) (Bowden and Jung-Beeman, 2003b)], and has been translated into a number of languages (Salvi et al., 2016b). The task challenge is

BOX 6 | Illustration of cryptic “double definition” clues: RAT-like mechanism.**Clue 6(a): Tea shop (5)** (Biddlecombe, 2009, attributed to Azed)

Answer = GRASS.

Synonym 1: “Tea” = slang for “marijuana” = GRASS

Synonym 2: “Shop” = slang for “betray to the police” = GRASS. “Shop” has to be taken as a verb in this meaning, in contrast to the nounal function in the clue itself.

Clue 6(b): Savings book (7) (Aarons, 2015, p. 365)

Answer = RESERVE.

Synonym 1: “Savings” = a RESERVE of money

Synonym 2: “book” = to RESERVE (a table etc.): again verbal (solution) rather than nounal (clue)

Clue 6(c): Quits flat (4) (Connor, 2011a, by Rufus)

Answer = EVEN

Synonym 1: “Quits” = “neither owing, nor owed” = EVEN: adjective, not verb

Synonym 2: “Flat” = “level” = EVEN: adjective, not noun

Clue 6(d): Left red wine in harbour (4) (Biddlecombe, 2009; Aarons, 2015, p. 366)

Answer = PORT, a triple-definition

Synonym 1: “Left” = “on PORT side”: adjective, not verb

Synonym 2: “Red wine” = fortified PORT wine

Synonym 3: “Harbour” = PORT

Clue 6(e) Soldier even fixed uniform (7) - Daily Telegraph 28392

Answer = REGULAR, a quadruple definition with a misleading military surface reading

Synonym 1: “Soldier” = REGULAR (i.e. member of permanent forces)

Synonym 2: “even” = “level” = REGULAR (adjective, not adverb)

Synonym 3: “fixed” = “at set intervals” = REGULAR (adjective, not verb)

Synonym 4: “Uniform” = “unvarying” = REGULAR (adjective, not noun)

for the participant to consider a triad of apparently unconnected words (e.g., *Cottage, Swiss, Cake*) and to come up with a fourth word (here *Cheese*) which is related to all three through some type of associative connective link.

Although no longer commonly used as a test of creativity *per se* (Salvi et al., 2016b), RAT are frequently used to study facets of creative problem-solving such as insight (Bowden et al., 2005; MacGregor and Cunningham, 2008; Cranford and Moss, 2012; Jarosz et al., 2012; Chein and Weisberg, 2014; Salvi et al., 2015; Webb et al., 2016), incubation effects (Smith and Blankenship, 1991; Cai et al., 2009; Sio and Ormerod, 2015), and fixedness upon the wrong solution (Smith and Blankenship, 1989, 1991).

RAT puzzles are thought to operate through a serendipitous spreading neuronal network (Collins and Loftus, 1975) akin to three ripples, whereby each triad member simultaneously but independently activates a retrieval search of semantic memory (Smith et al., 2012; Kenett et al., 2014; Olteteanu and Falomir, 2015). This global search operates as a multiple constraint problem, each cue word indicating a different attribute of the target word to be satisfied; the solution is arrived at by confluence of the ripples upon a jointly shared node (Gupta et al., 2012; Smith et al., 2013).

Alternatively, participants can adopt a more controlled generate-and-test strategy by considering just one of the three cues at a time, and testing out candidate solutions against each

constraint for suitability, to ensure all requirements are met (Bowden and Jung-Beeman, 2007; Smith et al., 2013). This type of analytic, step-wise process is associated with lower insight ratings and different patterns of neural activity and eye movements when compared to sudden, non-methodical solutions (Bowden and Jung-Beeman, 2003a, 2007; Subramaniam et al., 2009; Cranford and Moss, 2012; Salvi et al., 2016b; Webb et al., 2016).

Impasse in solving RAT puzzles can arise from a fixation upon incorrect words, particularly those which are closely associated, syntactically or semantically, with one or more of the target words, and which therefore spring easily to mind (Harkins, 2006; Gupta et al., 2012). This blocks access to more remotely associated words needed for the solution (Gupta et al., 2012). Indeed, fixation in RAT problem-solving can be deliberately induced by priming commonplace associations which are unhelpful to the correct solution of the problem (Smith and Blankenship, 1991).

Consequently, one factor leading to higher performance on RAT puzzles is the ability to avoid a bias toward high-frequency candidate answers, thus allowing more remotely associated possibilities to be accessed (Gupta et al., 2012). This accords well with Mednick's conceptualization of an uncreative person as one who possesses a “steep associative hierarchy” containing an initially high number of stereotypical responses which rapidly

tail off. By contrast, the highly creative individual will possess a “flat associative hierarchy” containing many more items, and fewer stereotypical responses (Mednick, 1962, p. 223). Creative individuals are thus argued to possess more associative links, leading to a more complex and less rigid lexical network (Gruszka and Necka, 2002; Kenett et al., 2014).

In general terms, RAT puzzles pose a similar challenge to the “definition” in cryptic crosswords, which may reference the target word with considerable concealment. In many cases, the sense required will not be the dominant association, but a secondary meaning (sometimes quite obscure) which will come much less readily to mind, and fixation upon the wrong sense is often deliberately induced by contextual means (Cleary, 1996—see **Box 1c**). Breaking free from the stereotypical interpretation in order to consider a range of potentially remote synonym options is therefore key to lighting upon the correct solution (cf. Dominowski, 1995).

Even closer to the format of the RAT puzzle, however, is the “double definition” clue (Biddlecombe, 2009; Connor, 2011a; Aarons, 2015), whereby the solver is presented with two words, both of which can be defined by the same polysemic or homographic solution word (Aarons, 2015; Pham, 2016). Occasionally, triad cryptic definitions (or even quadruple/quintuple) are also found (Connor, 2011a—see **Box 6**). As in jokes, double definition clues operate through “bisociation” and an unexpected pay-off: “the fun of seeing two disparate concepts suddenly become one” (Connor, 2011a).

Although the mechanism illustrated in **Box 6** is very similar to that of RAT puzzles (“What one word links the following words?”), cryptic double definitions present extra difficulties, introducing elements of misdirection which are generally absent in RATs. First, in a dyad pairing, the two words are typically selected to form a familiar but unhelpful phrase with meaning of its own (e.g., 6(a) “tea shop”), creating a distracting red herring (Connor, 2011a). This automatically triggered impasse must be resolved by decomposing the unhelpful “chunked” phrase into its component features, allowing for an alternative parsing of the problem elements (Knoblich et al., 1999). Secondly, at least one of the words is usually “multicategorical,” meaning that it can be used as different parts of speech in each of the clue and the solution (Aarons, 2015). Finally, the solver must identify the “double definition” mechanism unaided, since there is no clue-type indicator for this class (Upadhyay, 2008a). For all these reasons, double definitions can be one of the hardest clue types to crack (Connor, 2011a), requiring multiple constraining misconceptions about the meaning, form and function of the clue elements to be resolved.

Advanced Cryptic Crosswords

So far, this article has only considered cryptic clues which might appear in daily “block-style” cryptic puzzles (Friedlander and Fine, 2016). However, a second type of cryptic crossword—advanced cryptics—also exists, which raises the difficulty still further (Friedlander and Fine, 2016). Advanced cryptic crosswords are found in weekend newspapers and some magazines, and the grids generally use bars rather than blocked grids (Friedlander and Fine, 2016). Of these, the *Listener Crossword* is the most notoriously difficult, employing a high

degree of clue mechanism concealment, obscure vocabulary, grids of startling originality and a thematic challenge, often involving a number of tricky lateral thinking steps on the basis of minimal guidance (Listener Editorial Team, 2013; Alberich, n.d.). Solvers submit weekly solutions for the distinction of appearing on an annual roll of honor, but few achieve an all-correct year (Friedlander and Fine, 2016). The *Magpie*,¹ a monthly specialist magazine with five highly challenging advanced cryptic crosswords (and one mathematical puzzle) per issue, runs a similar all correct/roll of honor system, and is broadly of *Listener* standard (Friedlander and Fine, 2016).

It is difficult to pigeon-hole the challenges set by advanced cryptics: there is an acute thirst for originality among the aficionados of these puzzles which drives setters to produce ever more creative designs, mechanisms and themes which “require original thinking by the solver over and over again” (Anthony, 2015), and annual awards for the most admired crossword in the *Magpie* and *Listener* series are presented to setters on the basis of solver recommendation (e.g., the *Listener* “Ascot Gold Cup²”). However, two particularly prominent sources of challenge are described below.

Thematic Challenge: Acquisition of Incidental Hints

Many advanced cryptic puzzles contain a thematic challenge, lending extra difficulty to the puzzle. In one common approach, a number of thematically related entries may have no clue, requiring the solver to deduce the answers gradually from cross-checking letters, as the grid is populated. Additionally, entire areas of the grid—such as the complete perimeter—may need to be completed with thematically relevant items or messages. In other puzzles, letter sequences spelling out thematic material may be concealed in the grid (for example on the diagonals), requiring the solver to find and highlight them through a “wordsearch” process (Alberich, n.d.).

Thematic puzzles rely upon the solver’s ability to make cross-connections between seemingly disparate items drawn from unpredictable and often obscure fields of knowledge: in this they share similarities with lateral thinking quizzes such as BBC2’s *Only Connect* and BBC Radio 4’s *Round Britain Quiz* (Connor, 2016). Once again, the problem space is ill-defined: the nature of the connection, the goal state and the pathway to achieve coherence are all unspecified.

In order to solve these puzzles, solvers have to accumulate incidental information along the way: hints in the title or preamble might point obliquely to the theme; suggestive word fragments might appear in the grid, and thematic material might be gradually spelled out by other means—such as corrections to misprints in the clues. The PDM comes at the instant when all the disparate pieces of information suddenly come together to make sense. It is therefore comparatively rare for the theme to be deduced from the start (indeed this element of the puzzle is often termed the “endgame”): the solver must be able to tolerate—or even enjoy—the sensation of working for some time

¹<http://www.piemag.com/about/>

²http://www.listenercrossword.com/List_Awards.html

The Title is Wrong by Ifor

Answers to four clues have a thematically related connection (RAMSAY, FLORA, JEANETTE and MALCOLM - all turn out to share the surname 'MacDonald').

Four other answers are to be replaced with words, not necessarily of the same length: this resulted in the answers PIG, GOAT, SHEEP, COW being replaced by their noises (OINK, MAA, BAA, MOO) in the grid.

A third group of normal clues had entries containing one of the twelve letters which solvers had to highlight to 'denote the theme'. Careful examination of the grid enabled more astute solvers to spot the opening notes of *Old MacDonald* arranged as if on a musical stave - a gimmick based upon a punning definition of 'denoted'.



Corrections to misprints in the remaining clues spell out instructions for correcting the title: 'CLEAR ALL THE CONSONANTS'. Following this instruction modifies 'The Title is Wrong' to 'E I E I O', the refrain from the song. This had to be written below the grid in the final submission.

FIGURE 2 | Magpie crossword issue 130.4 (Ifor, 2013).

with unclear goals and incomplete, potentially conflicting and imprecise data. This may imply that advanced cryptic solvers tend toward personality traits such as a low "Need for Closure"—the desire for definite knowledge and resolution of an issue (Webster and Kruglanski, 1994); and a high "Tolerance of Ambiguity"—the perceiving of ambiguous situations as desirable, challenging, and interesting (Furnham, 1994; Zenasni et al., 2008). Earlier research (Friedlander and Fine, 2016) has also found that cryptic crossword solvers generally have a high "Need for Cognition," relating to a person's tendency to seek out, engage in and enjoy effortful thinking (see Cacioppo et al., 1984; Furnham and Thorne, 2013; Von Stumm and Ackerman, 2013).

An example of a thematic cryptic crossword challenge is shown in **Figure 2**. Here the well-known children's song "Old MacDonald Had a Farm" is used as a source of thematic material:

"the super-familiar hiding under a thick cloak of obscurity, waiting to reward the determined solver with a PDM that feels like a surprise from an old friend" (Editorial Notes, 2013, p. 10).

Given the richness of the thematic material in this puzzle, which is expressed through multiple different devices (MacDonalds, animal noises, EIEIO title and the notation in the grid), it is likely that solvers experienced a number of PDMs—a series of mini "insight moments"—en route to a final solution. Some PDMs would almost certainly have come out of the blue: in particular, the concealed instruction to correct the title by deleting consonants "hides in a simple statement of fact a truly surprising vowel-only 'correct' title that nobody could possibly have seen coming" (Editorial Notes, 2013, p. 10). The finding of the tune proved trickier:

"The common experience was an initial search (often for 'MacDonald'), followed by some confusion, followed by careful examination of the letters in the appropriate area, followed maybe by re-reading the preamble, combined with spotting some suspect letter duplications ... in other words, a penny that did drop, but did it slowly" (Editorial Notes, 2013, p. 10).

As with RAT puzzles, thematic challenges appear to operate through a ripple of spreading activation (Collins and Loftus, 1975). Each "clue to coherence" (Bowers et al., 1990) embodies a different attribute of the target connection to be made; when these unconscious activations achieve confluence, the pattern emerges quite suddenly into consciousness, leading to the perception of coherence, and the PDM (a process described as "intuitive guiding"—Bowers et al., 1990). Individual differences will again arise in the speed, complexity and gradient of the available interassociative connections (Bowers et al., 1990; Gruszka and Necka, 2002; Smith et al., 2012; Kenett et al., 2014).

Individual differences in the ability to assimilate chance hints may also be relevant: as Louis Pasteur famously remarked of his ostensibly fortuitous scientific discoveries, "Chance favors only the prepared mind" (Lecture, University of Lille, 7 December 1854—Seifert et al., 1995). "Opportunistic assimilation" (Seifert et al., 1995; Sio and Ormerod, 2015) refers to the ability to absorb new and serendipitously presented information, and to allow these additional jigsaw pieces to resolve or reframe one's understanding of a problem which has previously reached impasse. Much may depend on the initial preparation stage in which the solver becomes attuned to salient or important features they have already noted (Seifert et al., 1995; Ormerod et al., 2002) which they maintain at a heightened level of activation, leading to priming effects (Sio and Ormerod, 2015). Although potentially experiencing a number of failures and false leads in the process (Ormerod et al., 2002), progress is then made when the solver becomes intrigued by further patterns or anomalies (Kolodner and Wills, 1996), or stumbles across other relevant information (Weisberg, 2006) during completion of the grid.

The process is well-illustrated by the editorial feedback on *Magpie* 151/2 "Five-a-side (on Tour)" by Wan, which was themed around a subset of the 72 names of French scientists, engineers and mathematicians engraved on the Eiffel Tower (five from each side):

"In solving terms, there was a single critical, and memorable, moment of realization when the set of names suddenly made sense. This was normally preceded by a number of less memorable moments of thinking that there was some other reason for grouping, by nationality, or by specialization, or by university affiliation, or whatever. All the false trails had some value, because you were always going to be alert to French scientists or engineers once a few showed up. The feeling was of constant small steps forward, always with some difficulty, but never with that feeling of brick-wall despair that can accompany certain thematic endgames." (Editorial Notes, 2015, p. 9).

Individual differences in openness to experience and sensitivity to external stimuli could be relevant in these contexts, regulating the degree to which a person inhibits or remains subconsciously

receptive to ostensibly incidental information (Laughlin, 1967; Carson et al., 2003; Simonton, 2003; Weisberg, 2006; Carson, 2010; Russ and Dillon, 2011). A reduced tendency to pre-filter extraneous information as irrelevant (i.e., reduced latent inhibition) may enhance the ability to make lateral associations, and has been associated with both psychometrically and behaviorally assessed creativity, openness to experience, and richer, more diverse associative networks (Simonton, 2003; Carson, 2010).

Spatial or Transformational Challenges: Reconceptualizing the Layout

An additional source of difficulty in many advanced cryptic crosswords lies in the transformation of some elements. For example, some or all of the answers might need to be encoded or otherwise thematically altered before being entered in the grid. As in American-style "variety puzzles," such as those appearing periodically in the Sunday edition of the *NY Times* (Wikipedia, 2017a), this might involve anagramming, reversing or curtailing entries (resulting in non-words in the grid); but more complex adjustments might also be required. For example the solver might deduce that all overlong items, such as APHID (to fit a grid space of 3) and CHINWAG (to fit 5), might need to be entered using Greek characters to replace the English names for the Greek alphabet (i.e., AΦD and XNWAG Alberich, n.d.). Or all entries might need to be encoded using a Playfair cipher, with the keyword to be deduced (Upadhyay, 2015). Once again, the problem space is ill-defined: the solver has to assimilate key hints or salient features as the puzzle progresses in order to deduce what adjustments need to be made, and may pursue a number of false leads before hitting upon the correct solution. Meanwhile, the completion of the grid is made much harder by the absence of securely confirmed cross-checking letters while the entry mechanism remains unresolved.

Further to this, some advanced cryptics require a type of restructuring in which the dimensions, layout or salient features of the grid itself are changed (see **Figure 3**). In these puzzles, there is a need to reconceptualize spatial assumptions involving placement and layout constraints, and to dismantle an existing array in favor of a new, radically different format. Cunningham highlights these two characteristics as strong features of classic spatially-oriented insight puzzles such as the nine-dot problem, the ten-coin triangle and the chain necklace puzzle (Cunningham et al., 2009 - **Figure 1**). Difficulty is also heightened in many of these classic puzzles by the need to identify and verify what the eventual solution would look like (MacGregor et al., 2001; Cunningham et al., 2009): this prevents steady progress toward a concrete and visualizable goal state (MacGregor et al., 2001), even if the eventual solution criteria and constraints are made clear.

So, for example, in **Figure 3**, the solver is made aware by means of a hidden message that the grid must be cut up and reassembled; but the purpose of this transformation, the eventual grid layout and even the cutting line must all be deduced. Additional difficulty is introduced by the elliptical reference to a "saw"; given the need to cut the grid and the zig-zag nature of the cut, the required interpretation of the

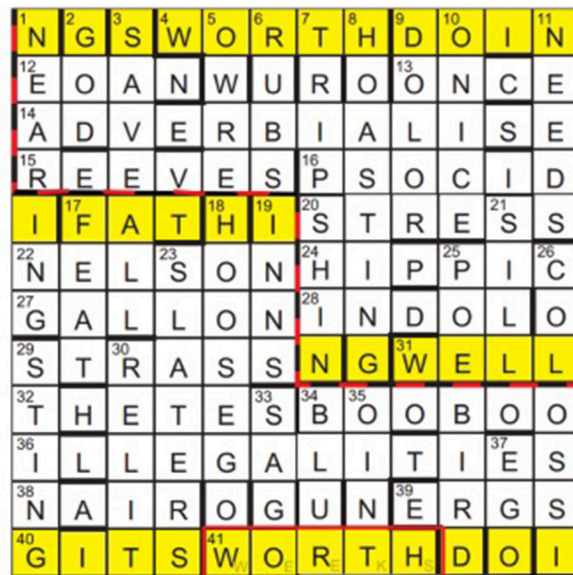
Saw by Chalicea

Instructions:

The wordplay in across clues yields the answer, along with an extra letter that is not entered in the grid.

In clue order these give instructions.

Before carrying these out, solvers need to change one word in the solution grid to the word that appears symmetrically opposite it in the grid, leaving real words.



Solution:

The original grid is completed as shown above, with one word replaced: WEEKS (41a) is changed to match WORTH (4a). The extra letters spell out the instruction 'CUT IN TWO, REFORM, SHADE SAW'. The solver must deduce the cutting line (shown by red dashes above).



Cutting along the red dotted line as indicated, in a step pattern, and reassembling the two halves in the new layout produces the 'saw' (= 'saying, proverb'): 'If a thing's worth doing, it's worth doing well'. All answers are real words.

FIGURE 3 | Magpie crossword issue 166.1 (Chalicea, 2016).

term ("saw" = a maxim, saying) might not spring to mind. Without understanding this hint, the unspoken endgame (that of reconstructing a well-known phrase along the top and bottom line) cannot be interpreted correctly.

INCIDENTAL SUPPORT FOR CRYPTIC CROSSWORD CLUES AS A FORM OF INSIGHT PUZZLE

The paper review set out above plausibly suggests that cryptic crosswords can function as insight problems, using a variety

of techniques, such as misdirection and an ill-defined problem space, to increase the likelihood of an "Aha!" response. However, following the methodology set out in the "Grounded Expertise Components Approach" (GECA—Friedlander and Fine, 2016), the first step in the current research program was to secure empirically based corroboration for this *a priori* assumption.

Confirmation was therefore sought as part of an 84-item broad-based questionnaire, intended to characterize the cryptic crossword solving population across a wide number of dimensions. The full methodology for this research was set out in a previous publication (Friedlander and Fine,

2016). In total, 805 solvers across the full range of solving ability took part, although there was some attrition toward the end of the survey. Solvers were objectively assigned to research categories on the basis of benchmarked criteria, resulting in both a 2-way (Ordinary/Expert—O/E) and a 3-way (Ordinary/High ability/Super-Expert—O/H/S) categorization of participant expertise. For full details of the categorization rationale, see Friedlander and Fine, 2016.

One key hypothesis of the survey was that “cryptic crossword solving regularly generates ‘Aha!’ or insight moments, supporting the hypothesis that the cryptic clue is a type of insight problem through misdirection; and that this pleasurable experience is a salient driver of cryptic crossword participation” (Friedlander and Fine, 2016, p. 7). To this end, the survey included a number of questions pertinent to the current discussion: results are presented below. All chi-square analyses are bootstrapped and 95% confidence intervals are reported in square brackets.

EVIDENCE FOR THE “PENNY-DROPPING MOMENT” (PDM) AND INCUBATION EFFECTS

PDM as a Motivating Experience

Participants were asked to rate 26 statements relating to their motivation for solving cryptic crosswords on a 5-point Likert scale (1 = “Completely Disagree”; 5 = “Completely Agree”). There were 786 responses (O: $n = 388$; H: $n = 221$; S: $n = 177$). **Table 1A** shows the five highest responses to these 26 statements (with abbreviated descriptions). As previously reported (Friedlander and Fine, 2016) all groups rated the “Aha!” moment (PDM) as a key motivational factor for solving cryptics; closely allied with this was the statement “Solving well-written clues gives me a buzz—it makes me smile or laugh out loud” which was ranked 4th in importance. The feeling of fulfillment—whether with the completed grid or with the “uniquely satisfying” cryptic crossword puzzle format—was also ranked highly (2nd and 5th most important). There were no statistically significant differences between the expertise groups for any of these statements. This suggests that—as for jokes—an important part of the crossword puzzle-solving experience lies in the pleasurable emotional reward bound up with the resolution of incongruity at the moment of insight. Studies of jokes and humor have found that laughter is associated with the release of endorphins which may be important in this context: the opiate effects of endorphins create a sense of wellbeing, pleasure and a sense of satisfaction (Dunbar et al., 2011). By contrast, extrinsic motivators, such as prizes, competitions, or public acclaim, were not important to participants across the board (Friedlander and Fine, 2016).

Incubation Effect

In a separate series of questions intended to capture the solving preferences of participants, respondents were invited to rate statements on a 3-way Likert scale (“No/Never”–“Perhaps/Sometimes”–“Yes/Always”; together with a null response option “Don’t know/Not applicable”). 796 responses were made (O: $n = 395$; H: $n = 223$; S: $n = 178$). Results are

given in **Table 1B**: figures represent the summed percentage of “Sometimes” and “Always” responses unless otherwise indicated.

Nearly 95% of solvers (94.6%; O: 95.7%; H: 95.5%; S: 91.1%) confirmed that “incubation effects”—setting the crossword aside for a while, in order to resolve periods of impasse—were a feature of the solving process. Indeed, 80.3% of participants agreed with the full “Yes” option: “*Yes—the answer is often obvious when I return to the crossword*” with a further 14.3% agreeing that “*I sometimes find it helpful to take a break, but I often return to the thoughts I was having previously.*” S solvers were least likely to have taken advantage of incubation breaks; even so, differences in the distribution of incubation effect between groups failed to reach statistical significance ($\chi^2_{(4)} = 8.681$, $p = 0.070$, Cramer’s $V = 0.074$ [0.040, 0.135]).

Conversely, S participants were most likely (84.8%) to have found that solutions occurred to them at least occasionally when they were engaged in totally unrelated activities (e.g. shopping, driving, taking a bath). Overall 79.8% of participants agreed with this statement (O: 77.4%; H: 79.9%; S: 84.8%), but differences between the groups again failed to reach statistical significance ($\chi^2_{(4)} = 5.393$, $p = 0.249$, Cramer’s $V = 0.058$ [0.032, 0.115]).

Impasse and the “Aha” Moment

Most participants also agreed that their enjoyment of the PDM was enhanced if they had needed to struggle with a clue (79.6%; O: 83.8%; H: 78.0%; S: 72.5%) although some respondents claimed that the “Aha!” moment was unaffected by the effort expended (16.3%; O: 13.7%; H: 17.0%; S: 21.3%). Very few participants claimed either that it decreased with effort expended (2.6%) or that they had never experienced a PDM (1.4%) when solving cryptics. Differences between groups approached, but did not achieve statistical significance ($\chi^2_{(6)} = 11.796$, $p = 0.067$, Cramer’s $V = 0.086$ [0.059, 0.153]) and inspection of standardized residuals indicated that this was driven by the higher number of S solvers in the “Makes no difference” group ($z = 1.7$).

DIFFERENCES IN SOLVING APPROACH BETWEEN CRYPTIC CROSSWORD EXPERTISE GROUPS

Participants were also asked about their approach to solving cryptics in order to explore potential differences between the expertise groups; **Table 2** highlights a number of key findings.

Suppression of the Misleading Surface Reading

Survey participants were asked to indicate whether they noticed the surface reading of a clue first, or read it purely as code. Two response options (“I always read the surface meaning first,” “I tend to read the surface first”) favored the surface reading; two options indicated that deliberate attempts were made to exclude “reading for sense” (“I try to exclude the misleading context,” “I always read as code: the surface meaning could be gobbledygook”); and there was one mid-way option

TABLE 1 | Responses by expertise category to questions about “insight” properties of crossword clues.

	O	H	S	All groups
(A) Top responses to “crossword motivation” question (mean scores, out of 5)				
Number of responses to question	388	221	177	786
1. Enjoy “Penny-Drop Moment”	3.92	3.92	4.07	3.96
2. Cryptics are uniquely satisfying	3.89	4.05	3.91	3.94
3. Mental exercise to keep brain sharp	3.88	3.83	3.85	3.86
4. Makes me smile or laugh	3.79	3.80	3.64	3.76
5. Satisfaction of filled grid	3.46	3.61	3.36	3.48
(B) % Participants agreeing with the following statements				
Number of responses to question	395	223	178	796
“Setting the crossword aside for a while helps” ^a	95.7	95.5	91.1	94.6
- “sometimes, though not always”	12.4	15.7	16.8	14.3
- “always (answer is obvious on return)”	83.3	79.8	74.3	80.3
“I have solved clues when I’m doing something else”	77.5	79.8	84.8	79.8
“The Aha! feeling is most intense after a long struggle”				
- “Yes”	83.8	78.0	72.5	79.6
- “No difference one way or the other”	13.7	17.0	21.3	16.3

^aThere were 797 responses to this question; *S n* = 179.

TABLE 2 | Differences in approach to solving cryptics.

	O	H	S	All groups
Number of responses	395	223	179	797
(% Participants agreeing with the following statements)				
(A) Do you notice the surface reading or the codes of a clue first?				
Surface first	33.2	25.6	25.7	29.4
Bit of both: surface and codes	50.4	42.6	38.0	45.4
Read as code, not for meaning	16.5***	31.8*	36.3**	25.2
(B) What do you look for in an Advanced Cryptic crossword?^a				
I don’t do Advanced Cryptics	n/a	12.1*	2.8*	8.0
Great clues	n/a	35.9*	16.8**	27.4
Good balance of clues and endgame	n/a	38.6	50.3	43.8
Tricky and satisfying Endgame	n/a	13.5*	30.2**	20.9
(C) Are you disappointed if you solve a crossword rapidly?^b				
No: I enjoy rapid solving	9.9	14.3	16.9	12.7
Don’t mind either way	41.8	33.6	41.0	39.3
Yes: I like to wrestle with the clues	48.4	52.0	42.1	48.0
(D) I Would change my crossword if the challenge got too easy (“Yes”)				
	70.1	71.7	66.3	69.7

(*/**/***indicates significance at the 0.05/0.01/0.001 level).

^aOrdinary solvers, by definition, do not solve Advanced Cryptic crosswords. %s relate to 402 participants (*H* = 223; *S* = 179).

^bThere were 796 responses to this question; *S n* = 178.

(“Bit of both; not sure which predominates”). There were 797 responses (*O*: *n* = 395; *H*: *n* = 223; *S*: *n* = 179); summarized details (Surface/ Bit of Both /Code) are given in Table 2A.

Most solvers (45.4%; *O*: 50.4%; *H*: 42.6%; *S*: 38.0%) selected the mid-way point, though this decreased with expertise: *S* solvers were most likely to suppress “reading for sense” in favor of “reading for code” (36.3%); the opposite was true for *O* solvers, who tended to read much more for sense (33.2%). Differences between the groups were significant ($\chi^2_{(4)} = 33.21$, $p < 0.001$, Cramer’s *V* = 0.144 [0.105, 0.199]) and inspection of

standardized residuals indicated that this was driven by higher levels of *H* (31.8%, $z = 2.0$, $p < 0.05$) and *S* (36.3%, $z = 3.0$, $p < 0.01$) solvers who suppressed the surface reading; and lower levels of *O* solvers who did this (16.5%, $z = -3.5$, $p < 0.001$).

Personal Preferences Leading to Greater Enjoyment of Advanced Cryptic Crosswords

Solvers were asked to identify whether they solved Advanced Cryptic crosswords, and, if so, whether the quality of the clueing

or the tricky endgame (or a bit of both) was their primary source of enjoyment (**Table 2B**). A small proportion of both expert groups chose not to solve Advanced Cryptic crosswords, although this was higher for H solvers than for S (“I don’t do Advanced Cryptics”: 8.0%; H 12.1%; S 2.8%). O solvers, by definition, do not solve this type of crossword (Friedlander and Fine, 2016, p. 8) and were omitted from this analysis. Where a preference was indicated, for H solvers the quality of the clueing was paramount (27.4%; H 35.9%; S 16.8%) whereas, for a larger number of S solvers, the lateral-thinking endgame was the most important attraction (20.9%; H 13.5%; S 30.2%). Differences between the groups were significant ($\chi^2_{(3)} = 40.47$, $p < 0.001$, Cramer’s $V = 0.317$ [0.226, 0.407]) and inspection of standardized residuals indicated that this was driven by higher levels of H (12.1%, $z = 2.2$, $p < 0.05$) and lower levels of S (2.8%, $z = -2.5$, $p < 0.05$) who failed to tackle Advanced Cryptics; higher levels of H (35.9%, $z = 2.4$, $p < 0.05$) and lower levels of S (16.8%, $z = -2.7$, $p < 0.01$) whose main target for enjoyment was the smooth clueing; and higher levels of S (30.2%, $z = 2.7$, $p < 0.01$) and lower levels of H (13.5%, $z = -2.4$, $p < 0.05$) whose primary focus was the endgame.

Speed-Solving and Challenge

Solvers were also asked whether they would be disappointed if they solved a crossword rapidly (**Table 2C**). Although chi-square showed a significant association overall ($\chi^2_{(4)} = 9.99$, $p = 0.041$, Cramer’s $V = 0.079$ [0.050, 0.139]), inspection of the standardized residuals revealed no stand-out elements. As expected, S solvers (among whom were a number of competition-focused “Speed Solvers”—see Friedlander and Fine, 2009) would be least troubled by a rapid solve (“No: I enjoy speed-solving”: 12.7%; O 9.9%, $z = -1.6$; H 14.3%, $z = 0.7$; S 16.9%, $z = 1.6$), but, even for this group, numbers were low, and standardized residuals were non-significant. Nearly half the solvers indicated that they would be disappointed without a good challenge to wrestle with, and although there was some variation across the expertise groups (48.0%; O 48.4%, $z = 0.1$; H 52.0%, $z = 0.9$; S 42.1%, $z = -1.1$) inspection of the standardized residuals were once again non-significant.

Indeed, when asked whether they might switch newspapers if the crossword challenge became routinely easy (**Table 2D**), nearly 70% of solvers indicated that they would consider this (69.7%; O 70.1%; H 71.7%; S 66.3%), with differences between the groups being statistically non-significant.

POTENTIAL CONTRIBUTION OF CRYPTIC CROSSWORDS TO INSIGHT RESEARCH

The above review suggests that the cryptic crossword domain could prove a useful addition to the repository of insight problem paradigms. That they are capable of triggering insight on a regular basis is quite clear: survey results reported above indicate that cryptic crossword solvers were primarily

motivated to solve cryptics because of the “Aha!” or “Penny-Drop” moment, and also reported that the “laugh-out-loud” moment at the point of solving the clues was highly enjoyable. Furthermore, the detailed review of cryptic clues set out above demonstrates that they use a broad variety of insight-triggering mechanisms shared in common with a wide range of other insight problem formats. A single cryptic crossword puzzle thus presents a unique compendium of heterogeneous challenges which sets it apart from all other methodologies currently available; and this should facilitate the comparison of outcomes between device types within the crossword itself, as well as with other insight puzzle challenges external to the crossword.

One small caveat is that cryptic crosswords are primarily restricted to a number of English language speaking countries, although a few cryptic type puzzles do exist in Dutch and German. This may reduce the flexibility of cryptic crosswords as an insight puzzle paradigm. Straight-definition crosswords are, of course, available in all languages, but lack the cryptic elements described in detail in this paper which set this puzzle form apart and trigger the insight moment.

Cryptic crossword clues thus reliably trigger insight experiences, but (as for all insight puzzles) this is not exclusively the case. In cryptic crossword trials filmed for transcription using Verbal Protocol Analysis (VPA), casual inspection of the recordings suggests that not every clue produces as many PDMs; and not every solver follows the same path to solution. Systematic analysis of the video recordings (on which see further Friedlander and Fine, 2016) will allow us to take full advantage of the think-aloud protocol to capture a wide range of strategically important factors such as intuitive vs. analytical approaches to clue solution; the length of time spent in impasse on each clue before moving onto another; the frequency of return to an obstinately resistant item; perseveration with an incorrect solution pathway; the antecedents of “Aha!” solution moments; the use of cross-checking letters as opportunistic solution prompts; the suppression of the surface meaning on initial reading; the certainty of correctness (without double-checking) on solution; and the use of jottings such as candidate anagram letters (see **Box 5** above) to facilitate solution (on the use of VPA in the GECA methodological approach, see further Friedlander and Fine, 2016). These aspects are all highly relevant to the discussion of insight problem solving across a wide range of problem domains.

As a precursor to the analysis, the clues used in the crossword trials will be individually analyzed to identify salient features, such as the mechanisms employed, the level, and number of the constraints preventing solution, and the predicted difficulty which flows from this (following e.g., Knoblich et al., 1999; Cunningham et al., 2009; MacGregor and Cunningham, 2009). It is very possible that the clues vary in difficulty on a principled basis, and if so, this might lead to a better understanding of what makes a cryptic crossword clue enjoyable, and more likely to trigger insight, to lead to impasse, or to invoke “Immediate Insight” solutions. Given the cross-over between cryptic crossword clue types and other insight puzzles, this

should shed helpful light on insight mechanisms in other areas, too.

Logistically, cryptic crosswords also offer a number of advantages over other puzzle types. In the first place, there is no lack of material: cryptic crosswords appear daily in all of the British newspapers, and widely across the world in countries with historically strong connections to Britain (e.g., Canada, Ireland, Australia, New Zealand, India, and Malta: Friedlander and Fine, 2016). It is thus entirely possible to commission a professionally composed, high-quality puzzle specifically for a research study thus guaranteeing that all participants will be naïve to the challenge. Clue solution rates are high, too: in trials involving 28 solvers (both expert and non-expert) tackling a commissioned 27-clue crossword of medium difficulty, 682 of the 756 clues (90.2%) were solved correctly within the 45 min time limit (Fine and Friedlander, in preparation). Solving times for those who finished the entire puzzle ($n = 19$) could be very rapid indeed (range solving times: 10m47s–40m30s; mean solving time for finishers 23m:43s, median 22m15s) resulting in solutions occurring, on average, approximately once a minute (Fine and Friedlander, in preparation).

Fast solvers in this trial were all highly expert in the field (Fine and Friedlander, in preparation), and the survey results set out above also indicate that experts may approach the solving of cryptic clues in subtly different ways to less expert solvers of equivalent experience. What could be seen as a disadvantage for this methodology (that cryptic crossword solving is a niche activity requiring inside knowledge of and experience with the clue mechanisms) thus becomes a compelling strength: there is much that might be gained from studying expert insight puzzle solvers at work, and this is currently impossible in other insight domains (such as RAT puzzles or matchstick math) which, by necessity, always use naïve populations.

Lamenting the lack of expertise studies in the insight area, Batchelder and Alexander (2012) even suggested artificially training groups of individuals to produce “expert” solvers of such problems, commenting that experts “might have the capacity to rapidly shift their search spaces until the type of space that contains the solution occurs to them” (Batchelder and Alexander, 2012, p. 88). However, this proposal overlooks the potential role of individual differences: MacGregor and Cunningham argue that there may be reliable variations in the ability of individual subjects to solve insight problems (2008; see also DeYoung et al., 2008; Ovington et al., 2016) which may undermine the ecological validity of training “experts” from a randomly selected sample of individuals. Within the crossword field we found naturally-occurring expertise groupings—all with equivalent levels of experience over many decades in the field, but with quite different expertise outcomes (Friedlander and Fine, 2016)—and this presents a unique opportunity for exploration.

The cryptic crossword survey data set out in Tables 1, 2 above hints at some interesting differences between the various expertise groups and their approach to solving this form of puzzle. Most intriguing of all is the possibility

that experts have an enhanced capacity to resist the red-herring set for them, by electively divorcing the reading of the clue from its surface meaning (“the surface meaning could be gobbledygook”), and thus shielding the mind from the deliberate misdirection. Whether expert solvers therefore experience the full phenomenological experience of the “Aha!” moment upon solution of the clue is thus an interesting angle for further investigation: experts claim to be equally motivated by the promise of the “Aha!” moment (Table 1), yet, paradoxically, appear to suppress that very need for Representational Change which might have been considered fundamental to the insight experience. Experts also solve more rapidly, with speed prowess being a primary focus for some (Friedlander and Fine, 2009), and this affords an opportunity to explore rapid “pop-out” solutions and the relevance of “Immediate Insight” to the exploration of the “Aha!” moment.

It is also notable that significantly more Super-Experts engage in Advanced Cryptic puzzles than High Expert solvers, and that their primary focus in doing so is significantly more often linked, not with the appreciation of the smooth misdirection of the clueing itself, but with the complexity, novelty and lateral thinking challenge of the Advanced Cryptic endgame, which is more akin to the “classic” insight puzzle format in its use of thematic or spatial features. This again affords opportunities to examine the multi-dimensional nature of the demands posed by different insight problem types, as described in the body of this article, and the interplay with individual differences shown by problem solvers, in terms of their thinking and personality styles.

CONCLUSION

In sum, this preliminary review suggests that cryptic crossword puzzles may be a promising source of insight problems offering a number of potential advantages over some of the puzzles and riddles previously used: for example, they are readily obtainable in potentially unlimited supply, solvable within acceptable time limits and suited to the simultaneous exploration of a variety of puzzle types and their potentially distinct solving mechanisms. Uniquely among existing paradigms, they also afford us the opportunity to study insight-solving expertise in action and to identify the characteristics and methodological approaches of those with a particular propensity to solve these puzzles effectively. There is therefore much to explore, and the discussion above suggests a number of particularly interesting avenues which we are currently pursuing. We believe that this new paradigm may prove to be a useful source of theoretically and empirically grounded, heterogeneous insight challenges; and that it is well-placed to shed a unique light on the workings of this elusive and intriguing aspect of human cognition.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the British Psychological Society. All

subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the School of Science and Medicine Ethics Committee, University of Buckingham.

AUTHOR CONTRIBUTIONS

KF drafted the article and KF and PF reviewed and finalized it. KF designed the survey and analyzed data via an Access database. KF and PF reviewed data and agreed coding treatments.

REFERENCES

- Aarons, D. L. (2012). *Jokes and the Linguistic Mind*. London: Routledge.
- Aarons, D. L. (2015). Following orders: playing fast and loose with language and letters. *Aust. J. Linguist.* 35, 351–380. doi: 10.1080/07268602.2015.1068459
- Alberich (n.d.). *The Listener Crossword. Free Crosswords Online*. [Website] Available online at: <http://www.alberichcrosswords.com/pages/listener.html>
- Amir, O., Biederman, I., Wang, Z., and Xu, X. (2015). Ha Ha! Versus Aha! A direct comparison of humor to nonhumorous insight for determining the neural correlates of mirth. *Cereb. Cortex* 25, 1405–1413. doi: 10.1093/cercor/bht343
- Ansburg, P. I., and Dominowski, R. L. (2000). Promoting insightful problem solving. *J. Creat. Behav.* 34, 30–60. doi: 10.1002/j.2162-6057.2000.tb01201.x
- Anthony, S. (2015). Editorial. *The Magpie (Cryptic Crossword Magazine)*, Vol. 13, eds S. Anthony, M. Goodliffe, and S. Shabankareh (Lewes). Available online at: www.piemag.com
- Anthony and Goodliffe vlog. (n.d.). *Cracking the Cryptic*. Available online at: https://www.youtube.com/channel/UCC-UOdK8-mIjxBQm_ot1T-Q
- Ash, I. K., Cushen, P. J., and Wiley, J. (2009). Obstacles in investigating the role of restructuring in insightful problem solving. *J. Probl. Solv.* 2:3. doi: 10.7771/1932-6246.1056
- Ash, I. K., Jee, B. D., and Wiley, J. (2012). Investigating insight as sudden learning. *J. Probl. Solv.* 4:2. doi: 10.7771/1932-6246.1123
- Attardo, S., Hempelmann, C. F., and Di Maio, S. (2002). Script oppositions and logical mechanisms: modeling incongruities and their resolutions. *Humor* 15, 3–46. doi: 10.1515/humr.2002.004
- Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W., Franklin, M. S., and Schooler, J. W. (2012). Inspired by distraction: mind wandering facilitates creative incubation. *Psychol. Sci.* 23, 1117–1122. doi: 10.1177/0956797612446024
- Bartolo, A., Benuzzi, F., Nocetti, L., Baraldi, P., and Nichelli, P. (2006). Humor comprehension and appreciation: an fMRI study. *J. Cogn. Neurosci.* 18, 1789–1798. doi: 10.1162/jocn.2006.18.11.1789
- Batchelder, W. H., and Alexander, G. E. (2012). Insight problem solving: a critical examination of the possibility of formal theory. *J. Probl. Solv.* 5, 56–100. doi: 10.7771/1932-6246.1143
- Biddlecombe, P. (2009). *Yet Another Guide to Cryptic Crosswords (YAGCC) - Clue Types*. Available online at: <http://www.biddlecombe.demon.co.uk/yagcc/YAGCC2.html>
- Bowden, E. M., and Jung-Beeman, M. (2003a). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychon. Bull. Rev.* 10, 730–737. doi: 10.3758/BF03196539
- Bowden, E. M., and Jung-Beeman, M. (2003b). Normative data for 144 compound remote associate problems. *Behav. Res. Methods Instrum. Comput.* 35, 634–639. doi: 10.3758/BF03195543
- Bowden, E. M., and Jung-Beeman, M. (2007). Methods for investigating the neural components of insight. *Methods* 42, 87–99. doi: 10.1016/j.jymeth.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
- Bowers, K. S., Regehr, G., Balthazard, C., and Parker, K. (1990). Intuition in the context of discovery. *Cogn. Psychol.* 22, 72–110. doi: 10.1016/0010-0285(90)90004
- Cacioppo, J. T., Petty, R. E., and Kao, C. E. (1984). The efficient assessment of need for cognition. *J. Pers. Assess.* 48, 306–307. doi: 10.1207/s15327752jpa4803_13
- Cai, D. J., Mednick, S. A., Harrison, E. M., Kanady, J. C., and Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10130–10134. doi: 10.1073/pnas.0900271106
- Canestrari, C., and Bianchi, I. (2012). Perception of contrariety in jokes. *Discourse Process.* 49, 539–564. doi: 10.1080/0163853X.2012.710524
- Carson, S. H., Peterson, J. B., and Higgins, D. M. (2003). Decreased latent inhibition is associated with increased creative achievement in high-functioning individuals. *J. Pers. Soc. Psychol.* 85, 499–506. doi: 10.1037/0022-3514.85.3.499
- Carson, S. H. (2010). “Latent inhibition and creativity,” in *Latent inhibition: Data, Theories, and Applications to Schizophrenia*, eds R. E. Lubow and I. Weiner. (New York, NY: Cambridge University Press), 183–198.
- Chalicea (2016). Saw. *The Magpie (Cryptic Crossword Magazine)*, Vol. 14, eds S. Anthony, M. Goodliffe, and S. Shabankareh (Lewes). Available online at: www.piemag.com
- Chambers (2014). *Chambers Dictionary, 13th Edn*. London: Chambers Harrap Publishers Ltd.
- Chan, Y.-C. (2016). Neural correlates of sex/gender differences in humor processing for different joke types. *Front. Psychol.* 7:536. doi: 10.3389/fpsyg.2016.00536
- Chein, J. M., and Weisberg, R. W. (2014). Working memory and insight in verbal problems: analysis of compound remote associates. *Mem. Cogn.* 42, 67–83. doi: 10.3758/s13421-013-0343-4
- Chronicle, E. P., MacGregor, J. N., and Ormerod, T. C. (2004). What makes an insight problem? The roles of heuristics, goal conception, and solution recoding in knowledge-lean problems. *J. Exp. Psychol. Learn. Mem. Cogn.* 30:14. doi: 10.1037/0278-7393.30.1.14
- Chu, Y., and MacGregor, J. N. (2011). Human performance on insight problem solving: a review. *J. Probl. Solv.* 3, 119–150. doi: 10.7771/1932-6246.1094
- Cleary, J. (1996). Misleading contexts: the construction of ambiguity in the cryptic crossword clue. *Edinb. Work. Pap. Appl. Linguist.* 7, 14–29.
- Collingridge, P. (2010). 16th February. *Cryptic Crossword Clues and Christmas Crackers [Blog]*. Peter's Website. Available online at: <http://archive.petercollingridge.co.uk/blog/cryptic-crossword-clues-and-christmas-crackers>
- Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407.
- Connor, A. (2011a). *Cryptic Crosswords for Beginners: Double Definitions. Crossword Blog*. Available online at: <https://www.theguardian.com/crosswords/crossword-blog/2011/nov/17/cryptic-crosswords-for-beginners-double-definitions>
- Connor, A. (2011b). *Cryptic Crosswords for Beginners: Soundalikes. Crossword Blog*. Available online at: <https://www.theguardian.com/crosswords/crossword-blog/2011/dec/08/cryptic-crosswords-for-beginners-soundalikes>
- Connor, A. (2013). *Crossword Blog: Meet the Setter-Philistine*. Available online at: <http://www.theguardian.com/crosswords/crossword-blog/2013/jan/31/crossword-blog-meet-the-setter-philistine>

ACKNOWLEDGMENTS

We are indebted to the editorial team at the Magpie crossword magazine (www.piemag.com) for allowing us to reproduce the crossword puzzles, editorial comments, and solutions in Figures 2, 3 and the related discussion. The survey was made available on the Internet via SurveyMonkey® (www.SurveyMonkey.com, Palo Alto, CA); and we are grateful to all the owners and administrators of the websites who allowed us to advertise for participants, and to those who took part so enthusiastically. There was no grant funding associated with this research.

- Connor, A. (2014). *Two Girls, One on Each Knee* (7). London, UK: Penguin Books.
- Connor, A. (2016). *The Joy of Quiz*. London: Particular Books Penguin, Random House.
- Cranford, E. A., and Moss, J. (2011). "An fMRI study of insight using compound remote associate problems," in *Paper Presented at the Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (Austin, TX), 3558–3563.
- Cranford, E. A., and Moss, J. (2012). Is insight always the same? a protocol analysis of insight in compound remote associate problems. *J. Probl. Solv.* 4, 128–153. doi: 10.7771/1932-6246.1129
- 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
- Cunningham, J. B., MacGregor, J. N., Gibb, J., and Haar, J. (2009). Categories of insight and their correlates: an exploration of relationships among classic-type insight problems, rebus puzzles, remote associates and esoteric analogies. *J. Creat. Behav.* 43, 262–280. doi: 10.1002/j.2162-6057.2009.tb01318.x
- Danek, A. H., Fraps, T., von Müller, A., Grothe, B., and Öllinger, M. (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., von Müller, A., Grothe, B., and Öllinger, M. (2014a). 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., Fraps, T., Von Müller, A., Grothe, B., and Öllinger, M. (2014b). Working wonders? Investigating insight with magic tricks. *Cognition* 130, 174–185. doi: 10.1016/j.cognition.2013.11.003
- Danek, A. H., Wiley, J., and Öllinger, M. (2016). Solving classical insight problems without Aha! Experience: 9 Dot, 8 Coin, and Matchstick Arithmetic Problems. *J. Probl. Solving* 9, 47–57. doi: 10.7771/1932-6246.1183
- Davidson, J. E. (1995). "The suddenness of insight," in *The Nature of Insight*, eds R. J. Sternberg and J. E. Davidson (Cambridge, MA: MIT Press), 125–156.
- Davidson, J. E. (2003). "Insights about insightful problem solving," in *The Psychology of Problem Solving*, eds J. E. Davidson and R. J. Sternberg (Cambridge, MA: University Press), 149–175.
- DeYoung, C. G., Flanders, J. L., and Peterson, J. B. (2008). Cognitive abilities involved in insight problem solving: an individual differences model. *Creat. Res. J.* 20, 278–290. doi: 10.1080/10400410802278719
- Dienhart, J. M. (1999). A linguistic look at riddles. *J. Pragmat.* 31, 95–125.
- Dominowski, R. L., and Buyer, L. S. (2000). Retention of problem solutions: the re-solution effect. *Am. J. Psychol.* 113, 249–274. doi: 10.2307/1423730
- Dominowski, R. L. (1969). The effect of pronunciation practice on anagram difficulty. *Psychon. Sci.* 16, 99–100.
- Dominowski, R. L. (1995). "Productive problem solving," in *The Creative Cognition Approach*, eds S. M. Smith, T. B. Ward, and R. A. Finke (Cambridge, MA: MIT Press), 73–95.
- Dunbar, R. I. M., Baron, R., Frangou, A., Pearce, E., van Leeuwen, E. J. C., Stow, J., et al. (2011). Social laughter is correlated with an elevated pain threshold. *Proc. Biol. Sci.* 279, 1161–1167. doi: 10.1098/rspb.2011.1373
- Dunbar, R. I. M., Launay, J., and Curry, O. (2016). The complexity of jokes is limited by cognitive constraints on mentalizing. *Hum. Nat.* 27, 130–140. doi: 10.1007/s12110-015-9251-6
- Editorial Notes (2013). *Solutions for Issue 130, October 2013. The Magpie (Cryptic Crossword Magazine)*, Vol. 11, eds Anthony, M. Goodliffe, and S. Shabankareh (Lewes). Available online at: www.piemag.com
- Editorial Notes (2015). *Solutions for Issue 151, July 2015. The Magpie (Cryptic Crossword Magazine)*, Vol. 13, eds S. Anthony, M. Goodliffe, and S. Shabankareh (Lewes). Available online at: www.piemag.com
- Ellis, J. J., and Reingold, E. M. (2014). The Einstellung effect in anagram problem solving: evidence from eye movements. *Front. Psychol.* 5:679. doi: 10.3389/fpsyg.2014.00679
- Ellis, J. J., Glaholt, M. G., and Reingold, E. M. (2011). Eye movements reveal solution knowledge prior to insight. *Conscious. Cogn.* 20, 768–776. doi: 10.1016/j.concog.2010.12.007
- Erdelyi, M. H. (2014). The interpretation of dreams, and of jokes. *Rev. Gen. Psychol.* 18, 115–126. doi: 10.1037/gpr0000002
- Evans, J. S., and Stanovich, K. E. (2013). Dual-process theories of higher cognition: advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241. doi: 10.1177/1745691612460685
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Evanston, IL: Row Peterson.
- Forabosco, G. (2008). Is the concept of incongruity still a useful construct for the advancement of humor research? *Lodz Papers Pragmatics* 4, 45–62. doi: 10.2478/v10016-008-0003-5
- Friedlander, K. J., and Fine, P. A. (2009). "Expertise in cryptic crossword performance: an exploratory survey," in *Proceedings of the International Symposium of Performance Science, Auckland*, eds A. Williamon, S. Pretty, and R. Buck (Utrecht: European Association of Conservatoires (AEC)), 279–284.
- Friedlander, K. J., and Fine, P. A. (2016). The grounded expertise components approach in the novel area of cryptic crossword solving. *Front. Psychol.* 7:567. doi: 10.3389/fpsyg.2016.00567
- Furnham, A. (1994). A content, correlational and factor analytic study of four tolerance of ambiguity questionnaires. *Pers. Individ. Diff.* 16, 403–410. doi: 10.1016/0191-8869(94)90066-3
- Furnham, A., and Thorne, J. D. (2013). Need for cognition. *J. Individ. Diff.* 34, 230–240. doi: 10.1027/1614-0001/a000119
- 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: Bradford Books/MIT Press), 197–228.
- Gilhooly, K. J., Ball, L. J., and Macchi, L. (2015). Insight and creative thinking processes: routine and special. *Think. Reason.* 21, 1–4. doi: 10.1080/13546783.2014.966758
- Gilhooly, K. J. (2016). Incubation and intuition in creative problem solving. *Front. Psychol.* 7:1076. doi: 10.3389/fpsyg.2016.01076
- Gruszka, A., and Necka, E. (2002). Priming and acceptance of close and remote associations by creative and less creative people. *Creat. Res. J.* 14, 193–205. doi: 10.1207/S15326934CRJ1402_6
- Gupta, N., Jang, Y., Mednick, S. C., and Huber, D. E. (2012). The road not taken: creative solutions require avoidance of high-frequency responses. *Psychol. Sci.* 23, 288–294. doi: 10.1177/0956797611429710
- Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *J. Pers. Soc. Psychol.* 91, 436–455. doi: 10.1037/0022-3514.91.3.436
- 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].
- Hurley, M. M., Dennett, D. C., and Adams, R. B. (2011). *Inside Jokes: Using Humor to Reverse-Engineer the Mind*. Cambridge, MA: MIT press.
- Hélie, S., and Sun, R. (2010). Incubation, insight, and creative problem solving: a unified theory and a connectionist model. *Psychol. Rev.* 117, 994–1024. doi: 10.1037/a0019532
- Ifor. (2013). "The Title is Wrong. *The Magpie (Cryptic Crossword Magazine)*, Vol. 11, eds S. Anthony, M. Goodliffe, and S. Shabankareh (Lewes). Available online at: www.piemag.com
- Jarosoz, A. F., Colflesh, G. J. H., and Wiley, J. (2012). Uncorking the muse: alcohol intoxication facilitates creative problem solving. *Conscious. Cogn.* 21, 487–493. doi: 10.1016/j.concog.2012.01.002
- Johnstone, A. H. (2001). Can problem solving be taught? *Univ. Chem. Educ.* 5, 69–73.
- Jones, G. (2003). Testing two cognitive theories of insight. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1017–1027. doi: 10.1037/0278-7393.29.5.1017
- Kenett, Y. N., Anaki, D., and Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Front. Hum. Neurosci.* 8:407. doi: 10.3389/fnhum.2014.00407
- 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
- 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, 1534–1555. doi: 10.1037/0278-7393.25.6.1534
- Koestler, A. (1964). *The Act of Creation*. London: Hutchinson.
- Kolodner, J. L., and Wills, L. M. (1996). Powers of observation in creative design. *Design Stud.* 17, 385–416. doi: 10.1016/S0142-694X(96)00021-X
- Kounios, J., and Beeman, M. (2014). The cognitive neuroscience of insight. *Annu. Rev. Psychol.* 65, 71–93. doi: 10.1146/annurev-psych-010213-115154
- Kounios, J., and Jung-Beeman, M. (2009). The aha! moment: the cognitive neuroscience of insight. *Curr. Dir. Psychol. Sci.* 18, 210–216. doi: 10.1111/j.1467-8721.2009.01638.x

- Kounios, J., Fleck, J., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M., et al. (2008). The origins of insight in resting-state brain activity. *Neuropsychologia* 46, 281–291. doi: 10.1016/j.neuropsychologia.2007.07.013
- Kozbelt, A., and Nishioka, K. (2010). Humor comprehension, humor production, and insight: an exploratory study. *Humor Int. J. Humor Res.* 23, 375–401. doi: 10.1515/humr.2010.017
- Laughlin, P. R. (1967). Incidental concept formation as a function of creativity and intelligence. *J. Pers. Soc. Psychol.* 5, 115–119.
- Listener Editorial Team (2013). *Listener Crossword: Introduction*. [Website]. Available online at: http://www.listenercrossword.com/List_Intro.html
- MacGregor, J. N., and Cunningham, J. B. (2008). Rebus puzzles as insight problems. *Behav. Res. Methods* 40, 263–268. doi: 10.3758/BRM.40.1.263
- MacGregor, J. N., and Cunningham, J. B. (2009). The effects of number and level of restructuring in insight problem solving. *J. Probl. Solv.* 2, 130–141. doi: 10.7771/1932-6246.1062
- MacGregor, J. N., Ormerod, T. C., and Chronicle, E. P. (2001). Information processing and insight: a process model of performance on the nine-dot and related problems. *J. Exp. Psychol.* 27, 176–201. doi: 10.1037/0278-7393.27.1.176
- Manley, D. (2014). *Chambers Crossword Manual (5th Revised Edn)*. London: Chambers.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* 69, 220–232.
- Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *J. Exp. Psychol. Learn. Mem. Cogn.* 12, 288–294.
- Moorey, T. (2009). *Cryptic Crosswords: Clue Types and Indicators*. [Website]. Available online at: http://www.timmoorey.info/pdfs/clue_types.pdf
- Navon, D. (1988). The seemingly appropriate but virtually inappropriate: notes on characteristics of jokes. *Poetics* 17, 207–219.
- Novick, L. R., and Sherman, S. J. (2008). The effects of superficial and structural information on online problem solving for good versus poor anagram solvers. *Q. J. Exp. Psychol.* 61, 1098–1120. doi: 10.1080/17470210701449936
- Novick, L. R., and Sherman, S. (2003). On the nature of insight solutions: evidence from skill differences in anagram solution. *Q. J. Exp. Psychol. A* 56, 351–382. doi: 10.1080/02724980244000288
- Ohlsson, S., Ernst, A. M., and Rees, E. (1992). The cognitive complexity of learning and doing arithmetic. *J. Res. Math. Edu.* 23, 441–467. doi: 10.2307/749565
- Ohlsson, S. (2011). *Deep Learning: How the Mind Overrides Experience*. Cambridge: University Press.
- Öllinger, M., and Knoblich, G. (2009). “Psychological research on insight problem solving,” in *Recasting Reality: Wolfgang Pauli’s Philosophical Ideas and Contemporary Science*, eds H. Atmanspacher and H. Primas (Berlin: Springer), 275–300.
- Öllinger, M., Jones, G., and Knoblich, G. (2008). Investigating the effect of mental set on insight problem solving. *Exp. Psychol.* 55, 269–282. doi: 10.1027/1618-3169.55.4.269
- Öllinger, M., Jones, G., and Knoblich, G. (2014). The dynamics of search, impasse, and representational change provide a coherent explanation of difficulty in the nine-dot problem. *Psychol. Res.* 78, 266–275. doi: 10.1007/s00426-013-0494-8
- Olteteanu, A.-M., and Falomir, Z. (2015). comRAT-C: A computational compound remote associates test solver based on language data and its comparison to human performance. *Pattern Recognit. Lett.* 67, 81–90. doi: 10.1016/j.patrec.2015.05.015
- Ormerod, T. C., MacGregor, J. N., and Chronicle, E. P. (2002). Dynamics and constraints in insight problem solving. *J. Exp. Psychol.* 28, 791–799. doi: 10.1037/0278-7393.28.4.791
- Ovington, L. A., Saliba, A. J., and Goldring, J. (2016). Dispositional insight scale: development and validation of a tool that measures propensity toward insight in problem solving. *Creat. Res. J.* 28, 342–347. doi: 10.1080/10400419.2016.1195641
- Ovington, L. A., Saliba, A. J., Moran, C. C., Goldring, J., and MacDonald, J. B. (2018). Do people really have insights in the shower? The when, where and who of the Aha! moment. *J. Creat. Behav.* 52, 21–34. doi: 10.1002/jocb.126
- Patrick, J., Ahmed, A., Smy, V., Seeby, H., and Sambrooks, K. (2015). A cognitive procedure for representation change in verbal insight problems. *J. Exp. Psychol.* 41, 746–759. doi: 10.1037/xlm0000045
- Pham, T. (2016). “The register of English crossword puzzles: studies in intertextuality,” in *Variational Text Linguistics: Revisiting Register in English*, eds C. Schubert and C. Sanchez-Stockhammer (Berlin; Boston: Walter de Gruyter GmbH), 111–137.
- Pretz, J. E., Naples, A. J., and Sternberg, R. J. (2003). “Recognizing, defining and representing problems,” in *The Psychology of Problem Solving*, eds J. E. Davidson and R. J. Sternberg (New York, NY: Cambridge University Press), 3–30.
- Ramachandran, V. S. (1998). The neurology and evolution of humor, laughter, and smiling: the false alarm theory. *Med. Hypotheses* 51, 351–354. doi: 10.1016/S0306-9877(98)90061-5
- Robertson, S. I. (2001). *Problem Solving*. Hove: Psychology Press.
- Russ, R. W., and Dillon, J. A. (2011). “Associative theory,” in *Encyclopedia of Creativity, 2nd Edn. Vol. 1*, eds M. A. Runco and S. R. Pritzker (Boston, MA: Academic Press), 66–71.
- Salvi, C., Bricolo, E., Franconeri, S. L., Kounios, J., and Beeman, M. (2015). Sudden insight is associated with shutting out visual inputs. *Psychon. Bull. Rev.* 22, 1814–1819. doi: 10.3758/s13423-015-0845
- Salvi, C., Bricolo, E., Kounios, J., Bowden, E., and Beeman, M. (2016a). Insight solutions are correct more often than analytic solutions. *Think. Reason.* 22, 443–460. doi: 10.1080/13546783.2016.1141798
- Salvi, C., Costantini, G., Bricolo, E., Perugini, M., and Beeman, M. (2016b). Validation of Italian rebus puzzles and compound remote associate problems. *Behav. Res. Methods* 48, 664–685. doi: 10.3758/s13428-015-0597-9
- Schulman, A. (1996). “The art of the puzzler,” in *Cognitive Ecology: Handbook of Perception and Cognition 2nd Edn Vol. 16*, eds M. P. Friedman and E. C. Carterette (San Diego, CA: Academic Press), 293–321.
- Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., and Yaniv, I. (1995). “Demystification of cognitive insight: opportunistic assimilation and the prepared-mind hypothesis,” in *The Nature of Insight*, eds R. J. Sternberg and J. E. Davidson (Cambridge, MA: MIT Press), 65–124.
- Shen, W., Yuan, Y., Liu, C., and Luo, J. (2015). In search of the ‘Aha!’ experience: elucidating the emotionality of insight problem-solving. *Br. J. Psychol.* 107, 281–298. doi: 10.1111/bjop.12142
- Shen, W., Yuan, Y., Liu, C., Zhang, X., Luo, J., and Gong, Z. (2016). Is creative insight task-specific? A coordinate-based meta-analysis of neuroimaging studies on insightful problem solving. *Int. J. Psychol.* 110, 81–90. doi: 10.1016/j.ijpsycho.2016.10.001
- Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives. *Psychol. Bull.* 129, 475–494. doi: 10.1037/0033-2909.129.4.475
- Simon, H. A. (1973). The structure of ill structured problems. *Artif. Intell.* 4, 181–201.
- Sio, U. N., and Ormerod, T. C. (2009). Does incubation enhance problem solving? a meta-analytic review. *Psychol. Bull.* 135, 94–120. doi: 10.1037/a0014212
- Sio, U. N., and Ormerod, T. C. (2015). Incubation and cueing effects in problem-solving: set aside the difficult problems but focus on the easy ones. *Think. Reason.* 21, 113–129. doi: 10.1080/13546783.2014.886626
- Smith, S. M., and Blankenship, S. E. (1989). Incubation effects. *Bull. Psychon. Soc.* 27, 311–314. doi: 10.3758/bf03334612
- Smith, S. M., and Blankenship, S. E. (1991). Incubation and the persistence of fixation in problem solving. *Am. J. Psychol.* 104, 61–87. doi: 10.2307/1422851
- Smith, K. A., Huber, D. E., and Vul, E. (2013). Multiply-constrained semantic search in the remote associates test. *Cognition* 128, 64–75. doi: 10.1016/j.cognition.2013.03.001
- Smith, S. M. (2003). “The constraining effects of initial ideas,” in *Group Creativity: Innovation Through Collaboration*, eds P. B. Paulus and B. A. Nijstad (New York, NY: Oxford University Press Inc), 15–31.
- Smith, S. M., Sifonis, C. M., and Angello, G. (2012). Clue insensitivity in remote associates test problem solving. *J. Probl. Solving.* 4, 28–49. doi: 10.7771/1932-6246.1124
- Sowden, P. T., Pringle, A., and Gabora, L. (2015). The shifting sands of creative thinking: connections to dual-process theory. *Think. Reason.* 21, 40–60. doi: 10.1080/13546783.2014.885464
- Subramaniam, K., Kounios, J., Parrish, T. B., and Jung-Beeman, M. (2009). A brain mechanism for facilitation of insight by positive affect. *J. Cogn. Neurosci.* 21, 415–432. doi: 10.1162/jocn.2009.21057
- Suls, J. M. (1972). *A Two-Stage Model for the Appreciation of Jokes and Cartoons: An Information-Processing Analysis*. New York, NY; London: Academic Press.

- Topolinski, S., and Reber, R. (2010a). Gaining Insight Into the “Aha” experience. *Curr. Dir. Psychol. Sci.* 19, 402–405. doi: 10.1177/0963721410388803
- Topolinski, S., and Reber, R. (2010b). Immediate truth–Temporal contiguity between a cognitive problem and its solution determines experienced veracity of the solution. *Cognition* 114, 117–122. doi: 10.1016/j.cognition.2009.09.009
- Topolinski, S., Bakhtiari, G., and Erle, T. M. (2016). Can I cut the Gordian tkok? The impact of pronounceability, actual solvability, and length on intuitive problem assessments of anagrams. *Cognition* 146, 439–452. doi: 10.1016/j.cognition.2015.10.019
- Underwood, G., MacKeith, J., and Everatt, J. (1988). “Individual differences in reading skill and lexical memory: the case of the crossword puzzle expert,” in *Practical Aspects of Memory: Current Research and Issues Vol. 2, Clinical and Educational Implications*, eds M. M. Gruneberg, P. E. Morris and R. N. Sykes (Chichester: Wiley), 301–308.
- Upadhyay, S. (2008a). *Decoding Double Definitions [Blog]. Crossword Unclued*. Available online at: <http://www.crosswordunclued.com/2008/10/decoding-double-definitions.html>
- Upadhyay, S. (2008b). *How To Spot Anagrams [Blog]. Crossword Unclued*. Available online at: <http://www.crosswordunclued.com/2008/08/how-to-spot-anagram.html>
- Upadhyay, S. (2015). *Playfair [Blog]. Crossword Unclued*. Available online at: <http://www.crosswordunclued.com/2015/07/playfair.html>
- Von Stumm, S., and Ackerman, P. L. (2013). Investment and intellect: a review and meta-analysis. *Psychol. Bull.* 139, 841–869. doi: 10.1037/a0030746
- Wallas, G. (1926). *The Art of Thought*. New York, NY: Harcourt Brace.
- 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
- Webster, D. M., and Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *J. Pers. Soc. Psychol.* 67, 1049–1062.
- Weisberg, R. W. (1995). “Prolegomena to theories of insight: a taxonomy of problems,” *The Nature of Insight*, eds R. J. Sternberg and J. E. Davidson (Cambridge, MA: MIT Press), 157–196.
- Weisberg, R. W. (2006). *Creativity: Understanding Innovation in Problem Solving, Science, Invention, and the Arts*. Hoboken, NJ: John Wiley and Sons.
- Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Think. Reason.* 21, 5–39. doi: 10.1080/13546783.2014.886625
- Wikipedia (2017a). *The New York Times Crossword Puzzle [Website]*. Available online at: https://en.wikipedia.org/wiki/The_New_York_Times_crossword_puzzle (Accessed June 22, 2017).
- Wikipedia (2017b). *Roy Walker (comedian) [Website]*. Available online at: https://en.wikipedia.org/wiki/Roy_Walker_comedian (Accessed May 30, 2017).
- Wiley, J. (1998). Expertise as mental set: the effects of domain knowledge in creative problem solving. *Mem. Cognit.* 26, 716–730.
- Worthen, B. R., and Clark, P. M. (1971). Toward an improved measure of remote associational ability. *J. Educ. Meas.* 8, 113–123.
- Yim, J. (2016). Therapeutic benefits of laughter in mental health: a theoretical review. *Tohoku J. Exp. Med.* 239, 243–249. doi: 10.1620/tjem.239.243
- Zenasni, F., Besançon, M., and Lubart, T. (2008). Creativity and tolerance of ambiguity: an empirical study. *J. Creat. Behav.* 42, 61–73. doi: 10.1002/j.2162-6057.2008.tb01080.x
- Zhao, Q., Zhou, Z., Xu, H., Chen, S., Xu, F., Fan, W., et al. (2013). Dynamic neural network of insight: A functional magnetic resonance imaging study on solving Chinese ‘Chengyu’ riddles. *PLoS ONE* 8:e59351. doi: 10.1371/journal.pone.0059351

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 Friedlander and Fine. 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.



Connect 4: A Novel Paradigm to Elicit Positive and Negative Insight and Search Problem Solving

Gillian Hill^{1*} and Shelly M. Kemp²

¹ Department of Psychology, University of Buckingham, Buckingham, United Kingdom, ² Learning and Teaching Institute, University of Chester, Chester, United Kingdom

OPEN ACCESS

Edited by:

Darya L. Zabelina,
University of Arkansas, United States

Reviewed by:

Hilde Haider,
Universität zu Köln, Germany
Laura Elizabeth Thomas,
North Dakota State University,
United States

*Correspondence:

Gillian Hill
gillian.hill@buckingham.ac.uk

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 28 April 2018

Accepted: 30 August 2018

Published: 25 October 2018

Citation:

Hill G and Kemp SM (2018) Connect
4: A Novel Paradigm to Elicit Positive
and Negative Insight and Search
Problem Solving.
Front. Psychol. 9:1755.
doi: 10.3389/fpsyg.2018.01755

Researchers have typically defined insight as a sudden new idea or understanding accompanied by an emotional feeling of Aha. Recently, examples of negative insight in everyday creative problem solving have been identified. These are seen as sudden and sickening moments of realization experienced as an Uh-oh rather than Aha. However, such experiences have yet to be explored from an experimental perspective. One barrier to doing so is that methods to elicit insight in the laboratory are constrained to positive insight. This study therefore aimed to develop a novel methodology that elicits both positive and negative insight solving, and additionally provides the contrasting experiences of analytic search solving in the same controlled conditions. The game of Connect 4 was identified as having the potential to produce these experiences, with each move representing a solving episode (where best to place the counter). Eighty participants played six games of Connect 4 against a computer and reported each move as being a product of positive search, positive insight, negative search or negative insight. Phenomenological ratings were then collected to provide validation of the experiences elicited. The results demonstrated that playing Connect 4 saw reporting of insight and search experiences that were both positive and negative, with the majority of participants using all four solving types. Phenomenological ratings suggest that these reported experiences were comparable to those elicited by existing laboratory methods focused on positive insight. This establishes the potential for Connect 4 to be used in future problem solving research as a reliable elicitation tool of insight and search experiences for both positive and negative solving. Furthermore, Connect 4 may be seen to offer more true to life solving experiences than other paradigms where a series of problems are solved working toward an overall superordinate goal rather than the presentation of stand-alone and un-related problems. Future work will need to look to develop versions of Connect 4 with greater control in order to fully utilize this methodology for creative problem solving research in experimental psychology and neuroscience contexts.

Keywords: creative problem solving, negative insight, Aha, Uh-oh, Connect 4

INTRODUCTION

An insight moment is defined as a sudden new understanding, idea or solution accompanied by an emotional Aha experience (Jung-Beeman et al., 2008; Klein and Jarosz, 2011). Insight has long been recognized as a desirable feature of creative problem solving, with many famous examples of discoveries in STEM (Science, Technology, Engineering, and Mathematics) being

attributed to it. Maryam Mirzakhani, winner of the Field's medal demonstrates this when asked about mathematics, "the most rewarding part is the 'Aha' moment, the excitement of discovery and enjoyment of understanding something new, the feeling of being on top of a hill, and having a clear view" (CMI, 2008, p. 12). A similar rewarding aspect to insight moments has recently been demonstrated by Friedlander and Fine (2016) whose Cryptic Crossword solving sample identified the Penny Dropping Moment (the Crossword solver community's term for insight moments) as the main motivation for pursuing their hobby. In both these examples the insight experience is a positive one, something that can be seen as a tacit assumption in the historical approach to insight research (Gick and Lockhart, 1995). More recently, however, it has been proposed that insight moments might incorporate negative realizations, with an accompanying Uh-oh moment rather than the prototypical Aha (Hill and Kemp, 2016; Hill and Kemp, unpublished a). This presents a problem for current methods that elicit insight for empirical exploration, which are only designed to produce positive solving experiences. Therefore the development of new methods that stimulate a full range of solving experiences is required to reflect and experimentally test these recent developments in the insight and creativity literature. As such this article describes a preliminary exploration of a new method to elicit experiences that incorporate both positive and negative insight and search solving.

Contemporary research has begun to take a renewed interest in the phenomenology of insight with a varying focus on emotional experiences (Danek et al., 2014a; Jarman, 2014). Danek et al.'s (2014a) participants attempted to solve the puzzle of how a magician had performed different tricks and demonstrated that the resulting solutions arose through both insight and search strategies. In a novel step, after they had completed all the trials participants reported their experiences whilst solving the tricks through insight using a visual analog scale (VAS) to rate against various components. The components of these scales were identified by the researchers and verified through qualitative, open solving descriptions from the participants given before they offered the ratings. Ratings were made for the level of impasse participants experienced before their Aha moment; how pleasant, sudden and surprising solutions were and; how certain they were of the insight solutions they found. Pleasantness was the highest rated feature, with impasse being interpreted as least indicative of Aha solutions. However, as recognized by Danek et al. (2014a), no ratings were recorded for search solutions meaning it was not clear if the phenomenological features identified were unique to insight solving and separable from more general responses to solving problems.

Webb et al. (2016) used the phenomenological rating scales developed by Danek et al. (2014a) across a variety of established tasks that elicit insight problem solving experience. Rather than use a dichotomous approach to labeling of solving experience (i.e., search or insight), their participants rated their feeling of Aha on a VAS (rating the feeling of Aha). They found that pleasantness was positively correlated with feelings of Aha and this effect was consistent across the different types of problem presented [classic insight, classic non-insight and Compound Remote Associates (CRA)]. Other features showed

less consistency, notably impasse either showed no correlation or a negative correlation. These ratings were made on a trial-by-trial basis offering further support for the scales' validity in capturing phenomenological components of insight. As such these studies provide converging evidence to support the importance of further exploration of the emotional component in insight using phenomenological ratings to do so.

Affective aspects of insight have been discussed historically, despite not being explored experimentally until recently. Gick and Lockhart (1995) raised the possibility that insight experiences may not be universally experienced as pleasant. They identified that some solutions might also be accompanied by chagrin, annoyance at the obviousness of the revelation they had previously missed. Hill and Kemp (2016) further explored the notion of negative aspects of insight in a qualitative study. They recorded reports of everyday, sudden realizations that did not represent the positive Aha experiences attached to solving a problem. Instead they demonstrated that negative insights, experienced as Uh-oh moments served to identify problems rather than resolve them. A notable example of this is described by software entrepreneur and philanthropist Dame Stephanie 'Steve' Shirley when outlining the coding process. She describes how she often identified mistakes in her computer coding as sudden negative insights that occurred early in the morning as she awoke (Al-Khalili, 2015); negative insight served to alert her to problems previously unforeseen that she would then work to solve. This demonstrates a proposed adaptive function of negative insight (Hill and Kemp, unpublished a), where identifying a problem has long been seen as an important element of problem solving (Guilford, 1951; Csikszentmihalyi and Sawyer, 1995; Runco and Chand, 1995).

However, whilst Hill and Kemp's (2016) research demonstrates experience of negative insight in everyday context this was based on qualitative reports, which leave a number of unanswered questions. There has been little exploration of how components of the insight moment that are considered emotional and cognitive are related. Topolinski and Reber (2010) asserted that emotional components are epiphenomena, occurring after the purely cognitive insight event. In such an account the negative flavor of some insights would result from subsequent appraisals, perhaps of disappointment or frustration. However, no experimental evidence has to date been provided to directly support this. Furthermore, examination of emotion literature highlights different theoretical perspectives that challenge the assumption that cognitive events necessarily precede an emotional evaluation. For example, Barrett's (2014) Conceptual Act Theory contends that the separation of mental processes to cognitive and emotional is a false dichotomy arguing that both are outcomes of integrated constructed experience rather than one being a consequence of the other. It positions valenced core affect as central to mental events that are then constructed as cognitive, emotional, or perceptual. By this account an insight moment would occur with intrinsic positive or negative core affect contingent on the insight context [whether the realization was 'good for me' or 'bad for me' (Gross, 2015)]. This study takes a first step to such experimental exploration through the development of a task

that can provide insight moments that are both positive and negative.

The types of task typically used to elicit insight were developed against the definition of insight, which carries the tacit assumption that insight is positive and represents a solving experience (for example see Gilhooly and Murphy, 2005; Cunningham et al., 2009; Salvi et al., 2016; Webb et al., 2016). However, the phenomenological scales developed by Danek et al. (2014a) do include the potential to measure negative insight, as they range from very pleasant (scored 100) to very unpleasant (scored as 0). Yet in their original study, participants' responses on average ranged in the positive half of the scale (well above 50), demonstrating that while the possibility to measure negative experiences is available, current paradigms do not elicit this full range of emotional insight responses. Webb et al.'s (2016) positive correlation suggests that as problems were solved with greater feeling of insight so were they generally rated more positively. However, any exceptions to this association could well be hidden by the overall trend. As such current tasks can be seen to offer limited opportunities to investigate negative insight moments that potentially occur at earlier stages of the problem solving process, for example representing sudden episodes of problem finding rather than solution finding. Therefore the full range of insight from negative to positive has yet to be fully explored through current experimental paradigms.

Current methods offer the opportunity for isolated and convergent solving experiences, with the solving moment signifying the culmination of the trial. For complex real-life problems, solving rarely happens in a single insight or search episode. Fleck and Weisberg (2013) and Weisberg (2014) proposed a model of problem solving to explain a continuum from insight to analysis when finding a solution. Within the stages of this model examples of mini-solving episodes can be seen that move the solver closer to their overall superordinate goal and may offer a model that better maps to real-life solving. In fact the subordinate, mini-solving episodes in this model might be considered as a series of problem solving events leading to an ultimate overall goal. In this context, the potential for negative insight moments can be identified, when a solving attempt fails but new information arises suddenly as a result of the failure. These Uh-oh moments initiate new problem solving efforts, perhaps in a different direction that may move the individual closer to their overall goal.

This illustrates that different levels of focus can be applied when considering problem solving, a point made by Perkins (2001) who identified a structure to break-through ideas common across different scales of problem solving. He outlined examples widening in scale from an individual's idea in the moment (more everyday insight) to 'great' profound realizations resulting from a life's work; for example Darwin's development of the theory of evolution. In the extreme Perkins (2001) even proposed consideration of problem solving on an evolutionary timescale. Such an approach again highlights a disparity between the types of tasks currently used to explore insight problem solving in the lab. and more naturalistic, real-life solving experience. Many current methods present discrete solving episodes that are

unconnected to each other, whilst solving in everyday life often sees related solving episodes moving toward an overall goal.

Table top games can be seen to mimic this, with a series of moves or turns working toward the overall goal of winning the game. Chess has been used by cognitive psychologists to explore problem solving and decision making and incorporates positive and negative experiences as a player builds a winning position and identifies potential negative threats from their opponent (Chase and Simon, 1973; Charness, 1992; Gobet and Simon, 1996; Leone et al., 2017). However, the need to learn the rules of chess and differing levels of player ability could introduce potential confounds when being used to explore problem solving behavior. A similarly dyadic game to chess, but with even simpler rules is Connect 4. Players take turns to drop counters (each player has separate colored counters) into a vertical grid, the standard version being seven positions wide and six counters deep. The counter falls to the lowest position, so the first to be dropped into a column will occupy the lowest row with subsequent counters sitting on top of each other. The winner of the game is the first to get four adjacent counters in a line; this can be horizontally, vertically, or diagonally. In playing the game both search and non-search intuitive strategies (potentially insight) can be employed to select moves (Mańdziuk, 2012). These moves like chess may be positively focused toward building a winning position or responding to a negative realization aimed at preventing an immediate loss. As such, Connect 4 would seem to be a candidate platform to elicit repeated episodes of positive and negative solving (selecting the best move) in the controlled environment of game play. These solutions being arrived at through analytic means or in an experience of insight congruent to those reported in other insight research (for example Bowden and Jung-Beeman, 2003a; Danek et al., 2014a)

Furthermore, Connect 4 with a maximum of 21 moves leading to a full grid and stalemate means that a game takes a much shorter time to play than for chess. Yet it retains the desirable features highlighted by researchers in problem solving and decision making of chess including turn-taking and competition leading to goal-oriented positive moves (solutions) and negative problem finding experiences. This would enable multiple, repeated solving experiences to be recorded within a relatively short participation period. Tasks that produce multiple within-participant comparisons over many trials are important, particularly for experimental approaches that incorporate physiological and neuroimaging data in the study of problem solving (Bowden and Jung-Beeman, 2003b; Shen and Yuan, 2016; Hill and Kemp, unpublished b). Despite this potential, little research has focused on Connect 4. The few papers that do are from the field of Applied Computing exploring algorithms to compute the best moves to win (e.g., Allis, 1988) or to develop a learning-based computer system to play Connect 4 (Mańdziuk, 2012). Therefore, this study in addition to developing a novel methodology to elicit both positive and negative problem solving experiences further aims to explore the potential for development of computer-based Connect 4 paradigms for uses beyond Applied Computing contexts.

The first aspect necessary in developing this novel problem solving task will be to check that the experiences elicited in

participants carrying out the task are those identified as relevant to the research question of interest. So in this case it will be necessary to demonstrate that a full range of solving experiences: positive and negative episodes of both insight and search are consistently reported across a range of participants and trials. As seen in the development of other problem solving paradigms (for example the CRA or magic tricks) participants are given definitions for experiences they are then asked to report having completed the task/problem (for example Jung-Beeman et al., 2004; Danek et al., 2016). A widely adopted definition given to help participants identify (positive) insight is that of Jung-Beeman et al. (2004):

A feeling of insight is a kind of 'Aha!' characterized by suddenness and obviousness. You may not be sure how you came up with the answer but are relatively confident that it is correct without having to mentally check it. It is as though the answer came into mind all at once-when you first thought of the word, you simply knew it was the answer. The feeling does not have to be overwhelming, but should resemble what was just described.

More recently an adapted version of this definition incorporated explicit description the alternative to insight describing analytic search as stepwise experiences, furthermore using the analogy of sudden lightbulb switching on for insight compared to gradual dimming up for search (Danek et al., 2016; Webb et al., 2016; Danek and Wiley, 2017). Yet, these studies only focus on insight as a positive experience, so a definition for this study will need to differentiate between Aha and Uh-oh experiences. However, further extending the already quite wordy definitions of insight may be problematic. Emerging evidence from qualitative work by Hill and Kemp (unpublished a) suggests that participants do not always pay attention to all aspects of the research definition of insight given. Qualitative responses were provided by participants some of which reported Uh-oh experiences that were responses to a surprising, negative external event. They appeared to ignore the given definition requiring their Uh-oh moment to be in relation to a new idea or understanding that is central to an insight moment. Furthermore, recent research has suggested that the Aha experience can be deconstructed into different dimensions and is separable from other aspects of insight solving such as solution generation (Kizilirmak et al., 2016; Danek and Wiley, 2017). For the purpose of verifying that Connect 4 elicits positive and negative experiences of insight and search solving the focus for this study is clearly on the experiential aspects of solving. Therefore the development of concise definitions should look to minimize the inclusion of material that may be distracting or less relevant and focus on the experiential components of insight and search solving.

Danek and Wiley (2017) identified three key aspects important in the experience of insight; pleasure, certainty, and suddenness. In addition they were able to demonstrate that elevated surprise ratings associated with false insight, when the participant experienced an insightful solution that was incorrect. In contrast the experience of relief was indicative of insight

solutions that were correct. In Connect 4 however, each move whilst representing a solving episode, does not have a binary correct/incorrect outcome. As such surprise and relief might be less useful in delineating solving experience in this context. Likewise, a feeling of certainty may also be problematic, as there is not such a concrete outcome to judge the efficacy of a move compared to the binary question of how certain someone is that their proposed solution (for example identified word in the CRA) is correct. Therefore a focus on the remaining aspects of suddenness and pleasure (termed more broadly as emotional valence to incorporate negative experience) will be used to develop working definitions for this paradigm.

This study therefore reports the implementation of a new domain of Connect 4 in problem solving research with the aim of eliciting positive and negative, insight and search experiences reliably in participants. It will further explore the validity of this method by using established scales (feelings of insight and phenomenological ratings) used in research paradigms that focus on positive insight and search solving to measure this experience. In addition, a behavioral measure (move time) will also be compared, as this has been shown to be a distinctive aspect in previous research; with insight moves being faster than search (Kounios et al., 2008; Subramaniam et al., 2009; Danek et al., 2014b; Shen et al., 2015). As such a series of hypotheses are proposed to meet these aims. Firstly, there will be a difference in speed of moves reported for different types of solving; specifically insight moves will be faster than search. Moves labeled as positive insight and positive search will be rated as more pleasant than negative insight and search ones. Insight moves will be rated as more surprising and sudden than search. Finally, there will be no influence of solving type or valence on ratings of move certainty.

MATERIALS AND METHODS

Participants

Eighty participants (54 female) were recruited via advertisement within the University and local community. Participants were all over 18 years old ($M_{\text{age}} = 30.63$ years, $SD_{\text{age}} = 12.64$, range age 18–66 years), with a mixture of native English speakers and those with English as an additional language ($n = 10$). Some participants were repeat participants in a longitudinal study that compared solving performance across different tasks (reported elsewhere). In addition to the data reported here, additional physiological (heart rate and interoceptive heart beat counting task) and psychological measures (emotionality self-reports) were recorded (also reported elsewhere).

Materials

A commercially developed, computer-based version of Connect 4 was used (Connect Four Fun developed by TMSOFT, tmsoft.com, copyright 2008–2016). The game has single and two player options, the former being used in this study. The 'night' theme was selected and used for all participants due to its relatively neutral background. In the multigame setting, the player who starts (human player or computer) is determined by the winner of the previous game, which could potentially introduce confounds,

therefore a single game setting was used meaning the human player (participant) always made the first move. The level of difficulty could be selected on a game by game basis choosing from: easy, medium, hard, pro, and expert. These represented subjective labels for the difficulty of play determined by the algorithms of the game (not available to the researcher). This was not deemed to be problematic as participants were self-identifying the level to play. See below in Discussion for further evaluation of this.

Measures

Feeling of Insight

Jung-Beeman et al. (2004) developed a forced choice response of either insight or non-insight. Participants made these self-report after each problem solving episode (in the original study's case after each CRA puzzle was solved). This study adapted the self-report measure to additionally incorporate valance, creating four solving experiences as shown in **Table 1**. Valence was differentiated in terms of motivations for the move, positive moves focused toward winning and negative moves avoiding losing. To distinguish between insight and search, the emotional descriptors of Aha and Uh-oh were used for insight along with the key idea that these occur suddenly. In contrast, search descriptions focused on gradually working out a move. The descriptions used were consistent with previous descriptions used to explore insight (see Hill and Kemp, 2016). A further option was included in line with Bowden and Jung-Beeman (2007) who enabled participants to choose 'other' to ensure that participants were not forced to choose an experience that was not congruent to them. This option was labeled as neutral/or no reason.

Phenomenological Self-Report Scales

Danek et al.'s (2014a) phenomenological self-report scales were used to measure self-reported ratings of pleasantness, surprise, suddenness, and certainty of the different solution types. As detailed above this measure has been further validated in relation to an established range of insight problems by Webb et al. (2016). Impasse was not measured as participants were unlikely to experience this in the context of Connect 4 (as they would always be able to make a move and not looking for a single correct

answer). Following the methodology of Danek et al. (2014a) these were presented at the end of the study after all games of Connect 4 had been played. Each VAS for phenomenological rating was presented one screen at a time in PsychoPy (Peirce, 2007, 2008) using the default VAS settings that presented the rating line in the center of the screen with labels for either end of the scale (see **Table 2** for the labels for each rating scale) and prompt question above. The position marked on the line by the participant provided a score between 1 and 0. Ratings were presented in a random order in terms of both the different types of solving and rating being given. This method minimized the chance that participants were simply responding in relation to the definitions given (although does not exclude this possibility – see further in Discussion). First, as the reports were presented separately and randomized, participants' attention was directed to the two specific aspects of each rating being requested (the solving type and phenomenological aspect being rated) reducing the likelihood of comparisons between ratings for different solving types. Second, as no numbers were used in the reports participants gave, simply a position on a line this again made it harder for participants to make reports relative to their previous ratings given.

Procedure

As highlighted in section Participants additional data (questionnaires and heart beat counting task) was collected before playing Connect 4, and a second heart beat counting measure was taken directly after playing and before completing the phenomenological ratings, these are reported elsewhere. The game of Connect 4 was introduced to participants both verbally and with written instructions immediately prior to playing. It was described as a game played in pairs who take turns in dropping counters in a grid with the winner being the first to get four in a row. An illustration of a Connect 4 grid with a winning game was provided and the different ways to win [horizontal, vertical, and diagonal (shown on picture) lines of four] were explained by the researcher. In addition the levels of difficulty that the game could

TABLE 1 | Self-reported feeling of insight: descriptions given to participants playing Connect 4.

Solving type	Description	Cue available whilst playing
Positive insight	You suddenly have an idea for your next move or how to win	Aha
Positive search	You work out your next move or how to win	I've worked out a good move
Negative insight	You suddenly see a problem or that you are in danger of losing	Uh-oh
Negative search	You work out a problem or that you are in danger of losing	I've worked out there's a problem

TABLE 2 | Questions asked of participants providing phenomenological ratings for the different solving types and labels for visual analog scale.

Phenomenological rating	Question	Label for extremes of VAS	
		0	1
Pleasantness	Please rate your <i>positive insight</i> experience:	Unpleasant	Pleasant
Surprise	Please rate your <i>negative insight</i> experience:	Not surprising	Surprising
Suddenness	The <i>negative search</i> idea came to me:	Slowly	Quickly
Certainty	I felt about the ideas I had through <i>positive search</i> :	Uncertain	Certain

Italic terms changed according to type of problem participants were rating: positive insight, positive search, negative insight, or negative search.

be played at were outlined. Descriptions were then provided for the different types of solving experience in the context of playing Connect 4 (**Table 1**).

Participants played a practice game set to the 'easy' level before selecting the difficulty level they wished to play their first block of three games. Participants indicated when they had chosen their move by pressing a button on a watch (Heart Rate monitor watch) recording the time of their move decision. Participants then verbally identified their selected move (each column was labeled with a number from one to seven) and their feeling of insight when making the move. They could indicate the four solving experiences identified in **Table 1** or select a neutral/no reason option. Reminders of these were provided whilst they were playing the game. The researcher recorded the experience for each move before making the move indicated, this was to avoid participants having to switch between pressing buttons on the watch and operate the Connect 4 game via the mouse or keyboard. Whilst playing the cursor was visible on the screen, therefore the researcher left the cursor in the position of the last move made (i.e., over the column of the last move) to avoid cuing the participant in any way. The participant was positioned facing the screen with the reminder sheet in front of them. They were seated next to the researcher, so no unintentional cues, such as eye movement could be detected by the participant whilst playing the game. After three games the participant had the opportunity to stay of the same level of difficulty or to change. The last three games were then played following the same protocols. The outcome of each of the six games (win, lose, or draw) was recorded by the researcher.

Statistical Analysis

As this study includes predictions for null hypotheses, for example in relation to certainty ratings, a Bayesian approach was taken to analysis as this enables direct testing of the fit of the data to the null (H_0) compared to alternative hypothesis (H_1) (Jarosz and Wiley, 2014). Therefore Bayesian Repeated Measures Analysis of Variance (Bayes RM-ANOVAs) were conducted using JASP (JASP Team, 2017) to analyze main effects and interactions for solving type (independent grouping variable of insight versus search) and valence (independent grouping variable of positive versus negative) on the dependent variables of solution time and phenomenological ratings (pleasantness, surprise, certainty, and suddenness). As little previous research is available on which to produce informed priors, default priors were used with the null hypothesis assumed to have an effect size of zero while the alternative an effect size that was not zero (Rouder et al., 2009). Bayes factors are ratios that express the likelihood of alternative comparative to null hypothesis (or vice versa), they can be reported in terms of the evidence toward the alternative (BF_{10}) or toward the null (BF_{01}). Bayes factors of 1–3 represent weak or anecdotal evidence, between 3 and 10 as moderate, 10 and 30 as strong, and above 30 as very strong evidence toward the hypothesis indicated (i.e., BF_{10} or BF_{01}) (Jeffreys, 1961; but for slightly different interpretation see Raftery, 1995). These interpretations have been adopted by researchers taking a Bayes approach within the field of experimental problem solving and insight (for overview of Bayesian approaches in the context of

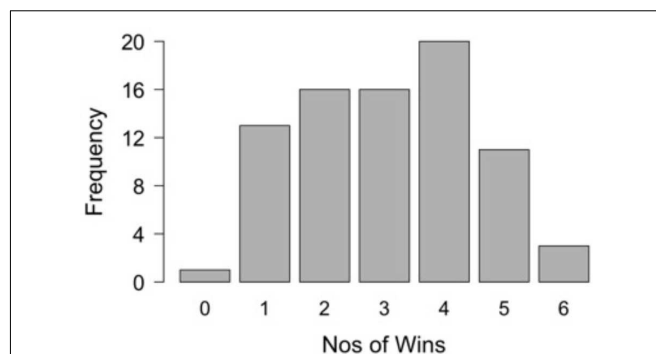


FIGURE 1 | Frequency distribution of participant's wins in Connect 4.

problem solving research see Jarosz and Wiley, 2014 and for an example of application of this analytical approach see Webb et al., 2016).

Ethics Statement

This study was carried out in accordance with the recommendations from the University Science and Medicine Ethics Committee. All participants gave written informed consent in line with the guidelines from the British Psychological Society and in line with the Declaration of Helsinki.

RESULTS

Participants on average won 3.1 ($SD = 1.46$) of the six Connect 4 games they played. **Figure 1** shows the distribution of number of games won that approximates to being normally distributed.

Connect 4 Frequency of Solving Types

Of all moves made, 74% were active solving experiences (search or insight rather than moves identified as neutral/no reason). 22% of these moves were insight (11% positive and 11% negative) and 78% were search (62% positive and 16%). **Table 3** shows the range of solving types reported by participants whilst playing Connect 4. Just under two thirds allocated moves to all four solving types (positive insight, positive search, negative insight, and negative search) whilst over 90% experienced at least three.

TABLE 3 | Breakdown of participants' reported solving as positive insight (+i), positive search (+s), negative insight (−i), and negative search (−s).

Reported	Nos. of participants	+i	+s	−i	−s
4 Solving types	54	✓	✓	✓	✓
3 Solving types	19		✓	✓	✓
	6	✓		✓	✓
	1	✓	✓		✓
	5	✓	✓		✓
	7	✓	✓	✓	
2 Solving types	6		✓		✓
	2	✓	✓		
	2		✓	✓	
1 Solving type	1		✓		

One question of specific interest might be whether all negative insights were reported as a direct response to losing or an imminent loss of a game. Comparing negative insight reporting across all games played showed that roughly equal reporting of negative insight was seen for games that were subsequently won or drawn (41%) compared to lost (59%). Furthermore, only 14% of the total negative insight moves were for the last move in a game that was lost.

Move Times Across Different Types of Solving

For nine participants timing data recorded on the watch was not available due to a recording fault with the equipment they were therefore excluded from analysis exploring move times. The overall mean time for a move across the remaining participants was 11.6 s ($SD = 4.4$ s). A repeated measures Bayesian ANOVA was conducted for participants who reported all four solving types ($n = 45$). Bayes factors (BF) were below three for all main effects of solving type (IV) and valence (IV) on move time (DV) and when comparing a null model incorporating the main effects to the interaction. As such this presents weak evidence of effects of solution type or valence of moves on the time taken to make them.

Phenomenological Self-Reports

For pleasantness ratings a repeated measures Bayesian ANOVA (IVs: Solving type and valence. DV: pleasantness) provided strong evidence of a main effect of valence ($BF_{10} = 5.77e + 38$) and moderate evidence of no main effect of solving type ($BF_{01} = 6.88$). Positive moves were rated as more pleasant than negative for both types of solving. On viewing the graph (Figure 2) presenting these findings it might appear that there was an interaction effect of solving type and valence, with insight moves rated as more positive and more negative than search. However, by adding the main effects to a null model and comparing to one with interaction effects there was seen to be weak evidence toward either model ($BF = 2.35$).

There was strong evidence ($BF_{10} = 266.70$) for a main effect of solving (IV) on surprise ratings (DV), with insight solutions being rated as more surprising than search for both positive and negative moves. There was moderate evidence of no main effect of valence (IV: $BF_{01} = 3.36$) or interaction effects ($BF = 3.71$ toward a null model including main effects compared to interaction effects) on surprise ratings.

For suddenness (DV) there was strong evidence ($BF_{10} = 527.77$) for a main effect of solving (IV), with insight solutions reported as more sudden than search. There was moderate evidence toward a null effect of valence (IV: $BF_{01} = 5.67$) and toward no interaction effects ($BF = 3.57$ toward the null model incorporating main effects).

For certainty ratings (DV) weak evidence was provided for all comparisons (main effects of IVs solving and valence, and interaction of the two: all BF 's < 2), meaning no conclusions could be made regarding evidence toward the null or alternative hypothesis. Graphs with ratings for the four solving types for each phenomenological scale are shown in Figure 2.

DISCUSSION

This study demonstrates that Connect 4 represents a naturalistic task that elicits insight and search problem solving experiences as a player make moves dropping counters into a grid, working toward the overall winning goal of getting four counters in a row. Importantly, it has demonstrated for the first time the elicitation of negative insight in a laboratory setting, meaning that validation of negative insight from an experimental perspective can be undertaken to compliment current research taking a qualitative approach (Hill and Kemp, 2016, unpublished a). The full range of solving was experienced in the majority of participants, with over 90% experiencing at least three of the four solving types. As such the utility of Connect 4 to render multiple incidences of within participant comparisons of different solving is apparent that is particularly important for experimental approaches and those that incorporate neuroimaging and physiological approaches (Bowden and Jung-Beeman, 2003b; Shen and Yuan, 2016; Hill and Kemp, unpublished b). Varying proportions of insight to search are seen for different types of elicitation task. For CRA problems around half of solved trials lead to insight reports (e.g., Jung-Beeman et al., 2004; Cranford and Moss, 2010). Magic tricks conversely gave a higher proportion of non-insight trials, ranging from 41% reported as insight by Danek et al. (2014b) to 29% by Hedne et al. (2016). It can therefore be seen that different methods elicit insight and search solutions to different degrees. Connect 4 in this study showed a lower rate of insight solving than other methods. However, whilst magic tricks and CRA paradigms provided solving experiences in under 60% of the trials, 74% of moves in Connect 4 provided reported solving experience.

Participants' post-game phenomenological reports verified hypothesized characteristics of the experiences elicited whilst playing Connect 4 in line with previous research (Danek et al., 2014a; Webb et al., 2016), finding that positive search and insight were rated as more pleasant than negative search and insight. Furthermore showing that insight (both negative and positive) moves were experienced as more surprising and sudden. Finally, there was not sufficient evidence to support the alternative or null hypothesis exploring certainty ratings across solving and valence. As such this demonstrates that Connect 4 serves as a useful potential method to explore aspects experimentally across the full range of positive and negative insight and search solving as it performs in line with a range other insight elicitation methods that are limited to eliciting positive solving experiences.

As discussed in the Introduction, Danek et al. (2014a) identified a limitation relating to their phenomenological ratings as participants did not provide ratings for non-insight, search solutions against which to compare. Subsequent papers, however, have tended to adopt the feeling of Aha or insight measured reported on a VAS (e.g., Webb et al., 2016) again meaning comparisons between phenomenological aspects of solving experienced as insight or search was not conducted. This paper therefore offers additional support, directly testing the predictions seen in previous literature relating to aspects of pleasantness, suddenness, surprise and certainty attached to insight compared to search solving.

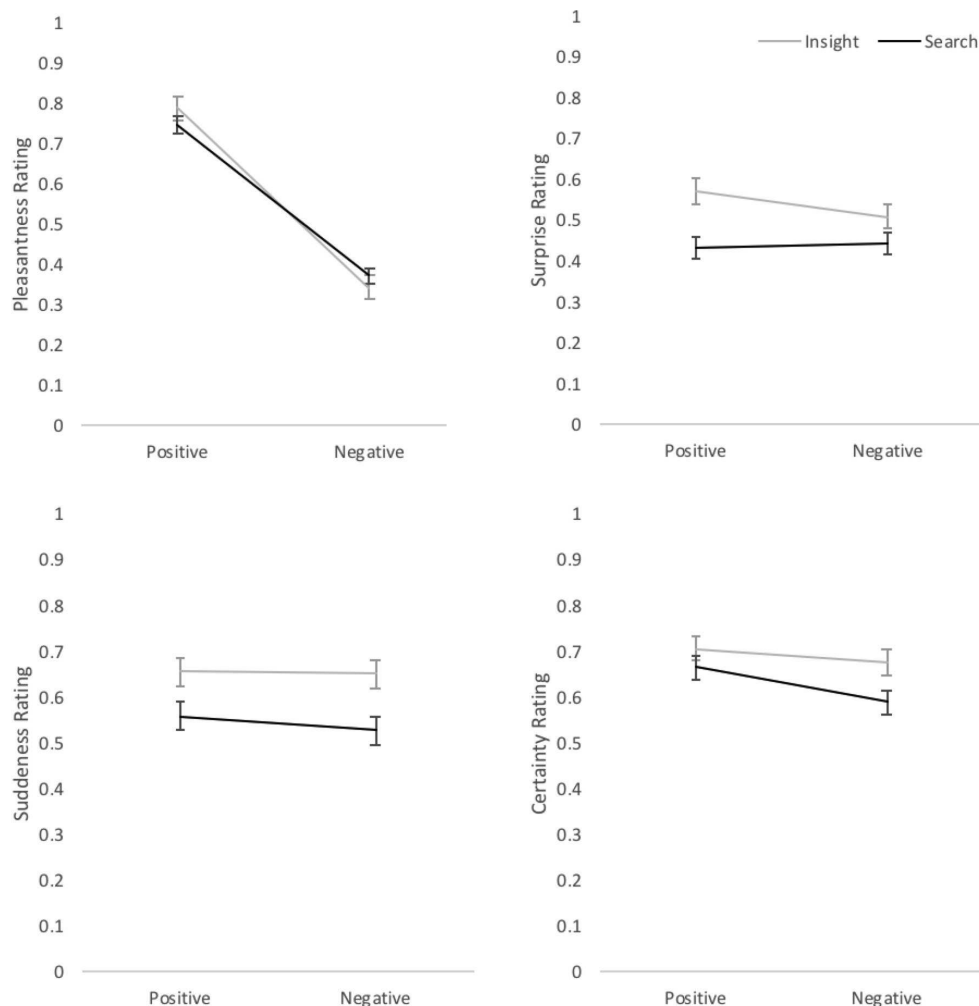


FIGURE 2 | Mean phenomenological ratings for solving type (insight/search) and valence (positive/negative) for Pleasantness, Surprise, Certainty, and Suddenness. Error bars = SE.

In terms of pleasantness, as hypothesized in this study positive insight and search solving were rated as more pleasant than negative solving. However, in previous literature it is suggested that positive emotions of happiness or pleasure were particularly associated with insight moments (Danek et al., 2014a; Shen et al., 2015). Danek et al.'s (2014a) participants, before providing the phenomenological ratings for their insight solutions also gave free reports describing their insight experiences. One of the resulting themes from this related specifically to emotional happiness, this was by far the most reported aspect relating to the insight experience. Shen et al., 2015 showed a direct comparison of happiness ratings [using different rating scales from Danek et al.'s (2014a)] for CRA insight and search solutions, showing that insight trials were rated higher for happiness than search. As such it might be predicted that positive insight would be rated as more pleasant than positive search. As little previous research has considered negative insight it is less easy to make predictions in relation to this. As shown in **Figure 2**, there is a pattern that suggests that positive insight might be seen as more pleasant than

positive search, and negative insight be seen as more unpleasant than negative search solving. However, as highlighted by the accompanying Bayesian analysis, no definitive conclusion for or against this pattern can be reached from the current data. This is therefore something to further explore in future research.

In addition to insight being more pleasant, insight solutions are also proposed to be more sudden. Connect 4 moves labeled as insight were rated as being more sudden than search for both positive and negative solving. Danek et al. (2014a) found suddenness to be less important in insight ratings than factors of pleasantness, surprise, and certainty, but as previously mentioned did not directly compare ratings to those non-insight ratings. Shen et al. (2015) did not have a measure of suddenness but found that participants rated greater hesitation for search trials than insight, so greater hesitation would map to reduced feelings of suddenness, making this finding congruent to the current results. Corroborating behavioral findings to these perceived ratings can be seen from many early CRA studies that show faster responding for trials labeled as insight than search (e.g., Kounios et al., 2008;

Subramaniam et al., 2009; Danek et al., 2014b; Shen et al., 2015 but also see critique of this by Cranford and Moss, 2010, 2011, 2012). One caution to this finding echoes that identified by Danek et al. (2014a) that suddenness formed a key part of the definition given to participants, so their ratings may simply reflect this rather than their experience of insight and search. Indeed, contrary to these self-reports there was insufficient evidence from behavioral measures of Connect 4 move speed (but see limitations below for further evaluation of this measure). Furthermore, Webb et al. (2016) highlighted that it is unclear if suddenness is an aspect of insight that generalizes across problem types. Results here would again suggest further work be necessary to be confident regarding this aspect in relation to insight compared to search in Connect 4 solving.

Previous research in the role of surprise in insight is even less clear. For example, Danek et al. (2014a) and Shen et al. (2015) found conflicting results in respect of surprise, with Shen et al. (2015) not finding that it featured in free responses participants gave in an exploratory study, whilst Danek et al. (2014a) found it was the second most important emotion after happiness. Likewise, Webb et al. (2016) demonstrated that feelings of Aha were more related to surprise than accuracy of the solution. This study again demonstrated congruent results, that insight solving was rated as more surprising than search for both positive and negative solving. Danek and Wiley (2017) suggested that surprise could further distinguish between true and false insight (where solutions were correct or incorrect), with higher surprise ratings for false insight. However, as identified in Connect 4, each move does not result in a dichotomous outcome that is either correct or incorrect, meaning such a relationship would be harder to quantify using the Connect 4 paradigm.

The absence of clear right/wrong outcomes for Connect 4 moves was again reflected in the lack of support from the data in effects for certainty ratings. Future work using the Connect 4 paradigm might consider introducing an objective measure of quality of moves that could be seen as comparable to correct/incorrect in other paradigms (e.g., Danek and Wiley, 2017). In the current study an overall marker of quality might be suggested in examining the number of games won. However, participants were able to self-select the level of difficulty they played at, meaning that the overall win rates of players were not comparable. Asking participants to play at set levels of difficulty would not make sense in terms of the aims of the study which was to elicit within participant solving experiences; if a level was too difficult or easy this would limit the solving that could take place. **Figure 1** demonstrates that participants were indeed selecting a level of play of appropriate challenge, as the approximate normal distribution of winning games with no ceiling or floor effects suggests participants were not playing at a level that was too easy or difficult. Furthermore, it is the within participant efficacy of each move relating to phenomenological experience that is of interest and therefore future research should look to develop such a measure of quality of moves similar to that seen in chess research (Sigman et al., 2010). However, such a measure would require firstly all the moves made to be recorded and compared to the options on the grid at each play point, something that was not possible

using the commercial version of Connect 4 employed in this study.

This highlights a current limitation of this paradigm, which is the need for a better, more fit for purpose version of Connect 4 to be developed. In addition to not being able to measure and quantify move quality the commercial version used ran a game without breaks in play. This meant that data collected whilst playing had to be done verbally requiring the presence of a researcher. Furthermore, the move time data relied on button presses on a watch which incorporated participants' responses to the type of solving, meaning the accuracy of these is questionable. This potentially introduced confounds (although precautions were taken to minimize the experimenter effects – see Method) and for the future complete automation of the task would be desirable. For example, this study took the approach introduced by Danek et al. (2014a) of obtaining phenomenological ratings post task. More recent work has obtained these ratings for each trial of solving (see Webb et al., 2016; Danek and Wiley, 2017), which is preferable as it means the ratings are made close to the actual solving experience, minimizing memory effects and likely confounding influences of definitions on ratings obtained. In order to do so with the current Connect 4 version would require interrupting each move in the game and switching to a different software or computer to collect this data; having a bespoke Connect 4 version would enable such data collection features to be incorporated. Furthermore, heart rate data collection (reported elsewhere) that took place whilst participants played Connect 4 was compromised. There were not long enough breaks between moves to adequately ascribe heart rate effects to individual solving experiences, again adding adequate time breaks between moves is something that could be built in to a bespoke Connect 4 version.

It could be questioned if the negative insights reported in this study are true instances of negative insight or the result of negative appraisals due to losing a game. As reported in section Connect 4 Frequency of Solving Types negative insight was not only reported as a result of losing a game, with a small amount of the overall reported negative insight moves being the final move in a lost game. In fact just under half the reported negative insight moves were in winning games. This would support that participants were reporting moves reflective of their experience of problem solving rather than in response to the outcome of a game (i.e., winning or losing).

A further matter for discussion is whether the methods used in this study (and previous work in the field) simply represent circularity in relation to definitions given to participants producing corresponding phenomenological reports. However, the authors believe that several factors mitigate these concerns. Firstly, participants were not forced to choose one of the four solving types, but had the additional option of neutral/no reason. This means that if the solving descriptions given did not match participants' experience they could indicate as such. Whilst some participants selected the no reason/neutral option for some moves, particularly early in the game (verbally for example many suggested that they always took the same first move) none exclusively selected it. This suggests the solving descriptions did map to genuine experience rather

than representing a demand characteristic of a forced choice. Specifically addressing the possibility of phenomenological ratings representing demand characteristics reflecting definitions given. Firstly steps were taken to reduce this possibility (see section Phenomenological Self-Report Scales) in terms of limiting the comparisons participants could make in the ratings they provided. Furthermore, whilst definitions given did explicitly include descriptions of suddenness, they did not describe things in terms of pleasantness, surprise or certainty. Future research could further look to reduce the possibility of circularity in a number of ways. As highlighted above, a more advanced version of Connect 4 that enabled phenomenological ratings to be taken for each move made (at the time of the move rather than at the end of the study) should improve the quality of these reports. As discussed recently by Laukkonen and Tangen (2018) self-reports made as close to the solving experience as possible reduce the influence of confounds such as those from memory reflecting earlier descriptions of experience given. In addition, the effect of giving definitions on subsequent phenomenological reports in problem solving paradigms could further be explored.

In summary, this study represents a proof of concept for the utility of Connect 4 as a paradigm to elicit problem solving experiences across valence (positive to negative) and solving type (insight to search). This should enable further experimental investigation of problem solving that incorporates the recently

described negative insight, contrasting this to positive insight and search-based solving. Future work is required to develop better computer hosted versions of the game that would enable the incorporation of bespoke features for research designs to: minimize confounding effects such as the presence of an experimenter; enable synchronization with other equipment, for example fMRI or physiological recording and; enable within task data collection for instance as discussed above, phenomenological ratings for each move (trial).

DATA AVAILABILITY

The dataset of the present study will be made available via the Open Science Framework.

AUTHOR CONTRIBUTIONS

All authors designed, analyzed the data, and wrote up this report. GH collected the data.

ACKNOWLEDGMENTS

This research was part of a Ph.D. thesis (Hill, 2017).

REFERENCES

- Allis, V. (1998). *A Knowledge-Based Approach of Connect-Four — the Game is Solved: White Wins*. Master's Thesis, Department of Mathematics and Computer Science, Vrije Universiteit, Amsterdam.
- Al-Khalili, J. (2015). Stephanie Shirley on Computer Coding [Radio Program]. Anna Buckley (Producer) Life Scientific. BBC. Available at: <http://www.bbc.co.uk/programmes/b05pmvl8>
- Barrett, L. F. (2014). The conceptual act theory: a précis. *Emot. Rev.* 6, 292–297. doi: 10.1177/1754073914534479
- 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., and Jung-Beeman, M. (2003a). Aha! insight experience correlates with solution activation in the right hemisphere. *Psychon. Bull. Rev.* 10, 730–737. doi: 10.3758/BF03196539
- Bowden, E. M., and Jung-Beeman, M. (2003b). Normative data for 144 compound remote associate problems. *Behav. Res. Methods Instrum. Comput.* 35, 634–639. doi: 10.3758/BF03195543
- Charness, N. (1992). The impact of chess research on cognitive science. *Psychol. Res.* 54, 4–9. doi: 10.1007/BF01359217
- Chase, W. G., and Simon, H. A. (1973). Perception in chess. *Cognit. Psychol.* 4, 55–81. doi: 10.1016/0010-0285(73)90004-2
- CMI. (2008). *Interview with Research Fellow Maryam Mirzakhani*. Available at: http://www.claymath.org/library/annual_report/ar2008/08Interview.pdf
- Cranford, E., and Moss, J. (2010). “Investigating insight using compound remote associate problems,” in *Proceedings of the Thirty-second Annual Conference of the Cognitive Science Society*, Starkville, MS.
- Cranford, E. A., and Moss, J. (2011). “An fMRI study of insight using compound remote associate problems,” in *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, Starkville, MS.
- Cranford, E. A., and Moss, J. (2012). Is insight always the same? *J. Probl. Solving* 4, 128–153.
- Csikszentmihalyi, M., and Sawyer, K. (1995). “Creative insight: the social dimension of a solitary moment,” in *The Nature of Insight*, eds R. J. Sternberg and J. E. Davidson (Cambridge, MA: MIT Press), 329–363.
- Cunningham, J. B., MacGregor, J. N., Gibb, J., and Haar, J. (2009). Categories of insight and their correlates: an exploration of relationships among classic-type insight problems, rebus puzzles, remote associates and esoteric analogies. *J. Creat. Behav.* 43, 262–280. doi: 10.1002/j.2162-6057.2009.tb01318.x
- Danek, A., 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
- Danek, A. H., Fraps, T., Von müller, A., Grothe, B., and Öllinger, M. (2014a). 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., Fraps, T., Von Muller, A., Grothe, B., and Öllinger, M. (2014b). Working wonders? Investigating insight with magic tricks. *Cognition* 130, 174–185. doi: 10.1016/j.cognition.2013.11.003
- Danek, A. H., Wiley, J., and Öllinger, M. (2016). Solving classical insight problems without Aha! Experience: 9 Dot, 8 coin, and matchstick arithmetic problems. *J. Probl. Solving* 9, 47–57. doi: 10.7771/1932-6246.1183
- Fleck, J. I., and Weisberg, R. W. (2013). Insight versus analysis: evidence for diverse methods in problem solving. *J. Cognit. Psychol.* 25, 436–463. doi: 10.1080/20445911.2013.779248
- Friedlander, K. J., and Fine, P. A. (2016). The grounded expertise components approach in the novel area of cryptic crossword solving. *Front. Psychol.* 7:567. doi: 10.3389/fpsyg.2016.00567
- 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: The MIT Press), 197–228.
- Gilhooly, K. J., and Murphy, P. (2005). Differentiating insight from non-insight problems. *Think. Reason.* 11, 279–302. doi: 10.1080/1354678042000187
- Gobet, F., and Simon, H. A. (1996). Templates in chess memory: a mechanism for recalling several boards. *Cognit. Psychol.* 31, 1–40. doi: 10.1006/cogp.1996.0011
- Gross, J. J. (2015). The extended process model of emotion regulation: elaborations, applications, and future directions. *Psychol. Inquiry* 26, 130–137. doi: 10.1080/1047840X.2015.989751
- Guilford, J. (1951). Creativity. *Am. Psychol.* 5, 444–454. doi: 10.1037/h0063487

- Hedne, M. R., Norman, E., and Metcalfe, J. (2016). Intuitive feelings of warmth and confidence in insight and noninsight problem solving of magic tricks. *Front. Psychol.* 7:1314. doi: 10.3389/fpsyg.2016.01314
- Hill, G. (2017). *Emotion and Insight: The physiological and Psychological Experience of Positive and Negative Insight Through Qualitative Exploration of Everyday Experience and Quantitative Investigation using Problem Solving Tasks*. Ph.D. thesis, University of Buckingham, Buckingham.
- 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/job.142 [Epub ahead of print].
- Jarman, M. S. (2014). Quantifying the qualitative: measuring the insight experience. *Creat. Res. J.* 26, 276–288. doi: 10.1080/10400419.2014.929405
- Jarosz, A., and Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *J. Probl. Solving* 7, 2–9. doi: 10.7771/1932-6246.1167
- JASP Team. (2017). *JASP (Version 0.8.2) [Computer Software]*. Available at: <https://jasp-stats.org/>
- Jeffreys, H. (1961). *Theory of Probability*, 3rd Edn. New York, NY: Oxford University Press.
- Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biol.* 2:500–510. doi: 10.1371/journal.pbio.0020097
- Jung-Beeman, M., Collier, A., and Kounios, J. (2008). How insight happens: learning from the brain. *Neuroleadersh. J.* 1, 20–25.
- Kizilirmak, J. M., Galvao Gomes da Silva, J., Imamoglu, F., and Richardson-Klavehn, A. (2016). Generation and the subjective feeling of “Aha!” are independently related to learning from insight. *Psychol. Res.* 80, 1059–1074. doi: 10.1007/s00426-015-0697-2
- Klein, G., and Jarosz, A. (2011). A naturalistic study of insight. *J. Cognit. Eng. Decis. Mak.* 5, 335–351. doi: 10.1177/1555343411427013
- Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M., et al. (2008). The origins of insight in resting-state brain activity. *Neuropsychologia* 46, 281–291. doi: 10.1016/j.neuropsychologia.2007.07.013
- Laukkonen, R. E., and Tangen, J. M. (2018). How to detect insight moments in problem solving experiments. *Front. Psychol.* 9:282. doi: 10.3389/fpsyg.2018.00282
- Leone, M. J., Slezak, D. F., Golombek, D., and Sigman, M. (2017). Time to decide: diurnal variations on the speed and quality of human decisions. *Cognition* 158, 44–55. doi: 10.1016/j.cognition.2016.10.007
- Mańdziuk, J. (2012). *Human-Like Intuitive Playing in Board Games*. In *International Conference on Neural Information Processing*. Berlin: Springer, 282–289. doi: 10.1007/978-3-642-34481-7_35
- Peirce, J. W. (2007). PsychoPy—psychophysics software in python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017
- Peirce, J. W. (2008). Generating stimuli for neuroscience using psychoPy. *Front. Neuroinform.* 2:10. doi: 10.3389/neuro.11.010.2008
- Perkins, D. N. (2001). *The Eureka Effect: The Art and Logic of Breakthrough Thinking*. New York, NY: WW Norton & Company.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociol. Methodol.* 25, 111–163. doi: 10.2307/271063
- Rouder, J. N., Speckman, 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
- Runco, M. A., and Chand, I. (1995). Cognition and creativity. *Educ. Psychol. Rev.* 7, 243–267. doi: 10.1007/BF02213373
- 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
- Shen, W., and Yuan, Y. (2016). Finding the somatic precursor of spontaneous insight. *Int. J. Psychol.* 51:207.
- Shen, W., Yuan, Y., Liu, C., and Luo, J. (2015). In search of the ‘Aha!’ Experience: elucidating the emotionality of insight problem-solving. *Br. J. Psychol.* 107, 281–298. doi: 10.1111/bjop.12142
- Sigman, M., Etchemendy, P., Fernandez Slezak, D., and Cecchi, G. A. (2010). Response time distributions in rapid chess: a large-scale decision making experiment. *Front. Neurosci.* 4:60. doi: 10.3389/fnins.2010.00060
- Subramaniam, K., Kounios, J., Parrish, T. B., and Jung-Beeman, M. (2009). Brain mechanism for facilitation of insight by positive affect. *J. Cogn. Neurosci.* 21, 415–432. doi: 10.1162/jocn.2009.21057
- Topolinski, S., and Reber, R. (2010). Gaining insight into the “Aha” experience. *Curr. Direct. Psychol. Sci.* 19, 402–405. doi: 10.1177/0963721410388803
- 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
- Weisberg, R. W. (2014). Toward an integrated theory of insight in problem solving. *Think. Reason.* 21, 5–39. doi: 10.1080/13546783.2014.886625
- doi: 10.1080/13546783.2014.886625

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 Hill and Kemp. 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.



Feelings-of-Warmth Increase More Abruptly for Verbal Riddles Solved With in Contrast to Without Aha! Experience

Jasmin M. Kizilirmak^{1*}, Violetta Serger¹, Judith Kehl², Michael Öllinger³, Kristian Folta-Schoofs^{1†} and Alan Richardson-Klavehn^{2†}

¹ Neurodidactics and Neuro Lab, Institute for Psychology, University of Hildesheim, Hildesheim, Germany, ² Memory and Consciousness Research Group, Clinic for Neurology, Otto-von-Guericke University, Magdeburg, Germany,

³ Parmenides Center for the Study of Thinking, Pullach, Germany

OPEN ACCESS

Edited by:

Ian Hocking,
Canterbury Christ Church University,
United Kingdom

Reviewed by:

Gillian Hill,
University of Buckingham,
United Kingdom
Kenneth James Gilhooly,
Brunel University London,
United Kingdom

*Correspondence:

Jasmin M. Kizilirmak
kizilirmak@uni-hildesheim.de

[†] These authors shared senior
authorship

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 28 April 2018

Accepted: 19 July 2018

Published: 13 August 2018

Citation:

Kizilirmak JM, Serger V, Kehl J, Öllinger M, Folta-Schoofs K and Richardson-Klavehn A (2018) Feelings-of-Warmth Increase More Abruptly for Verbal Riddles Solved With in Contrast to Without Aha! Experience. *Front. Psychol.* 9:1404. doi: 10.3389/fpsyg.2018.01404

When we are confronted with a new problem, we typically try to apply strategies that have worked in the past and which usually lead closer to the solution incrementally. However, sometimes, either during a problem-solving attempt that does not seem to lead closer to the solution, or when we have given up on problem-solving for the moment, the solution seems to appear out of nowhere. This is often called a moment of insight. Whereas the cognitive processes of getting closer to the solution are still unknown for insight problem-solving, there are two diverging theories on the subjective feeling of getting closer to the solution: (1) One that states that an intuitive feeling of closeness to the solution increases slowly, but incrementally, before it surpasses the threshold to consciousness and becomes verbalizable (=insight) (continuous approach), and (2) another that proposes that the feeling of closeness to the solution does not increase before it exceeds the threshold to consciousness (discontinuous approach). Here, we investigated the subjective feeling of closeness to the solution, assessed as feeling-of-warmth (FoW), its relationship to solving the problem versus being presented with it and whether a feeling of Aha! was experienced. Additionally, we tested whether Aha! experiences are more likely when the problem is solved actively by the participant or presented to the participant after an unsuccessful problem-solving attempt, and whether the frequency of Aha! experiences correlates with problem difficulty. To our knowledge, this is the first study combining the CRAT with FoW assessments for the named conditions (solved/unsolved, three difficulty levels, Aha!/no Aha!). We used a verbal problem-solving task, the Compound Remote Associates Task (CRAT). Our data revealed that Aha! experiences were more often reported for solutions generated by the participant compared to solutions presented after unsuccessful problem-solving. Moreover, FoW curves showed a steeper increase for the last two FoW ratings when problems were solved with Aha! in contrast to without Aha!. Based on this observation, we provide a preliminary explanation for the underlying cognitive process of solving CRA problems via insight.

Keywords: insight, problem solving, consciousness, feeling-of-warmth, intuition, cognition

INTRODUCTION

Problems can be solved in many different ways, but one gross categorization of simple problems used in research is solving problems stepwise and analytically or by a sudden insight (Metcalf and Wiebe, 1987). Analytical problem-solving refers to a gradual process of applying existing knowledge and available operators to a given problem representation. The best examples are probably mathematical equations for which one already knows the relevant formulas, or problems like the Tower of Hanoi. When prior knowledge fails to solve a problem, it is often necessary to turn away from known problem-solving approaches and invent something new. In such situations, people often get stuck in an impasse: a state of mind where the problem seems unsolvable. The driving force to overcome an impasse is thought to be a representational change, that either changes the given problem representation or the imposed goal representation (Ohlsson, 1992; Kershaw et al., 2013). A representational change is often accompanied by a deep insight into the solution of a novel problem. In our daily lives, such insights often occur when we have already turned our attention elsewhere, after being stuck with our unsuccessful problem-solving attempts for a very frustrating time. One of the earliest characterizations of insight proposes that a gap in the problem representation is detected and the problem solver is able to realize which components of the problem are essential for solving it (*selective encoding*), “synthesizing what might originally seem to be isolated pieces of information into a unified whole” (*selective combination*), and relating novel information to prior knowledge (*selective comparison*) (Davidson and Sternberg, 1984). Being able to realize which components of the problem are actually relevant for the solution is rather difficult for insight problems and is often thought to occur only after a representational change. Usually, those pieces of the problem are picked that seem the most promising based on prior experience (Knoblich et al., 2001). However, for insight problems, those are usually the ones that lead us into an impasse during our problem-solving attempt. A representational change needs to take place—the attentional focus needs to be shifted toward the actually relevant pieces of information which are usually less likely from our experience (Öllinger et al., 2014).

A recent study on representational change and insight assessed the dynamics of the representational change and whether they differ for problems solved with or without insight (Danek et al., 2018). The authors used videos of magic tricks and participants needed to figure out how they worked. Insight was operationalized as experiencing a subjective feeling of Aha! (solution being found suddenly, being confident it is correct). This operationalization has been frequently used since Jung-Beeman and colleagues introduced it (Jung-Beeman et al., 2004). The representational change was assessed by having participants rate the relevance of verbs for performing the tricks. The authors found that the shift from irrelevant to relevant verbs occurred gradually for no Aha! and more sudden for tricks solved with Aha!.

This pattern bears high similarity with the subjective feeling of closeness to the solution (Metcalf and Wiebe, 1987;

Reber et al., 2007; Hedne et al., 2016), sometimes operationalized as feeling-of-warmth (FoW, in the style of the children’s game pot hitting¹). Metcalfe and Wiebe (1987) compared the dynamics of FoW during solving classical insight problems (problems which are thought to lead to an initial impasse during problem-solving), incremental problems (e.g., the Tower of Hanoi), and algebra problems. They found that FoW increased incrementally for non-insight problems and more suddenly for insight problems.

The likeness between the dynamics of the representational change and FoW for insight problems may suggest FoW as an intuitive marker of a representational change in the right direction. Intuition can be defined as the ability to comprehend an idea or being able to judge stimulus characteristics without being consciously aware of the knowledge on which this judgment is based (Ilg et al., 2007). Seeing FoW as an intuitive marker of the representational change would be in line with Bowers’ proposal that there are two stages of intuition: (1) a guiding stage, that is, the implicit perception of coherence of thought (intuition), and (2) an integrative stage during which the problem components form a plausible solution that is available to consciousness (insight) (Bowers et al., 1990, 1995). However, this approach on intuition and insight is in conflict with another approach that regards insight, intuition and analytical/incremental problem-solving as three different processes (Reber et al., 2007). Reber et al. (2007) propose that during analytical problem-solving, subjective and objective closeness to the solution increase equally linearly. In contrast, when a problem is solved by insight, the subjective feeling of closeness is at first level and only increases just before the solution becoming consciously available. How the objective closeness to the solution increases in the case of an insight solution, is not specified. The intuitive problem-solving process differs from the insight process by the objective closeness increasing linearly, while the subjective closeness raises at first linearly but with a flatter slope than for analytical problem-solving, and surges suddenly just before the solution becomes verbalizable. Reber’s model of intuitive problem-solving seems to map Bowers’ idea of intuitive problem-solving attempts that culminate in an insight (Bowers et al., 1990, 1995).

Zander et al. (2016) discussed the two approaches on insight in a review on insight and intuition. They described continuous and discontinuous models for both and conclude that intuition researchers favor the continuous model of intuition. In the continuous model, intuition is based on an early assessment of initial semantic search processes for the solution, culminating in an insight when the solution becomes accessible to conscious thought. In contrast, insight researchers seem to favor a discontinuous model which sees intuitive feelings about the correct solution as a misdirection of the problem-solving attempts that lead into an impasse, from which only restructuring may lead to an insight. Here, we consider FoW as an equivalent of an intuition about the closeness to the solution. On the first

¹This game is traditionally played by hiding a prize underneath a pot. Of a group, one is designated the seeker, blindfolded, and equipped with a wooden spoon. The seeker needs to try and find the pot by hitting the floor (and eventually the pot) while the others call “cold,” “warmer,” “hot,” depending on how close the seeker is to the pot. At least in Germany it is very famous (German name “Topfschlagen”).

glance, the discontinuous model seems congruent with Reber's model curve of insight. However, if intuition were to lead the problem solver astray, FoW should increase before the problem solver gets stuck in an impasse, only to decrease again, when the participant realizes that their intuition was incorrect. This process would probably be repeated several times before reaching a solution, resulting in a zigzag curve of FoW with a sudden final surge at the end². If intuition were to culminate in insight, we would expect only one increase in the feeling of warmth, not an early increase followed by a decrease.

So far, we have only considered problems that are solved. What about problems that are not solved? Could insight also be involved when a solution is not found by the participant? There are very few studies we know of that looked at unsolved or incorrectly solved problems in the context of insight. Kizilirmak and colleagues report that Aha! experience are reported by participants also for unsolved problems for which the solution was presented (Kizilirmak et al., 2016a,c). However, a preceding attempt at problem solving seems important for the Aha! experience to occur, as it showed a higher prevalence for solutions that were presented after an unsuccessful attempt at problem solving (mean frequency = 0.41, $SD = 0.14$) as opposed to solutions that were immediately presented (0.31, $SD = 0.35$) (Kizilirmak et al., 2016c). Danek and Wiley (2017) investigated Aha! experiences for incorrect solutions and found that they were qualitatively different to Aha! experiences for correctly solved problems. That is, surprise was more strongly related to incorrectly solved problems with Aha!, whereas for correctly solved problems with Aha! it was tension relief. However, it is difficult to say whether the Aha! experience could be likened to insight or whether it is necessary for a problem solver to find the correct solution on his own, because there is no common definition of insight used by all insight researchers. Currently, however, most researchers think of solutions to problems that were solved with an Aha! experience as insight solutions, and this is what we will stick to in the present study.

Aims of the Current Study

The current study investigates the dynamics of the subjective perception of closeness to the solution during verbal-problem solving separated by solutions solved either by the participant or presented after an unsuccessful attempt. This classification is detailed by reported Aha! problem difficulty. Until now, FoW dynamics were tested for classical single-trial insight problems (i.e., a set of very different problems) but without considering Aha! (Metcalfe and Wiebe, 1987) and with magic tricks for problems solved with versus without Aha! (Hedne et al., 2016). We would like to add to these findings by showing how the subjective perception of closeness to the solution develops over time for problems solved with Aha! and without Aha! and for solved versus presented solutions. So in line with this research topic's aim of showcasing (a) either novel methods to

research creativity or (b) the application of tried and tested methods in a novel way, the current study represents one of the latter.

We assessed FoW ratings and subjectively reported Aha! experiences while participants tried to solve Compound Remote Associate Task (CRAT) problems of three levels of difficulty. The CRAT is a verbal problem-solving task during which three words are presented that on first glance seem unrelated (e.g., power, shoe, radish). A fourth word needs to be found that can be used to form compound words with each of the other three (horse). The task is thought to be well suited to provoke insight solutions, because close associations with the three problem words often lead to an impasse (e.g., power outage, power rangers, power point,...).

The CRAT was originally developed by Bowden and Jung-Beeman (2003) who based their task on the Remote Associates Task by Mednick (1962) who intended this task as a test of students' creativity. We believe that our study is a good extension of Hedne et al. (2016) in which magic tricks were used. We have shown that generating solutions to insight problems with Aha! are closely related to enhanced long-term memory for the problem and its solution (Kizilirmak et al., 2016a,c). The underlying mechanism is probably driven by reward-related processes. The sudden comprehension of difficult solutions is related to positive feelings such as tension relief (Danek et al., 2014), as well as the novel information (the solution) being easily integrated into prior knowledge (schema-based learning) (Kizilirmak et al., 2016b).

Gaining a better understanding of the dynamics of the subjective perception of closeness to the solution by means of FoW ratings will help us in understanding the cognitive process of insight, under which circumstances it occurs, and whether intuition can be seen as an antecedent of the Aha! experience, at least in the case of the CRAT. So far, this is the first study to use the CRAT for investigating FoW in general and in relation to the subjective feeling of Aha!.

Based on previous findings, we expected roughly equal distributions for generated and non-generated solutions. For FoW dynamics, we expected several potential outcomes: (a) Either a replication of Hedne's and Metcalfe's findings (Metcalfe and Wiebe, 1987; Hedne et al., 2016), that is, an almost level curve for problems solved with Aha! that rises very suddenly just before a solution is reported. Such a curve would also be in line with Reber and colleagues' model curve of insight (Reber et al., 2007). (b) Or a slow rise followed by a much steeper slope just before the solution is reported. This would be in line with Reber's intuition model which we consider as reflecting Bowers' idea that insight is the second stage of intuitive problem-solving. Regarding item difficulty, we expected a higher frequency of Aha! for difficult items. This hypothesis was based on a study of insight reports from real life, which suggests that problems for which Aha! experiences were reported were mostly so difficult that problem solvers got stuck in an impasse for a long time and turned to other

²Unfortunately, to accurately assess and map such a development, we would probably need more continuous FoW assessments, which is why we did not include this model as part of our testable hypotheses.

matters before suddenly realizing the solution (Klein and Jarosz, 2011).

MATERIALS AND METHODS

Participants

Thirty-six healthy young adults (six male) participated in the study after providing written informed consent in accordance with the Declaration of Helsinki (World Medical Association, 2013). The study was approved by the Ethics Committee of the University of Hildesheim, Germany. Participation was voluntary and compensated via course credits. Median age was 20.5 years (range: 18–35 years). All had normal or sufficient uncorrected vision for reading the stimuli with ease, as tested by letting participants read the instructions aloud. Five participants were left-handed, the remaining 31 participants were right-handed. However, as all conditions were assessed within-subjects, and button-assignments were counterbalanced across participants, handedness should have no confounding effect.

Stimulus Material

For each participant, we used 96 German CRAT items of a 144 item selection of our original 180 items used in earlier studies (Kizilirmak et al., 2016b,c). All CRAT items consist of four nouns, three words that make up the problem and one word that is the solution. The words are either nouns or color words. The solution word is one which can be used to form a compound word with each of the other three by appending it either as a prefix or suffix. To enable the investigation of the influence of item difficulty (i.e., the probability of an item to be successfully solved within the time limit), we categorized the items into three levels of difficulty: easy, medium, difficult. This categorization was based on data from a normative data sample ($N = 20$) collected at the Otto-von-Guericke University of Magdeburg, Germany. The 48 items with the lowest solution rate (primary sorting) and highest response time (secondary sorting, e.g. all items with a solution rate of 50 % were further ranked according to response time) were classified as “difficult,” the 48 items with the highest solution rate and lowest response time were classified as “easy,” and 48 items around the median solution rate were classified as “medium.” The remaining 36 items were not used in this study to ensure a more clear-cut difference between the difficulty levels.

The thus selected 144 items were divided into three sets (48 problems each) that were matched for probability to be solved (used to determine problem difficulty), to elicit a subjective Aha! response, and for plausibility according to a normative data sample that used a different set of 20 participants. For the current study, two sets were chosen, which item pools were chosen were counterbalanced across participants according to a reduced Latin square. From the 96 problems, six items (two of each of three levels of item difficulty) were drawn pseudo-randomly for six practice trials presented prior to the experiment proper. The third pool was not used. It should be noted that for each participant, plausibility, solution probability, and Aha! probability was equal, while specific stimulus characteristics like word frequency and emotional valence were counterbalanced

across participants, thereby preventing any confounding effects of those factors.

Design

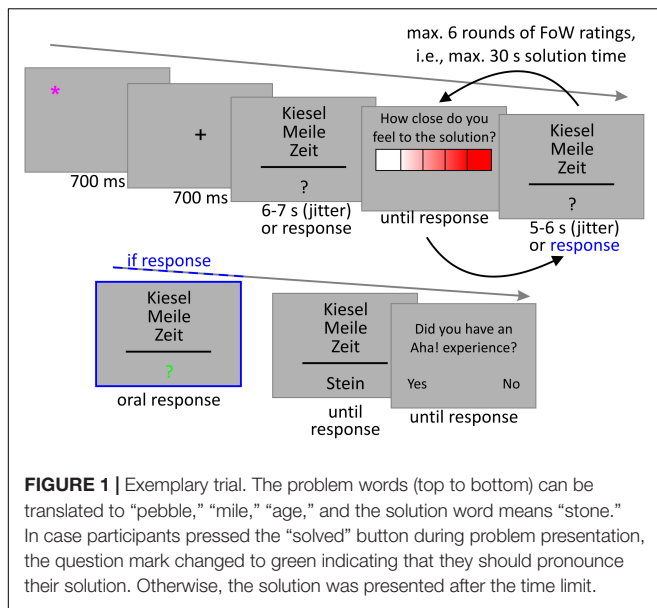
We investigated alleged differences in the course of the subjective feeling of closeness to the solution (operationalized as FoW) depending on (1) whether the solution to a CRAT items was generated or presented after unsuccessful generation (factor = GENERATION), (2) whether the solution was comprehended with or without a feeling of Aha! (AHA), and (3) depending on item difficulty (DIFFICULTY). Participants were asked to assess their subjective closeness to the solution by means of a FoW on a 5-point heat scale (from 0 = white = cold to 4 = red = hot). FoW was assessed for the first time 6–7 s after stimulus onset to provide additional time for initial reading of the words, and every 4.5–5.5 s (pseudo-random jitter) thereafter until either coming up with a solution or reaching an upper time limit of 30 s (time for FoW ratings not counted). The jittered assessment time of FoW was intended to decrease the disturbance of the solution process by anticipated FoW ratings. The occurrence of an Aha! experience was assessed for each item after the solution was found or provided after reaching the upper time limit. Participants were required to decide via button press whether they had an Aha! experience or not.

Task and Procedure

Firstly, participants were provided with oral and written information about the task and procedures as well as a consent form. After providing their written consent, they were asked to describe the task in their own words. This was done to check whether everything was understood as intended and to provide further instructions if necessary.

The main experimental task was conducted in a silent room with dimmed light inside a 1.3 deep, 4.0 m long, 2.0 m high box. The box serves as a shield against visual and partly auditory distractions. Participants were placed in a chair that was adjusted according to their height so that they could comfortably place their chin on a chin rest. The chin rest was placed exactly 1.0 m in front of a flat computer screen. The chin rest was part of a stationary 1250 Hz iView X eye-tracker (SensoMotoric Instruments, Teltow, Germany) with which we recorded additional gaze direction data which are, however, not part of the current report.

Stimulus presentation and behavioral data collection was controlled via the software Presentation, version 20 (Neurobehavioral Systems, Inc., Berkeley, CA, United States). The task began with 6 practice trials, followed by a break and the chance to ask questions. The practice trials did not differ from the main trials. The 90 main trials were presented in three blocks a 30 trials. Before each block started, a 9-point (3×3 matrix, 800×800 pixels) calibration field for the eye-tracker was presented and participants were required to fixate on each point in turn as orally instructed while the experimenter calibrated the eye-tracker. During the breaks between blocks, participants were allowed to pace around. As depicted in **Figure 1** (exemplary trial), the background was always a medium grey (RGB code 178, 178, 178), the font Calibri, font size 28, font color black



(RGB code 0,0, 0). During each trial, participants were presented with a star (*) symbol that could appear in each of the four edges of an 100 × 100 pixels field centered on the screen. The position for the star was distributed equally and pseudo-randomly across trials. The star was presented in pink (RGB code: 255, 0, 127) for 700 ms. It was followed by a fixation cross presented in black (RGB code: 0, 0, 0) in the center of the screen for another 700 ms. Participants were instructed to first fixate the star and then shift their gaze to the cross as soon as it appeared. This procedure was implemented to support the synchronization of gaze direction data and behavioral data, because both were recorded by different computers. Directly after the fixation cross, the CRAT item without its solution was presented. The three triad words were stacked, centered, and 50 pixels apart in height. The third word was presented centrally. Below the three problem words, a question mark was presented as a place holder for the solution, separated from the problem words by a black line. Participants should press the space bar as soon as they came up with the solution for the problem. Each problem was presented for a total of 30 s or until participants pressed space to indicate that they came up with a solution. In case they did not press space, during the first 6 to 7 s (pseudorandom jitter), the first FoW rating had to be made. The question “How close to the solution do you feel?” was presented in German above a 5-point heat scale that consisted of five boxes (assigned range: 0 – 4), ranging from white (RGB code 255, 255, 255) to red (255, 0, 0) across different lighter tones of red. Participants could choose the corresponding via left and right arrow keys and should confirm via pressing the space bar. The next five FoW ratings were presented after 5–6 s (pseudorandom jitter), if the space bar was not pressed during the presentation of the problem. After reaching the upper time limit, the solution was presented in place of the question mark until participants pressed the space bar to indicate that they had understood how the solution word could be used to build compound words with

all three triad words. In case participants indicated that they came up with the solution by pressing the space bar, the question mark changed color and became green (0, 255, 0), indicating that they should speak their solution out loud. The solution was then written down by the experimenter for data analysis. Either after providing a solution or after the solution was presented due to not solving the problem after 30 s problem presentation, participants were presented with the question “Did you have an Aha! experience? - Yes/No.” The left and right arrow keys were assigned Yes/No counterbalanced across participants. The Aha! experience was described in the written instructions in line with the four criteria proposed by Topolinski and Reber (2010): It was defined as the solution being comprehended suddenly, being convinced of the truth of the solution, feeling that the solution is easy to understand, once they know it. Moreover, it should be associated with a positive feeling. Like Bowden and Jung-Beeman (2003), we further emphasized that the described feeling of Aha! does not have to be overwhelming, but should closely correspond to this, because such laboratory insight tasks with a high number of trials of the same type will probably very rarely lead to the overwhelming feeling of Aha! in contrast to natural situations. At the end of the presentation, participants were asked to fill out a questionnaire that asked them about their strategies in solving the riddles and some other potential confounds, as well as demographic data. Median duration was 1 h 45 min ($SD = 22$ min).

Data Analysis

Data were analyzed statistically using SPSS 24.0.0 for Mac OS (International Business Machines Corp., Armonk, NY, United States). We report conditioned probabilities in regard to the occurrence of Aha! given the solution was generated or not, once in regard to all items, and in regard to the number of FoW ratings per item. The number of rounds of ratings per item is dependent on how fast participants solved an item, as the FoW rating was given in intervals of 5–6 s, that is, 6–7 s for the very first round. All items with incorrectly generated solutions were excluded from data analysis, leaving only correctly generated and not generated solutions (relative number of excluded items: median = 0.08, $SD = 0.07$). In the following, when using the term “generated” we are always referring to correctly generated solutions. In case the distribution did not deviate from normality as tested via Kolmogorov–Smirnov test, non-parametrical tests were used, otherwise, parametrical tests were used. Effect sizes are reported as follows: Cohen’s d for repeated-measures t -tests and partial η^2 for repeated-measures analyses of variance (ANOVAs). For Wilcoxon signed-rank tests, we $ES = \frac{z}{\sqrt{N}}$, as suggested by Pallant (2007), where N is the number of observations not participants. In case of a violation of the sphericity assumption as tested via Mauchly’s test, Greenhouse–Geisser corrected p -values and ϵ are reported together with uncorrected F -values and uncorrected degrees of freedom to enhance readability. In addition to effect sizes, we calculated the statistical power for each test *post hoc*. We did not use *a priori* power analyses for several reasons:

- We did not have any particular expectations about the effect size, as there is no prior feelings-of-warmth study using the Compound Remote Associates Task, and because having to reach, e.g., clinical relevance, no *a priori* threshold for power is necessary (besides, of course, for the effect reaching statistical relevance).
- We would have conducted the study even when we would have been unable to reach the optimal sample size, because strong effects would nevertheless be found, and those are the ones that are most likely reliable.
- Calculating the sample size for reaching a certain effect size typically does not take into account the fact that one may collect not one data sample per subject and per cell, but several, as we did. Estimating the true mean of the participant with several measurements per cell leads to a more accurate estimate and hence to a better estimate of the true population mean. Therefore, the power of such studies should also be higher. This is the standard procedure (to increase the number of trials per condition for each participant) for many psychophysiological, neuropsychological and Neuroscience studies, where it would take too many temporal, personal, and monetary resources to increase the sample size. However, as far as the authors are aware, standard power calculation tools like G*Power provide no way to take this into account.

We therefore went along with a sample size that based on prior experience from numerous experiments led to large effect sizes. And indeed, as can be seen in our report of the statistical results, the minimum significant effect size was large.

Because it is highly discussed whether Aha! experiences can occur for non-generated solutions, that is, solutions that were presented after reaching the time limit without solving the problem, we also looked at the number of participants with empty cells for any condition.

RESULTS

Frequencies of Conditions

Firstly, we computed the mean frequency of all combinations of GENERATION (generated, not generated), that is, whether a problem was solved or not solved), and AHA (aha, no aha), that is, whether participants reported an Aha! experience after they came up with a solution (generated) or after the solution was

presented (not generated). All frequencies of conditions are listed in **Table 1**.

Another potential dependency we looked at was DIFFICULTY (easy, medium, hard). As can be seen in **Figure 2**, although the relative frequency of Aha! differed for generated and non-generated solutions, it did not differ according to problem difficulty. This observation was corroborated by a 2×3 repeated-measures ANOVA with the factors GENERATION and DIFFICULTY. As can be seen in **Figure 2**, there was a main effect of GENERATION [$F(1,35) = 6.26, p = 0.017, \eta_p^2 = 0.152, power = 0.682$], but no main effect of DIFFICULTY [$F(2,70) = 1.71, p = 0.197, \epsilon_{G-G} = 0.732, \eta_p^2 = 0.046, power_{G-G} = 0.295$], nor an interaction [$F(2,70) = 1.08, p = 0.875, \epsilon_{G-G} = 0.886, \eta_p^2 = 0.003, power_{G-G} = 0.065$]. As reported in **Table 1**, significantly more Aha! experiences were reported for generated [$P(\text{aha} | \text{generated}) = 0.76, SD = 0.27$] compared to non-generated solutions [$P(\text{aha} | \text{non-generated}) = 0.57, SD = 0.29$], as tested via Wilcoxon signed-rank test [$T = 179, p = 0.016, ES = 0.285, power = 0.654$].

Secondly, we looked at the number of participants with empty cells, that is, zero cases of a certain combination of aha/no aha and generation/non-generation (see **Table 2**). There was only one participant who never reported Aha! experiences for non-generated solutions³. As can be taken from **Table 1**, Aha! experiences were reported for almost half of all problems that could not be solved. Interestingly, seven participants reported no case of solutions generated without Aha!, suggesting that the CRAT really might be more of an insight problem-solving task, that is, a task which is mostly solved via insight.

Feeling-of-Warmth Course

The development of FoW can only be analyzed for items that were either not solved or solved after at least three rounds, because there is no curve otherwise. For items that were not solved, it will be interesting to see, whether participants felt closer to the solution by the end of the six rounds of FoW ratings or rather the 30 s of attempting to generating a solution.

³This number further seems to depend on how long it took participants to solve an item, because for items solved within three rounds, there were 14 participants who never reported no Aha!, for four rounds, there were 18, and for 5 rounds there were 23, while the number of participants who never reported Aha! experiences for solved items with 3, 4, and 5 rounds was very low and always the same 2. However, closer evaluation of their post-experimental questionnaires revealed no striking differences to the other participants.

TABLE 1 | Absolute (abs.) frequencies and conditional relative frequencies (rel.) of all conditions (without incorrectly generated items).

Condition	Min		Max		Mean		Standard deviation	
	abs.	rel.	abs.	rel.	abs.	rel.	abs.	rel.
P(aha generated \cap correct)	1	0.03	46	0.76	25.4	0.76	11.5	0.27
P(no aha generated \cap correct)	0	0.00	38	0.97	7.9	0.24	9.4	0.27
P(aha non-generated)	0	0.00	59	1.00	27.1	0.57	14.5	0.29
P(no aha non-generated)	0	0.00	52	1.00	21.0	0.43	15.1	0.29

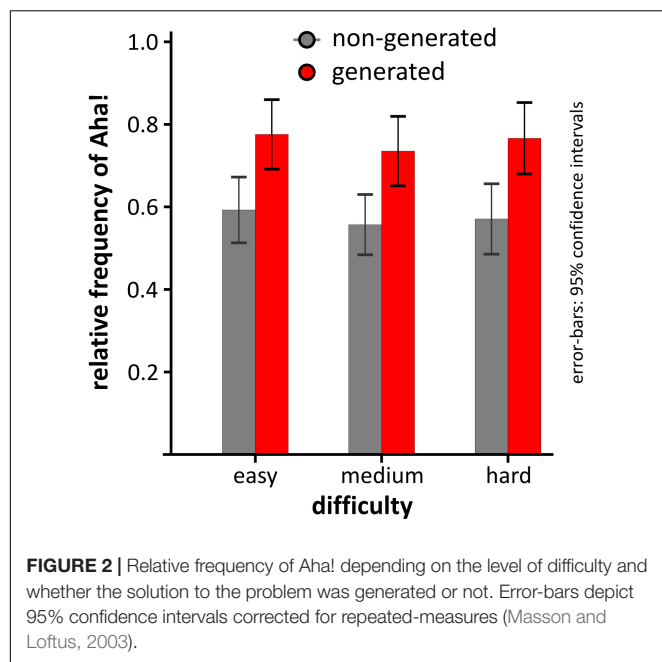
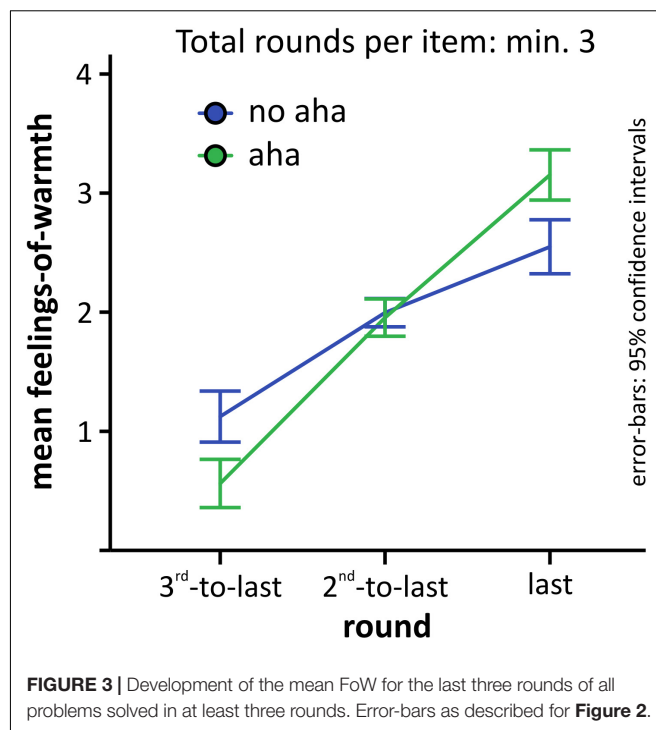


TABLE 2 | Number of participants with zero cases per condition.

Condition	Number of participants
aha \cap generated \cap correct	0
no aha \cap generated \cap correct	7
aha \cap non-generated	1
no aha \cap non-generated	1

Feeling-of-Warmth for Solved Items (Generated Solutions)

First of all, we looked at the last three rounds of any item that was solved after at least three rounds and compared FoW curves for items solved with versus without Aha!. All participants could be included, because all of them had at least one trial solved within three rounds. The mean number of trials was 4.0 ($SD = 4.1$) for no aha and 10.6 ($SD = 5.5$) for aha. As can be seen in **Figure 3** and conform with the idea that FoW would increase suddenly when the problem is solved via insight (i.e., with Aha! experience), the curve for problems solved with Aha! was below the one solved without Aha! for the third to last round, but increased highly and above those solved without Aha! for the last round, just before the solution was found. We computed a 3×2 repeated-measures ANOVA with factors ROUND(third-to-last, second-to-last, last) and AHA(aha, no aha) to compare mean FoW ratings, and found a highly significant main effect for ROUND [$F(2,52) = 132.27, p < 0.001, \epsilon_{G-G} = 0.845, \eta_p^2 = 0.836, \text{power}_{G-G} = 1.0$], no main effect of AHA [$F(1,26) = 1.05, p = 0.315, \eta_p^2 = 0.039, \text{power} = 0.167$], and a highly significant interaction [$F(2,52) = 15.63, p < 0.001, \epsilon_{G-G} = 0.642, \eta_p^2 = 0.375, \text{power}_{G-G} = 0.988$]. When comparing the difference between the means of the last minus third-to-last FoW ratings for problems solved with (2.59, $SD = 0.97$) versus without Aha! (1.43, $SD = 1.07$), we found a highly significant difference [$t(26) = 4.27, p < 0.001$, Cohen's



$d = 0.821, \text{power} = 0.956$], suggesting that the offset between the last and third-to-last FoW ratings may be a good marker for whether problem-solving is accompanied by a feeling of Aha! or not.

Secondly, we looked at FoW curves depending on the number of rounds needed until the solution was generated, and again compared them for items solved with versus without Aha!. We could only analyze problems solved within three (20 participants could be included, mean number of trials with aha = 4.3, $SD = 2.3$, mean number of trials with no aha = 2.15, $SD = 1.7$), four (13 participants, aha = 3.1, $SD = 1.9$, no aha = 1.9, $SD = 1.2$) and five rounds (9 participants, aha = 2.8, $SD = 2.5$, no aha = 1.8, $SD = 1.4$). This pattern, i.e., that most participants solved most items within the first three rounds, is typical for the CRAT, as Bowden and Jung-Beeman (2003) report that CRAT items are mostly solved within the first 15 s, which corresponds to three rounds in our design. Due to the low number of participants, we refrained from statistical inference testing, but report the data descriptively.

The pattern for problems solved within three rounds (**Figure 4A**) was highly similar to the pattern reported above and is in line with the idea that FoW rises suddenly for problems solved with aha. The curve for five rounds (**Figure 4C**) is also in line with this hypothesis, whereas the curves for items solved within four rounds (**Figure 4B**) seem to completely overlap for aha and no aha. The curves for four and five rounds suggest that the slope of the FoW curve is more of a second order polynomial function (tested with the curve fitting tool from <https://mycurvefit.com>, access date: 2018-03-28) rather than linear (as might be inferred from the three-point curves), in line with the model suggested by Reber et al. (2007).

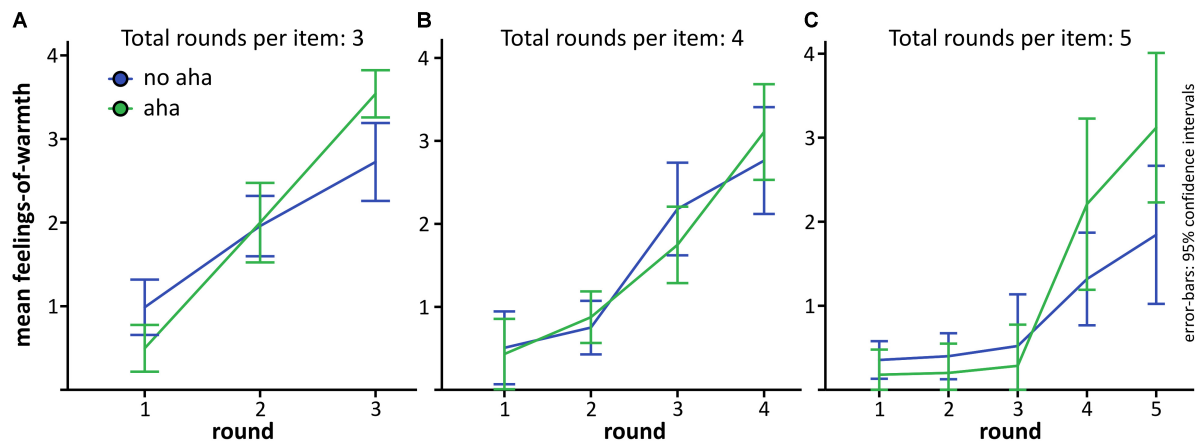


FIGURE 4 | Development of FoW depending on the number of rounds per problem until the solution was generated. **(A)** Problems solved after 3 rounds. **(B)** Problems solved after 4 rounds. **(C)** Problems solved after 5 rounds. Problems solved after 6 rounds are not depicted, because the number of participants who had at least one problem solved during the last round was very low. Error-bars as described for **Figure 2**.

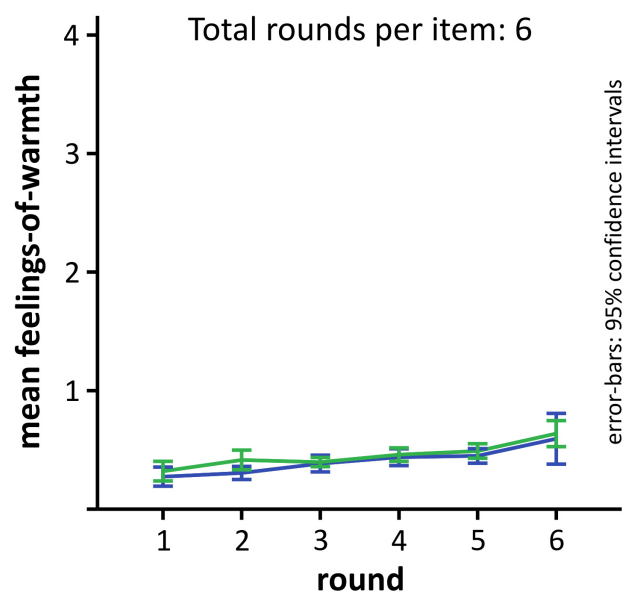


FIGURE 5 | Development of FoW for unsolved problems for comparison. The Aha/no Aha! decision was based on the solution that was presented after time-out. Error-bars as described for **Figure 2**.

Feeling-of-Warmth for Unsolved Problems (Non-generated Solutions)

For comparison, we also analyzed the development of FoW over time for unsolved problems, and compared the curves for problems solved with versus without Aha!. Thirty-four participants could be included in this analysis. Two participants had empty cells (one only reported Aha! experiences for non-generated solutions and the other only no Aha!). As expected, the curves show a flat course and did not differ for aha and no aha (**Figure 5**). A 6×2 repeated-measures ANOVA revealed a significant main effect of

ROUND [$F(5,165) = 9.78, p < 0.001, \epsilon_{G-G} = 0.393, \eta_p^2 = 0.229, power_{G-G} = 0.977$], but no main effect of AHA [$F(1, 33) = 1.66, p = 0.207, \eta_p^2 = 0.229, power = 0.239$], nor was there a significant interaction [$F(5, 165) = 0.342, p = 0.666, \epsilon_{G-G} = 0.323, \eta_p^2 = 0.010, power_{G-G} = 0.097$]. There was a low but significant increase of FoW over time, although it stayed between the lowest two values (0, 1), suggesting that participants did never feel particularly close to the solution, before it was presented.

DISCUSSION

The present study investigated the relationship between the subjective closeness to the solution, assessed as FoW ratings, the subjective Aha! experience, item difficulty, and the generation of solutions for CRAT problems. This is the first study to investigate the relationship between a measure of the subjective closeness to the solution (FoW) depending on whether an insight occurred or not (feeling of Aha!).

Feeling-of-Warmth Differ for Problems Solved With Versus Without Aha!

The observed FoW curves for problems solved in at least three rounds of 5–6 s each showed that insights, operationalized as experiencing a feeling of Aha! upon solving a problem, were characterized by a curve that showed a sudden increase of FoW during the last two FoW ratings (<10 s) before reporting a solution. The slope was much steeper for problems solved with than without Aha!. This finding is in line with an observation made by Metcalfe and Wiebe (1987) who measured FoW for solved insight problems as compared to analytical problems. However, as the authors defined insight problem-based and not process-based, we have to be careful when comparing their results with our findings. In terms of the continuous and discontinuous approaches on insight described by Zander et al. (2016), our

results seem to be more in support with the continuous model, which proposes a slow increase that ends in a sudden surge, similar to the curve proposed by Reber et al. (2007) for intuitive problem solving and we conceive a curve that depicts Bowers' approach on insight as the final stage of intuitive problem-solving (Bowers et al., 1990). However, because we have only enough trials with at least three FoW ratings and because FoW was assessed in intervals of 5–6 s, our curve is not fine-grained enough to say for sure whether the FoW development is more similar to Reber's intuition curve or his insight curve for the subjective closeness to the solution. Those two model curves only differ in regard to whether the slope is level (insight) or whether it rises just a little (intuition) before culminating in a sudden surge just before the solution is found. What we can derive with certainty from our data is that problems solved with Aha! do show more of a sudden increase at the end and those solved without Aha! show more of a gradual rise. Especially the curve with five FoW ratings suggests that there is a very sudden increase in FoW for problems solved with as compared to without Aha!. Although we have only few participants that solved problems after five FoW assessments, this suggests that if we were to assess FoW in a more continuous way, it would be in line with the insight model curve by Reber et al. (2007).

We propose that the observed FoW curves support the following cognitive process for insight solutions: When searching for the remote association that comprises the solution word of a CRA problem, the remote associations activated by means of spreading activation are at first not available to consciousness (see Öllinger and von Müller, 2017, for an alleged model of the underlying search process—combing spreading activation and constraint satisfaction). However, at the time when the associations are set up between all triad words and the solution word, its activation level becomes strong enough to become consciously available. This comprises the moment of Aha!.

Our findings are in contrast to those of Hedne et al. (2016) who measured FoW for magic tricks solved either with Aha! or without. They found no difference in FoW ratings (differential measure = last – first rating, angular measure = differential warmth/s) for tricks solved with or without Aha!. An important difference between Hedne and colleagues' and our study is the frequency of Aha! for solved problems. Whereas for our task 76% of all solved items were solved with Aha!, Hedne and colleagues report almost the reverse distribution, namely 29% of all solved items were solved with Aha!. The low number of problems solved via insight may have led to a less accurate estimation of the true mean of FoW, not allowing to find differences between FoW for insight and non-insight solutions, even if there were any. This low frequency of Aha! for magic tricks seems a little surprising at first, because Danek and colleagues, who pioneered magic tricks as a task to investigate insight problem solving, always report higher distributions: 41.1% (Danek et al., 2013b), and 66.5% (Danek and Wiley, 2017). However, Hedne et al. (2016) reported not the Aha! rate for all correctly solved items, as Danek et al. (2013b) and we did, but Aha! for all solved items (be it correct or incorrect) (personal communication with Hedne, 2018 March 25). So, to make our reported Aha! rate more comparable across studies, we additionally calculated $P(\text{Aha!} \mid \text{generated}(\text{correct} \cap \text{incorrect}))$

which was 72.9 % (SD = 21.6) and still deviated considerably from the other studies. There are other potential explanations of the diverging findings, such as differences of the Aha! definition participants were provided with, or that the tasks really differed considerably in their probability to induce an Aha! experience. Hedne et al. (2016) indeed defined the Aha! experience by only one criterion, that is, that the solution appeared “out of nowhere,” whereas the current study and Danek and colleagues included at least two of the four criteria suggested by Topolinski and Reber (2010): suddenness, being convinced of the truth of the solution, ease of understanding, and positive affect.

All in all, our findings support the idea that subjective feelings of closeness to the solution rise more suddenly for insight than for no insight. Moreover, they show the importance of how insight is defined (experimenter-based, participant-based) and if the participant-based approach is chosen, how the Aha! experience is described to the participants, when investigating differences in FoW curves for insight and no insight solutions. In terms of a more fine-grained differentiation between intuition, insight, and incremental problem solving as proposed by Reber et al. (2007), we unfortunately cannot draw any clear conclusions, because we ended up with too few trials for a statistical comparison between detailed FoW curves (4–5 ratings). It may be advised for future studies on the topic, to increase the number of trials.

The Aha! Experience Is Related to the Generation of a Solution but Not Problem Difficulty

We found that Aha! experiences were more often reported when CRAT problems were solved compared to when the solution was comprehended only after failing at generating it (76% versus 57%). However, Aha! experiences were still reported relatively often even for presented solutions, suggesting that insight-like experiences can even be felt when comprehension is induced. Another study using CRAT problems reported Aha! frequencies of 56% for correctly solved items (Jung-Beeman et al., 2004). Unfortunately, there is no published data from other labs on Aha! rates for solutions to problems that were presented after a failed solution attempt. Importantly, we are not referring to problems that were solved incorrectly, but problems for which no solution was generated within the time limit. In previous studies, we observed an equal distribution of Aha! for generated and non-generated solutions for Mooney stimuli, that is, pictorial riddles (Kizilirmak et al., 2016a), or the reverse pattern, that is, a higher frequency of Aha! for non-generated CRAT problems (Kizilirmak et al., 2016c). However, either the stimulus material differed considerably (verbal semantic problems here versus pictorial visual problems in Kizilirmak et al., 2016a) or the conditions used (solution process repeatedly interrupted at short intervals and only problems where participants had the chance to solve them here versus problems with or without the chance to solve them in Kizilirmak et al., 2016c). It is therefore difficult to compare our results. The diverging findings for Aha! rates of correctly solved CRAT problems nonetheless suggest that there are many different factors aside from the problem type that play a role in whether items are solved with or without Aha!.

In contrast to our hypothesis, the frequency of Aha! experiences was not dependent on the difficulty level of the CRAT problem. In other words, whether the solution to a difficult, medium, or an easy CRAT problem is comprehended, the probability of experiencing an Aha! moment was equal. This observation complements the observations made by Knoblich et al. (1999) who found a relationship between task difficulty and the probability of a representational change. In matchstick arithmetic tasks the degree that a chunk decomposition or a constraint relaxation requires determined the solution rates and solution times. Given this evidence, our results suggest that problem difficulty of the CRAT is not exclusively caused by the degree of representational change but by an additional source of problem difficulty such as semantic distance, that is not related to the feeling of aha!. This interpretation is in line with the multiple causes of difficulty approach (Kershaw and Ohlsson, 2004; Kershaw et al., 2013; Öllinger et al., 2014).

On the other hand, it could also be that the variation of problem difficulty for CRAT problems was too low to enable us to find any significant differences between difficulty levels and Aha! frequency even if they existed. Other studies which quantified the Aha! rather than recording binary occurrence, report significant correlations between the strength of the Aha! experience and solution rates (as an operationalization of problem difficulty). For example, Webb and colleagues report significant but weak correlations [$r(99) = 0.26-0.27$] between solution rates (accuracy) and Aha! ratings of classic insight problems (such as the rope problem) and also for an English version of the CRAT (Webb et al., 2017). Danek and colleagues further observed significant differences for mean Aha! ratings of correct versus incorrect solutions (Danek et al., 2013b; Danek and Wiley, 2017). Hence, it may be that only the strength of the Aha! is related to problem difficulty, similar to the complexity of the representational change required (Knoblich et al., 1999), but not whether it occurs or not. Future studies should focus on tasks with a larger variability between task difficulty and assess solution rates as well as Aha! rates and the strength of the Aha! to test this assumption.

Limitations

There are several limitations for the conclusions that can be drawn from the current manuscript. First, we do not know in how far our results can be generalized to other types of problems besides the CRAT and probably the incoherent triads that Zander et al. (2016) referred to in their review. Second, to assess the course of FoW, we interrupted the problem-solving process of our participants in intervals of 5–7 s. We do not know in which way this or even asking for a FoW rating in itself may influence the ratings. What we noticed is that the frequency of reported Aha! experiences differs from our other experiments using the CRAT with the same time for solving the problems (30 s in total). As we reported in 2016 in the *Journal of Problem Solving*, 24% of all items were solved with Aha!, 21% solved without Aha!, 41% were not solved with Aha!, and 14% were not solved without Aha! (Kizilirmak et al., 2016c). Thus, it looks like there may be

an influence of the interruptions or the FoW ratings *per se*. However, as the paradigm also differed in the conditions present, because in the 2016 study, we had items for which participants had the chance to solve CRA items and those whose solutions were presented immediately, we cannot be sure that the diverging findings are only due to the interruptions or consciously considering the subjective closeness to the solution, as they might also be due to not having a no-chance to solve condition.

CONCLUSION

Our results provide support for the idea that insight solutions pop into awareness suddenly, probably around 5–12 s before being able to indicate behaviorally that the problem has been solved. The slope for the last three FoW ratings (5–6 s apart) was significantly steeper for problems solved with Aha! compared to those without, lending support to the idea that the subjective feeling of closeness to the solution does not rise or only rises weakly until the solution is verbalizable. It is even conceivable that participants would be able to voice the solution at the time of the second-to-last FoW rating which is much higher than the third-to-last for insight, but only press the button after they have confirmed that their solution is a valid compound word for the three words comprising the CRA item. Future studies could instruct participants to voice a solution whenever they have a candidate, even when they are unsure, in addition to assessing FoW ratings, to test this hypothesis. We further found that CRA problems are mainly solved via insight (i.e., accompanied by a subjective feeling of Aha!) and that insight solutions do not depend on problem difficulty. This finding is very useful in regard to learning from insight, as other studies have shown that solving problems by insight facilitates long-term memory encoding (Danek et al., 2013a; Kizilirmak et al., 2016a): It is not necessary for the problem to be especially difficult to be solved with an Aha! experience. Hence, for the application of learning from insight, even easy problems can be used.

ETHICS STATEMENT

The study was approved by the Ethics Committee of the University of Hildesheim, Germany. Participation was voluntary and compensated via course credits. All participants gave their informed written consent and were explicitly told that they could abort the procedure at any time without the need for an explanation and without any negative consequences.

AUTHOR CONTRIBUTIONS

JMK, A-RK, and KF-S conceived and designed the study. JMK programmed the experiment and additional scripts for later data analysis. JK carried out a pilot study under JMK's and AR-K's supervision (Bachelor's thesis in Psychology) in A-K's lab. VS conducted the main study under JMK's supervision (Master's thesis as a pre-service teacher) in KF-S's lab.

JMK, JK, and VS carried out the statistical analyses. JMK wrote the first draft of the manuscript. JMK, MÖ, A-RK, and KF-S revised the manuscript. All authors read and approved the submitted version.

FUNDING

This work was partly supported by a grant assigned to AR-K by the German Research Foundation (Deutsche

Forschungsgemeinschaft) for project TPA10N, part of the collaborative research center “Neurobiology of Motivated Behavior” SFB779, awarded to the University of Magdeburg.

ACKNOWLEDGMENTS

We would like to thank Amory Danek, Margaret Webb, and Gillian Hill for valuable discussions, and David Mietzner who helped with data collection.

REFERENCES

- Bowden, E. M., and Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behav. Res. Methods Instrum. Comput.* 35, 634–639. doi: 10.3758/BF03195543
- Bowers, K. S., Farvolden, P., and Mermigis, L. (1995). “Intuitive antecedents of insight,” in *The Creative Cognition Approach*, eds S. M. Smith, T. B. Ward, and R. A. Finke (Cambridge, MA: MIT Press), 27–51.
- Bowers, K. S., Regehr, G., Balthazard, C., and Parker, K. (1990). Intuition in the context of discovery. *Cogn. Psychol.* 22, 72–110. doi: 10.1016/0010-0285(90)90004-N
- Danek, A. H., Fraps, T., von Müller, A., Grothe, B., and Öllinger, M. (2013a). 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., von Müller, A., Grothe, B., and Öllinger, M. (2013b). 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. (2014). 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
- Danek, A. H., Williams, J., and Wiley, J. (2018). Closing the gap: connecting sudden representational change to the subjective Aha! experience in insightful problem solving. *Psychol. Res.* doi: 10.1007/s00426-018-0977-8 [Epub ahead of print].
- Davidson, J. E., and Sternberg, R. J. (1984). The role of insight in giftedness. *Gift. Child Q.* 28, 58–64. doi: 10.1177/001698628402800203
- Hedne, M. R., Norman, E., and Metcalfe, J. (2016). Intuitive feelings of warmth and confidence in insight and noninsight problem solving of magic tricks. *Front. Psychol.* 7:1314. doi: 10.3389/fpsyg.2016.01314
- Ilg, R., Vogele, K., Goschke, T., Bolte, A., Shah, J. N., Pöppel, E., et al. (2007). Neural processes underlying intuitive coherence judgments as revealed by fMRI on a semantic judgment task. *Neuroimage* 38, 228–238. doi: 10.1016/j.neuroimage.2007.07.014
- Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biol.* 2:E97. doi: 10.1371/journal.pbio.0020097
- Kershaw, T. C., Flynn, C. K., and Gordon, L. T. (2013). Multiple paths to transfer and constraint relaxation in insight problem solving. *Think. Reason.* 19, 96–136. doi: 10.1080/13546783.2012.742852
- Kershaw, T. C., and Ohlsson, S. (2004). Multiple causes of difficulty in insight: the case of nine-dot problem. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 3–13. doi: 10.1037/0278-7393.30.1.3
- Kizilirmak, J. M., Galvao Gomes da Silva, J., Imamoglu, F., and Richardson-Klavehn, A. (2016a). Generation and the subjective feeling of “aha!” are independently related to learning from insight. *Psychol. Res.* 80, 1059–1074. doi: 10.1007/s00426-015-0697-2
- Kizilirmak, J. M., Thuerich, H., Foltz-Schoofs, K., Schott, B. H., and Richardson-Klavehn, A. (2016b). Neural correlates of learning from induced insight: a case for reward-based episodic encoding. *Front. Psychol.* 7:1693. doi: 10.3389/fpsyg.2016.01693
- Kizilirmak, J. M., Wiegmann, B., and Richardson-Klavehn, A. (2016c). Problem solving as an encoding task: a special case of the generation effect. *J. Probl. Solving* 9, 59–76. doi: 10.7771/1932-6246.1182
- Klein, G., and Jarosz, A. (2011). A Naturalistic Study of Insight. *J. Cogn. Eng. Decis. Mak.* 5, 335–351. doi: 10.1177/1555343411427013
- 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, 1534–1555. 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: 10.1016/j.actpsy.2008.08.008
- Masson, M. E. J., and Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Can. J. Exp. Psychol.* 57, 203–220. doi: 10.1037/h0087426
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* 69, 220–232. doi: 10.1037/h0048850
- Metcalfe, J., and Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Mem. Cogn.* 15, 238–246. doi: 10.3758/BF03197722
- Ohlsson, S. (1992). “Information-processing explanations of insight and related phenomena,” in *Advances in the Psychology of Thinking*, eds M. Keane and K. Gilhooly (London: Harvester-Wheatsheaf), 1–44.
- Öllinger, M., Jones, G., and Knoblich, G. (2014). The dynamics of search, impasse, and representational change provide a coherent explanation of difficulty in the nine-dot problem. *Psychol. Res.* 78, 266–275. doi: 10.1007/s00426-013-0494-8
- Öllinger, M., and von Müller, A. (2017). Search and Coherence-Building in Intuition and Insight Problem Solving. *Front. Psychol.* 8:827. doi: 10.3389/fpsyg.2017.00827
- Pallant, J. (2007). *SPSS Survival Manual*, 3rd Edn. New York, NY: McGrath Hill.
- Reber, R., Ruch-Monachon, M.-A., and Perrig, W. J. (2007). Decomposing intuitive components in a conceptual problem solving task. *Conscious Cogn* 16, 294–309. doi: 10.1016/j.concog.2006.05.004
- Topolinski, S., and Reber, R. (2010). Gaining insight into the “aha” experience. *Curr. Dir. Psychol. Sci.* 19, 402–405. doi: 10.1177/0963721410388803
- Webb, M. E., Little, D. R., and Cropper, S. J. (2017). Once more with feeling: normative data for the aha experience in insight and noninsight problems. *Behav. Res. Methods* doi: 10.3758/s13428-017-0972-9 [Epub ahead of print].
- 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
- Zander, T., Öllinger, M., and Volz, K. G. (2016). Intuition and insight: two processes that build on each other or fundamentally differ? *Front. Psychol.* 7:1395. doi: 10.3389/fpsyg.2016.01395

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 Kizilirmak, Serger, Kehl, Öllinger, Foltz-Schoofs and Richardson-Klavehn. 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.



How Working Memory Provides Representational Change During Insight Problem Solving

Sergei Korovkin^{1*}, Ilya Vladimirov¹, Alexandra Chistopolskaya^{1,2} and Anna Savinova¹

¹ Department of Psychology, Yaroslavl State University, Yaroslavl, Russia, ² Laboratory for Cognitive Studies, The Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia

OPEN ACCESS

Edited by:

Amory H. Danek,
Universität Heidelberg, Germany

Reviewed by:

Stellan Ohlsson,
University of Illinois at Chicago,
United States
Mareike Wieth,
Albion College, United States

*Correspondence:

Sergei Korovkin
korovkin_su@list.ru

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 27 April 2018

Accepted: 12 September 2018

Published: 01 October 2018

Citation:

Korovkin S, Vladimirov I,
Chistopolskaya A and Savinova A
(2018) How Working Memory
Provides Representational Change
During Insight Problem Solving.
Front. Psychol. 9:1864.
doi: 10.3389/fpsyg.2018.01864

Numerous studies of insight problem solving are focused on both the control and storage systems of working memory. We obtained contradictory data about how working memory systems are involved in insight problem solving process. We argue that measuring the dynamics of the control system and storage systems through the course of problem solving can provide a more refined view on the processes involved, as a whole, and explain the existing controversies. We theorize that specific insight mechanisms require varying working memory capacities at different stages of the problem solving process. Our study employed a dual task paradigm to track the dynamics of working memory systems load during problem solving by measuring the reaction time in a secondary probe-task during different stages of problem solving. We varied the modality (verbal, visual) and the complexity of the probe-task during insight and non-insight problem solving. The results indicated that the dynamics of working memory load in insight problems differs from those in non-insight problems. Our first experiment shows that the complexity of the probe-task affects overall probe-task reaction times in both insight and non-insight problem solving. Our second experiment demonstrates that the solution of a non-insight problem is primarily associated with the working memory control system, while insight problems rely on relevant storage systems. Our results confirm that insight process requires access to various systems of working memory throughout the solution. We found that working memory load in non-insight problems increases from stage to stage due to allocation of the attentional control resources to interim calculations. The nature of the dynamics of working memory load in insight problems remains debatable. We claim that insight problem solving demands working memory storage during the entire problem solving process and that control system plays an important role just prior to the solution.

Keywords: insight, working memory, representational change, probe-task, executive functions, storage and control systems

INTRODUCTION

For a long time, the problem of working memory role in problem solving, particularly in insight problems, was (and still is) a focus of numerous studies in the field. A number of reviews and original research articles have been devoted to working memory in problem solving (Hambrick and Engle, 2003; Wiley and Jarosz, 2012). An interest in the role of working memory during

insight problem solving stems from the information processing theories viewing insight as a representational change that can possibly occur within working memory (Ohlsson, 1992, 2011; Öllinger et al., 2013). Baddeley's working memory model describes both the storage systems (visuo-spatial sketchpad, phonological loop and episodic buffer) required to hold representations and the control system (central executive) enabling the restructuring process (Baddeley, 2002). Investigating the processes involved in working memory during problem solving can provide a unique perspective into its internal structure. The conclusions drawn from the working memory studies can be useful for answering the vital question: "Are there any specific mechanisms dedicated to insight solutions?"

Information processing theories seek to determine whether there is something special in insight phenomenon that makes it uniquely different from analytical problem solving; whether insight is a metacognitive epiphenomenon accompanying a broad range of unrelated processes, or whether it involves specific cognitive mechanisms. At first sight, comparing the information processing occurring in different types of problems is a good way to approach this question. Although this widespread approach seems encouraging, studies that employ the traditional experimental designs and paradigms commonly used in working memory research (e.g., distractors in the dual task paradigm, working memory span studies) often report controversial results.

Contradictions in Working Memory Effects

A number of studies have revealed contradictory results regarding the role of working memory in insight problem solving process (DeCaro et al., 2016, 2017; Chuderski and Jastrzębski, 2017). The discussion on the role of working memory in insight primarily focuses on the working memory control system in problem solving. Some studies claim that working memory is a crucial component of both insight and non-insight problem solving processes. Working memory capacity has a strong positive correlation with insight problem solving performance and creativity (Cinan and Doğan, 2013; Chuderski, 2014; Chuderski and Jastrzębski, 2018). De Dreu et al. (2012) demonstrated that creative task performance suffers under working memory load. DeYoung et al. (2008) showed that insight problems are as related to working memory as non-insight problems, but only insight problem solving is related to divergent thinking and breaking the frame. Murray and Byrne (2005) found that accuracy in insight problem solving is positively correlated with working memory storage as well as with attention switching processes, but not with selective and sustained attention. However, some studies revealed different effects of working memory control and storage systems on insight problems. Necka et al. (2016) claimed that insight problem solving positively correlates with the recognition of the already presented items in working memory (updating processing in working memory storage) rather than with the substitution of old items with new ones (executive control).

Other studies revealed that working memory affects insight problems less than non-insight problems. Concurrent counting

during the problem solving process shows a greater negative effect on non-insight than insight problems, and these findings were supported by ERP data via P300 amplitude analysis (Lavric et al., 2000). Ash and Wiley (2006) demonstrated that insight problems with reduced initial phase are not as related to working memory. Fleck (2008) found that insight problem solving correlates only with verbal working memory, but not with control system or spatial working memory. Verbal working memory may affect only the initial phases of problem comprehension without affecting specific insight processes.

Some studies clearly demonstrated that working memory deficits can be beneficial to insight problem solvers. For example, lateral frontal lobe damage patients solve matchstick problems better compared to healthy participants (Reverberi et al., 2005). Participants with mild alcohol intoxication perform remote associate tests better, faster, and experience more insight solutions (Jarosz et al., 2012). Higher working memory capacity is associated with lower matchstick problem accuracy due to inhibited constraint relaxation (DeCaro et al., 2016). Additionally, higher working memory also leads participants to employ complex ineffective strategies in water jar tasks despite the availability of simpler strategies (Beilock and DeCaro, 2007).

Moreover, there is different data regarding the role of storage systems of working memory in insight problem solving. Performance in insight problem solving is not linked to the control system but is associated with the verbal and visuo-spatial components of working memory (Gilhooly and Fioratou, 2009). Gilhooly and Murphy (2005) claimed that verbal insight problem solving rates are positively related to verbal working memory (vocabulary scores) and spatial insight problem solving rates are positively related to spatial working memory (spatial flexibility). Performance on the nine-dot problem is related to spatial but not verbal working memory (Chein et al., 2010). However, the storage systems of working memory are not involved in insight problem processing independently of the control system. Performance in Compound Remote Associate problems can be predicted by both verbal working memory and attention switching (Chein and Weisberg, 2014). On the other hand, verbal working memory distraction via articulatory suppression enhances insight problem solving because it reduces the verbal-based problem processing (Ball et al., 2015). Surprisingly, the preliminary load of spatial working memory enhances the solution rate in the T-puzzle insight problem (Suzuki et al., 2014).

Some controversies can be accounted for by the differences in the procedures and task materials used in these studies. However, the main source of these controversies might stem from two other major factors: heterogeneity of the problem solving process and the complex nature of the working memory model.

Heterogeneity refers to the idea that insight problem solving process consists of several phases (problem comprehension, impasse, and representation restructuring) that are not equally related to working memory. For example, the selective forgetting hypothesis claims that forgetting and memory clearing occurs during the impasse phase (Simon, 1977; Ohlsson, 1992). According to this hypothesis, reduced attention control should be less demanding on the control system of working memory during the impasse phase compared to other phases. The relationship

between working memory and insight problem solving can change from phase to phase during this process (DeCaro et al., 2017). The dynamics of insight problem solving processes are infrequently discussed within the working memory studies (Ash and Wiley, 2006; Korovkin et al., 2014; Yeh et al., 2014; Lv, 2015). At the same time, heterogeneity of the phases in insight problem solving was demonstrated in eye-movement studies (Knoblich et al., 2001; Ellis et al., 2011; Yeh et al., 2014). Thus, we propose that the role of working memory in problem solving should be discussed in regards to each phase separately.

The working memory model itself is a challenging theoretical framework featuring certain ambiguity in terms of relevant components and parameters. This challenge is aggravated by the lack of unity between theoretical models of working memory (Engle et al., 1999; Baddeley, 2002; Cowan, 2010). Two main approaches to working memory studies in problem solving are experimental and individual differences approaches (Hambrick and Engle, 2003). These approaches differ not only in their methodology but also in their theoretical basis. The experimental approach typically incorporates the distraction paradigm and is based on Baddeley's (2002) working memory model. Distractors selectively target one of the storage systems of working memory to isolate the modal-specific effects within the problem solving process. The individual differences approach is based on the concept of working memory capacity and focuses on the quantity of stored items. We consider it necessary to take all characteristics of working memory into account to shed light on the processes that make up insight. Understanding the control system is crucial to describing overcoming of the impasse. Additionally, understanding the modal-specific storage systems is necessary to reveal the mechanisms of representation restructuring. Finally, understanding the overall capacity is essential for assessing the information processing aspects of problem solving.

Probe-Task

Conventional methods used in working memory studies do not capture the dynamics of working memory load over time. We propose a technique that can accomplish this goal. This technique relies on the assumptions drawn from Kahneman's (1973) resource model. According to this model, cognitive resources are limited and distributed in concordance with subjective importance. Therefore, if two tasks are performed at the same time continuously, the performance drop in one of them, indicating that available resources have been allocated to the second task instead. If participants should engage in problem solving, while performing a monotonous secondary probe-task, the reaction time in the probe-task should increase whenever the primary problem solving process becomes particularly resource demanding, and vice versa.

Wieth and Burns (2014) clearly showed that both insight and non-insight problem solving processes suffer under multitasking conditions. This fact is in line with our assumptions that the problem solving process competes with the secondary task for resources. Moreover, the interference which occurs due to the competition does not appear to be very damaging to the problem solving process. The surprising result is that providing an incentive does not allow participants to overcome

the difficulties associated with multitasking. This may be due to limited attentional resource which cannot be significantly increased. Instead, the authors assume that high motivation leads to surface processing. This means that in the multitasking condition participants shift their attention to the simpler task, essentially making the secondary task the main task. This fact could be a limitation when only using reaction times as the only dependent variable in a dual-task paradigm. Thus, we used reaction times as a main dependent variable and solution rates, solution times, and probe-task accuracy as additional indicators.

The overall problem-solving trial time can be divided into several equal time stages. For example, if the problem was solved in 300 s, the data obtained within the first 100 s, middle 100 s, and last 100 s would represent three stages and corresponding dynamics. Splitting this process into three stages allows us to trace the temporal dynamics of working memory.

Based on the assumption that working memory resources are not unified, we can also vary the content of the secondary probe-task in such a way that it should compete with only some of the systems, but not others. For example, by varying the overall complexity of the probe-task we can investigate the overall working memory capacity demands in problem solving, while, by altering the content of the probe-task (e.g., modality of stimuli) we can isolate the effect of specific storage systems availability.

This technique allows us to answer the following questions on the role of working memory during the insight problem solving process:

- (1) Is working memory necessary for insight problem solving process? Does working memory load vary across insight and non-insight problems? Does the insight problem solving process add to working memory load in addition to single probe-task performance?
- (2) Are working memory storage systems, the control system, and their overall capacities that are involved in insight problems drastically different compared to non-insight problem solving?
- (3) Is there a specific pattern of the temporal dynamics of working memory load during the insight problem solving process? Do capacity, storage, and control systems demands differ across various phases of problem solving?

The study described below was designed to answer these questions regarding the role of working memory and its components in insight problem solving. It was operated under the aforementioned assumptions associated with the dual-task paradigm. This allowed us to operationalize the level of working memory load (low/high) caused by the problem solving process via the reaction time in the simultaneously performed probe task; the slower the reaction time, the higher the working memory load.

EXPERIMENT 1

Experiment 1 was conducted to test hypotheses about the role of working memory in insight problem solving. First, we hypothesized that working memory is necessary for insight

problem solving; although not to the same degree as for non-insight problem solving. We predicted that working memory load in insight problem solving will be significantly greater than baseline yet significantly lower than in non-insight problem solving. Second, we expected the probe-tasks to take up the working memory capacity proportionally to their complexity. Third, we predicted that different stages of the problem solving would require different amounts of working memory; more specifically, working memory load should be higher toward the end of problem solving in both problem types due to the accumulation of problem-related information.

To test these hypotheses, we employed a 2 (problem type) \times 2 (probe type) \times 3 (problem stage) full factorial within-subject design with the reaction time in the probe task serving as a dependent variable. The problem type variable consisted of two levels: insight problems and non-insight problems. The probe type variable featured two levels varying in the number of items held in working memory: a simple probe-task (two possible choices) and a complex probe-task (six possible choices). The problem stage acted as a grouping variable with three levels: the average reaction time in the probe task during the first, the middle, and the last part of overall problem solving time course. Full factorial design was incorporated leading to four (2×2) conditions that were later split into three stages each.

Method

Participants

Participants in the experimental group were 32 people (25 women), aged 18–34 ($M = 22.16$; $SD = 3.18$). Participants in the control group were 32 people (22 women), aged 18–28 ($M = 21.66$; $SD = 2.61$). The majority of the sample consisted of undergraduate and graduate students at Yaroslavl State University. All participants were tested individually, took part voluntarily, and were not paid for their participation.

Stimuli

We had two types of probe-tasks:

The Simple Probe-Task

Participants were shown the pictures of two alternatives: a circle and a square. Participants were instructed to respond by pressing the left key if they saw a circle and the right key button if they saw a square. The participants' goal was to perform the task as quickly and accurately as possible.

The Complex Probe-Task

Participants performed the same task, but had six alternatives choices instead. The alternatives were: a square, a circle, a triangle, a cross, a pentagon, and a hexagon. Participants were instructed to press the left key if they saw a circle, a triangle or a pentagon, and the right key in all the other cases.

All probe-tasks were presented in the center of the screen. All figures were black; the background was white. All trials were preceded by a brief (100 ms) blank screen. These probe-tasks were designed to be demanding, yet realistically possible to be performed simultaneously with the primary problem.

We used two types of problems as a primary task:

Non-insight Problems

These problems have clear conditions, a solution algorithm and a logical answer. Participants know all important operators for finding a correct solution and have the right representation of conditions. An example of a non-insight problem: "Given four coins of identical look and feel, two of which are slightly heavier and two are slightly lighter, how could one identify all of them when only allowed to use the balance scale twice?"

Insight Problems

These problems require a change of operators or representation, wherein the participant does not know a new system of operators. The solution occurs suddenly and is often associated with an emotional response. An example of an insight problem: "If you have black socks and brown socks in your drawer, mixed in a ratio of 4–5, how many socks will you have to take out to make sure that you have a pair the same color?"

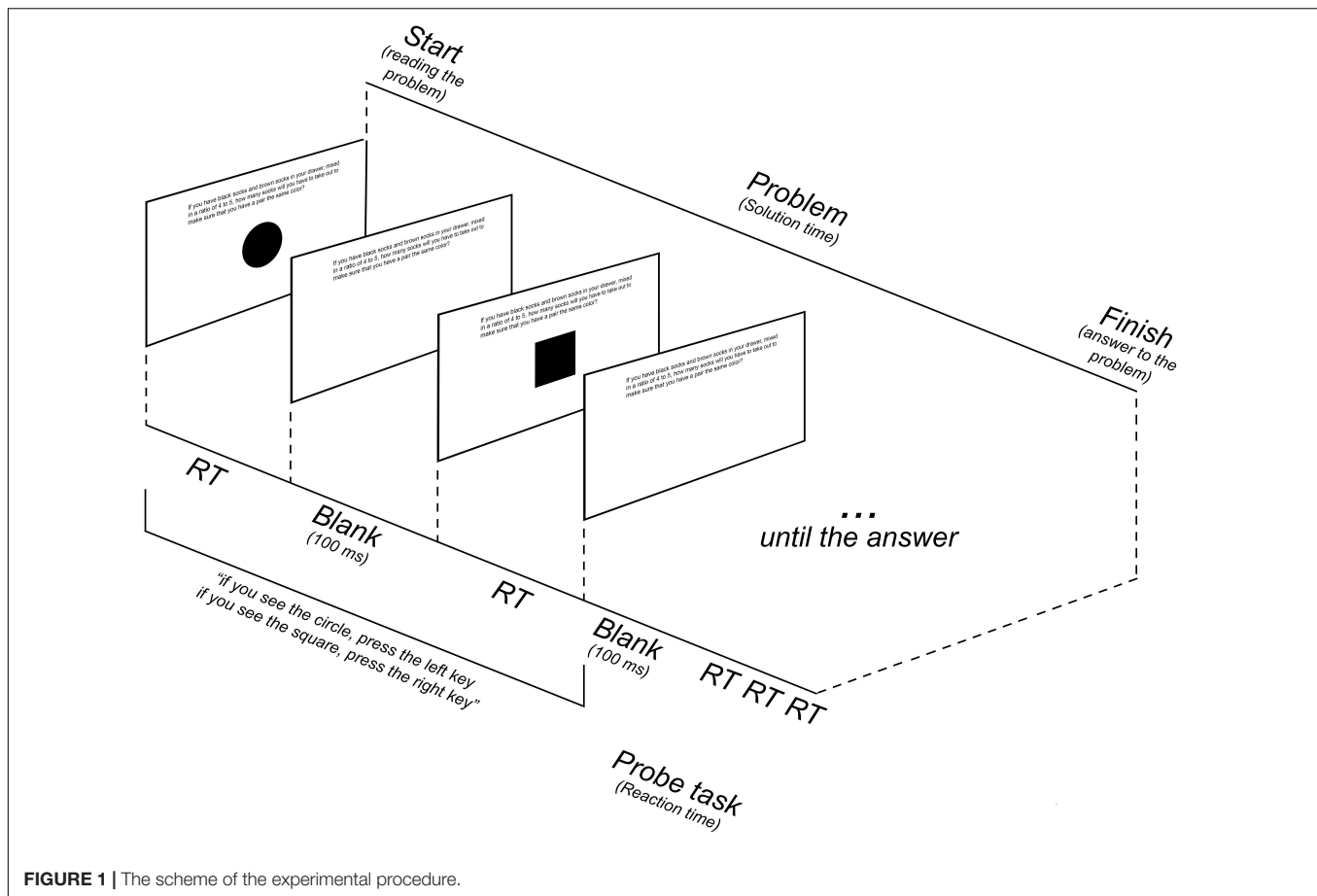
We selected problems with average solution time between 60 and 150 s. In this experiment we used verbal problems only. Participants were not allowed to use notes and write any information down because this would conflict with the probe-task performance. The problems were solved aloud, and participants answered verbally. All the problems are presented in the **Supplementary Materials**. The control group (no probe-task) was included in this study to verify whether or not problem solving was substantially altered by the dual-task itself and whether probe-task performance is affected by the problem solving process in the first place. Participants in the control group solved the same set of problems as in the experimental group but without any secondary task (4 insight and 4 non-insight problems).

The experiment was performed with PsychoPy2 scripts (Version 1.81.02; Peirce, 2008) on the HP Envy x360 15-ar001ur computer with a 15.6" screen.

Procedure

Each participant completed two parts of the experiment: practice trials and experimental trials. The purpose of the practice trials was to familiarize participants with the secondary probe-tasks. During the practice trials participants completed 30 trials of both types of probe-tasks – one at a time, not engaged in the problem solving process. There were 30 trials of each type of probe-tasks presented in random order. Average reaction time of the probe-tasks was calculated and served as a baseline for future comparisons. The scheme of the procedure is presented in **Figure 1**.

When participants finished the practice trials, they proceeded to the experimental trials. Each participant solved two insight and two non-insight problems per each of two probe-task levels in random order (eight problems total). The probe-task trials repeated indefinitely for as long as it took to finish the primary problem. Participants had up to 5 min to solve each problem and were instructed to report the proposed solution verbally. Unsolved trials were not included in the data analysis. Participants were provided with a short break (up to 1 min) after each problem trial.



Preliminary Analysis

Each of the 32 participants in the experimental group attempted to solve 8 problems (256 problems in total). Trials in which participants solved the problem in under 30 s were excluded from the analysis, since such a short thinking time might be indicative of participants' exposure to a given problem in the past. Trials that took more than 5 min were considered unsolved and were excluded as well. Besides those exclusions, extreme values of the probe-task reaction times above 3 IQR were considered indicative of participant's low engagement in the task and, therefore, were excluded from the analysis. Overall, 15 non-insight trials and 50 insight trials were excluded from the analysis. The rest of the trials constituted the obtained data set. The control group data was pre-processed the same way: 9 non-insight trials and 51 insight trials were excluded.

Each problem solving trial was split into three equal time intervals similar to the approach previously used by Knoblich et al. (2001). After that, we averaged the probe-task reaction time within each of those stages, resulting in three probe-task reaction time observations per problem trial. Data obtained from problems in the same condition were averaged across participants, giving us a single data point per each condition for each participant.

The decision to split the overall solution time into three stages was the result of a compromise: while having only two

stages would insufficiently represent the course of the problem solving process since it would leave the middle stage of the problem solving unobserved; having more stages can lead to over-conservative statistical estimations due to the aggressive multiple comparison correction, making it hardly possible to reach significance even with a profound effect. We consider the division into three stages theoretically plausible as well: the first stage represents the familiarization with a problem, the middle stage is representative of an impasse, and the final stage is related to overcoming the impasse as well as solution verification.

Results

The preliminary analysis revealed that participants typically successfully solve the majority of the problems (the average solution rate is 77.9%). Participants were successfully performing the probe-tasks as well (95.7% accuracy). This data suggests that participants were adequately focused on both the primary problem and secondary probe-tasks. We found that there are no significant differences between the control and experimental groups in solution times, $F(1,62) = 0.004$, $P = 0.952$, $\eta_p^2 < 0.001$; there is no main effect of problem type, $F(1,62) = 0.565$, $P = 0.455$, $\eta_p^2 < 0.009$; as well as no interaction between the group and problem type factors, $F(1,62) = 0.163$, $P = 0.687$, $\eta_p^2 = 0.003$. We, therefore, argue that the probe-task does not substantially alter the problem solving process itself. Despite

the difference between the solution rates of insight and non-insight problems, we suggest that the difficulty of problems has no major effect on reaction time because for both problem types, only trials of the approximately same duration (30–300 s) were analyzed. A brief overview of these results can be found in **Table 1**. For a detailed analysis refer to the **Supplementary Table S4**.

A $3 \times 2 \times 3$ repeated measures ANOVA with Greenhouse–Geisser correction was performed to test our hypotheses. The results are shown in **Figures 2, 3**. A main effect of the probe-task type was found for reaction time, $F(1.94, 40.72) = 184.18$, $P < 0.001$, $\eta_p^2 = 0.898$. *Post hoc* pairwise comparisons with the Bonferroni adjustment revealed that reaction time in all three groups were significantly different. The fastest condition was the practice trials with a single probe-task without parallel problem solving ($M = 0.79$; $SD = 0.15$); the slowest condition was non-insight problem solving with a parallel probe-task ($M = 1.93$; $SD = 0.43$). The difference between the practice trial and non-insight problem conditions was found to be significant [$t(27) = -14.83$, $p < 0.001$, $r = -0.874$]. The probe reaction time in the insight problem condition ($M = 1.67$; $SD = 0.42$) was significantly greater than in practice trials [$t(28) = 12.97$, $p < 0.001$, $r = 0.828$] and significantly less than in non-insight problems [$t(28) = -4.32$, $p < 0.001$, $r = -0.319$]. Thus we may conclude that insight problem processing competes with the probe-task for resources of working memory. This means that working memory is necessary for insight problem solving, but is not as crucial for non-insight problem solving.

A main effect of probe type was revealed [$F(1, 21) = 32.65$, $P < 0.001$, $\eta_p^2 = 0.609$]. The results are shown in **Figures 4, 5**. *Post hoc* analysis of the probe-tasks in practice trials showed that the simple probe-task was performed faster ($M = 0.57$; $SD = 0.06$) than the complex probe-task ($M = 0.99$; $SD = 0.26$), $t(29) = -9.25$, $p < 0.001$, $r = -0.736$. Moreover, the simple probe-tasks were significantly faster than the complex probe-tasks both in the insight [$t(24) = -2.53$, $p = 0.018$, $r = -0.247$] and non-insight problems [$t(28) = -2.93$, $p = 0.007$, $r = -0.253$].

As we expected, the analysis did not reveal any interaction between the probe type and the stage factor [$F(1.77, 37.21) = 0.5$, $P = 0.59$, $\eta_p^2 = 0.023$], between task type and probe type

[$F(1.7, 35.8) = 0.47$, $P = 0.601$, $\eta_p^2 = 0.022$], nor between probe type, task type, and the stage factors [$F(3.04, 63.76) = 0.9$, $P = 0.447$, $\eta_p^2 = 0.041$].

There was a significant main effect of the stage factor [$F(2, 41.95) = 76.04$, $P < 0.001$, $\eta_p^2 = 0.784$] and an interaction between the task type and stage factors [$F(3.13, 65.81) = 31.69$, $P < 0.001$, $\eta_p^2 = 0.601$]. Various task conditions of the probe-task performance revealed different dynamics. The reaction time decreased in the practice trial over time [the first and second stages: $t(30) = 3.21$, $p = 0.003$, $r = 0.278$; the first and third stages: $t(30) = 4.55$, $p < 0.001$, $r = 0.356$], representing a typical learning curve. At the same time, the reaction time increased over time in both insight and non-insight problems [the first and second stages of insight problems: $t(28) = -3.74$, $p < 0.001$, $r = -0.322$; the first and third stages of insight problems: $t(28) = -6.5$, $p < 0.001$, $r = -0.51$; the first and second stages of non-insight problems: $t(29) = -6.04$, $p < 0.001$, $r = -0.535$; the first and third stages of non-insight problems: $t(29) = -13.22$, $p < 0.001$, $r = -0.764$].

Post hoc pairwise comparisons with the Holm–Bonferroni adjustment revealed a gradual increase in reaction time values in all conditions. There were significant differences in non-insight problems when performing the simple probe-task between the first and second stages [$t(29) = -5.46$, $p < 0.001$, $r = -0.454$], the first and third stages [$t(29) = -9.28$, $p < 0.001$, $r = -0.681$], and the second and third stages [$t(29) = -5.26$, $p < 0.001$, $r = -0.416$]. The same effect was observed for the complex probe-task in non-insight problems between the first and second stages [$t(30) = -4.37$, $p < 0.001$, $r = -0.401$] and the first and third stages [$t(30) = -7.2$, $p < 0.001$, $r = -0.587$]. Reaction times for both simple and complex probes increased from stage to stage during non-insight problem solving. This may be due to a gradual increase of working memory load by analytical processes and the accumulation of problem-related information over time.

Surprisingly, we observed a stage-to-stage increase of the reaction time for insight problems as well. The reaction time for the simple probe in the first stage of insight problems was smaller than in the second stage [$t(27) = -4.64$, $p < 0.001$, $r = -0.272$] and the third stage [$t(27) = -4.18$, $p < 0.001$, $r = -0.351$]. Similarly, the reaction time for the complex probe in the first stage of insight problems was smaller than in the second stage [$t(26) = -2.56$, $p = 0.017$, $r = -0.304$] and the third stage

TABLE 1 | The descriptive statistics of solution time and solution rate of the problems in Experiment 1.

	Control group				Experimental group			
	<30 s	>300 s	Solution rate	Solution time, sec (SD)	<30 s	>300 s	Solution rate	Solution time, sec (SD)
Insight problems	31 (24.22%)	20 (15.625%)	77 (60.16%)	105.47 (48.88)	22 (17.19%)	24 (18.75%)	82 (64.06%)	107.98 (46.01)
Non-insight problems	3 (2.34%)	6 (4.69%)	119 (92.97%)	102.04 (34.82)	3 (2.34%)	4 (3.125%)	121 (94.53%)	110.52 (34.66)

<30 sec, number of previously known problems or problems solved in less than 30 s and excluded from the further analysis. >300 s, number of problems solved in more than 5 min and excluded from the further analysis.

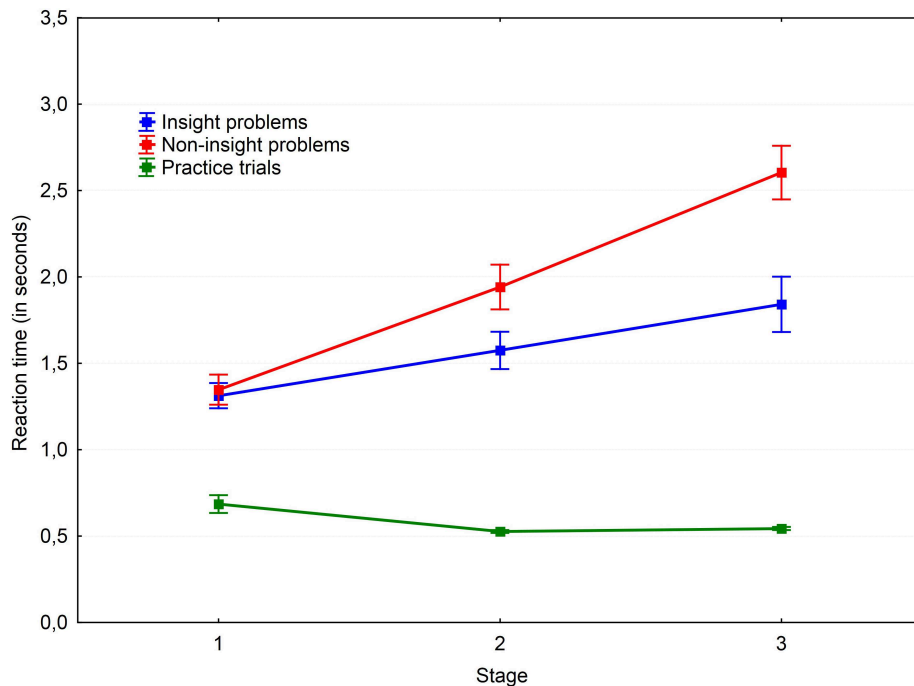


FIGURE 2 | Dynamics of working memory load via the simple probe-task. Vertical bars denote standard errors.

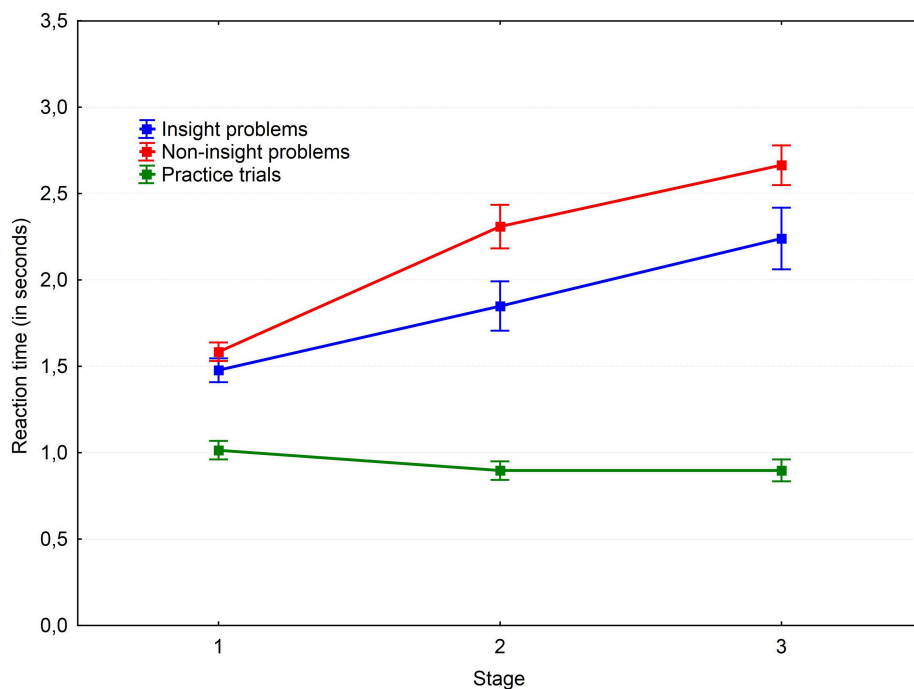


FIGURE 3 | Dynamics of working memory load via the complex probe-task. Vertical bars denote standard errors.

$[t(26) = -3.99, p < 0.001, r = -0.466]$. Nevertheless, the reaction times (presumably indicative of working memory load) were generally higher in non-insight problems. However, pairwise comparisons revealed that insight and non-insight problems

differ at the second stage $[t(26) = -2.4, p = 0.024, r = -0.274]$ and the third stage $[t(26) = -5.1, p < 0.001, r = -0.465]$ in the simple probe condition and at the second stage $[t(26) = -2.55, p = 0.017, r = -0.296]$ and the third stage $[t(26) = -3.06, p = 0.005,$

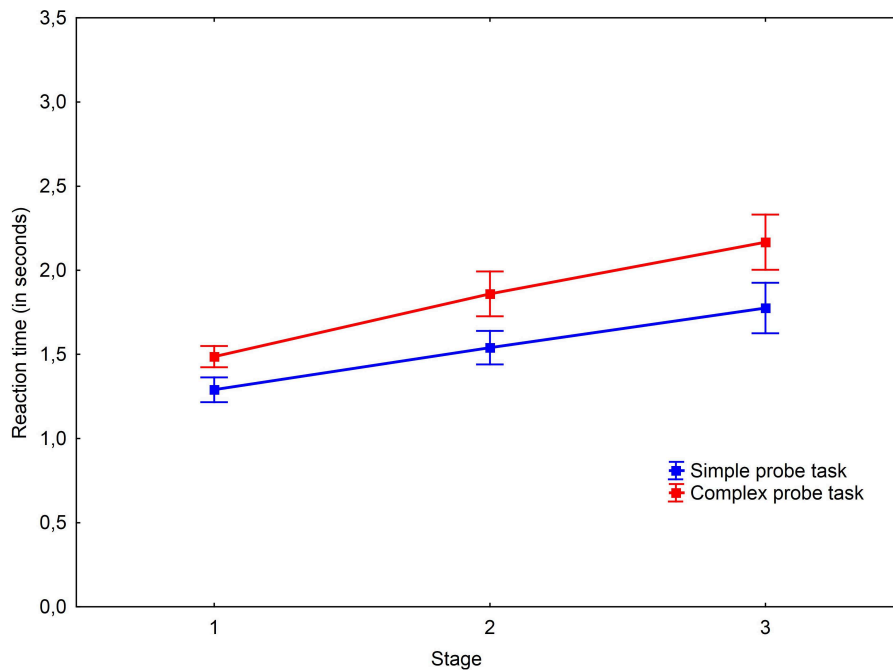


FIGURE 4 | Dynamics of working memory load in the insight problems. Vertical bars denote standard errors.

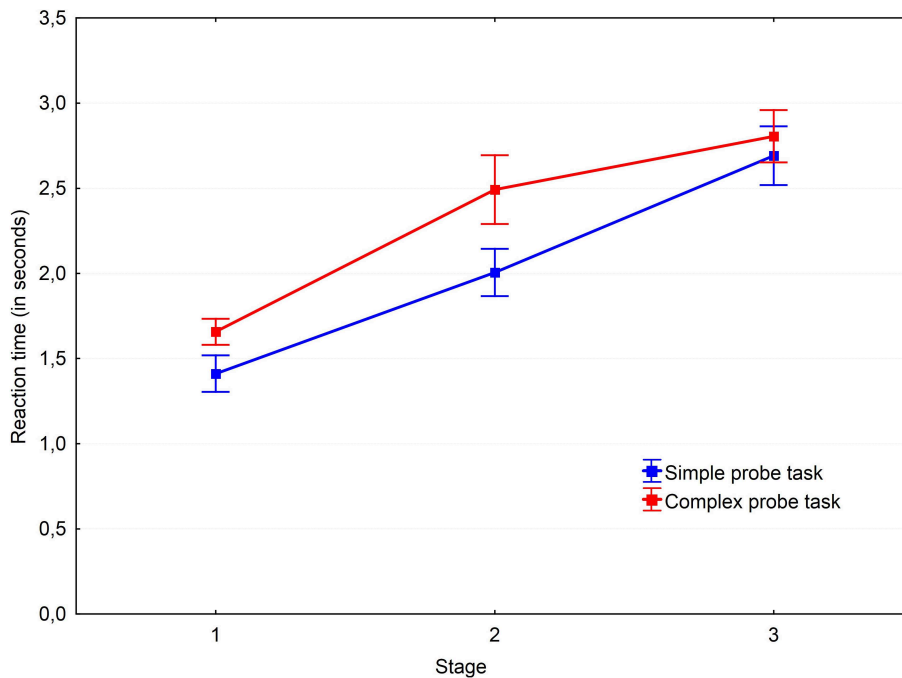


FIGURE 5 | Dynamics of working memory load in the non-insight problems. Vertical bars denote standard errors.

$r = -0.356$] in the complex probe condition. The reaction time for the same probe types in the first stage is equal for the insight and non-insight problems.

The complex probe-task was performed slower both in both insight and non-insight problems but not at the third stage. The

reaction times in non-insight problems were different between the probes at the first stage [$t(28) = -3.68, p < 0.001, r = -0.344$] and second stage [$t(28) = -2.5, p = 0.019, r = -0.267$]. The same results may be observed in insight problems where the probes were different at the first stage [$t(24) = -2.82, p = 0.009$,

$r = -0.277$] and second stage [$t(24) = -2.48$, $p = 0.021$, $r = -0.241$]. We argue that simple probes become harder during the later stages of the problem solving process because of the concurrent problem solving processes in the final stage of a solution.

Discussion

The obtained results generally confirmed our hypotheses. Hypothesis 1, that working memory is necessary for insight problem solving although not to the same degree as for non-insight problem solving, was completely confirmed. We found that working memory load in insight problem solving is higher than the baseline reaction time in practice trials. This leads to a conclusion that while insight problem solving is demanding in terms of working memory, non-insight problem solving is notably more so. While non-insight problem processing includes planning, holding interim calculations in memory, and control; solving insight problems may involve posing and testing hypotheses, problem comprehension, restructuring of a representation, and verification of solutions. These processes are cognitively demanding but are relatively rare, impermanent, and eventual.

Hypothesis 2 was confirmed by the main effect of probe-task type. Probe-task processing occupies a part of working memory capacity during the problem solving process proportionally to task complexity. Comparison of the probe-tasks in the practice trials revealed that these tasks initially differ by their complexity. The complex probe performance during the main problem solving process is slower than the simple probe performance in all problem types. On the one hand, this shows that the probes are performed well and do not crucially distract from the main problem solving process. On the other hand, it can be described as a modality-independent increase in working memory load under the complex condition because we used different modalities in the main problem (the problems were presented textually) and probe-tasks (the probes were presented visually).

Hypothesis 3 was confirmed by the main effect of the stage factor and an interaction of stage and task factors. We found that the patterns of reaction time dynamics are different in various conditions. We observe a clear learning curve in the practice trials for both probes where reaction times decrease from stage to stage. In contrast, working memory load in the insight and non-insight problems prominently increases. The notable difference between the first and third stages in both types of problems demonstrates that cognitively demanding processing accumulates during the problem solving process. Working memory load in the first stage is similar in insight and non-insight problems and is significantly higher than baseline. We theorize that the same processes related to problem comprehension and building a mental model of the problem are implemented at this stage. The further increases to reaction time in non-insight problem solving may be explained by the increasing processing. As mentioned earlier, the same pattern of working memory load is observed in insight problem solving; the closer one gets to insight solution, the more important of a role working memory plays in insight problem solving. Nevertheless, working memory load does not increase to the same degree in non-insight problems.

Unexpectedly, we found that the probe-tasks of different types are performed similarly at the third stage both in the insight and non-insight problems. Based on the qualitative analysis of the experimental sessions, we speculate that participants might have distracted themselves from the probe-tasks to continue engaging in the problem solving process during the later stages of the trial. This distraction might have obscured the difference between the probe-task types. It also means that parallel competition between the two tasks becomes impossible and turns into switching between the tasks. This also indicates the heavy load of working memory during the last stage of the insight solution.

There were some limitations in this experiment. First, increase in reaction time during the last stage could have been confounded by the process of the verbalization required to report the solution. Second, the obtained results do not allow us to draw any definitive conclusions regarding the role of working memory modal-specific systems. Some of such effects were reported to be found in previous studies (Gilhooly and Fioratou, 2009; Chein et al., 2010). We designed and conducted Experiment 2 to overcome the limitations of Experiment 1.

EXPERIMENT 2

To overcome the limitations of the first experiment, we modified the procedure and attempted to isolate the effect of solution verbalization and verification by separating it from the dual task performance. When a participant found a solution for a problem, they were instructed to press a pause button to report the solution and get the experimenter's response. If the participant's solution was incorrect, they resumed the dual task performance. Additionally, we attempted to identify the modality of the representational processing in insight problem solving. To do so, we introduced the variable of congruence – whether the problem and the probe-task were of the same modality or not. Representational change in insight problem solving can occur within the modal-specific storage systems while being relatively unaffected by the control system. Visual representational change in insight problems can be processed in the visuo-spatial sketchpad, while verbal restructuring – in the phonological loop. In other words, if the problem and the probe-task are both visual or both verbal – the competition occurs on the storage system level (congruent condition), while if the problem and the probe-task are presented in different format – they do not compete in the same storage systems, only for non-specific control system (non-congruent condition).

The general hypotheses of Experiment 2 were as follows:

- (1) Working memory storage systems are involved in both types of problem solving.
- (2) There is a modal specificity of working memory storage system load in insight problem solving. Insight problem solving is expected to be more demanding in terms of working memory storage systems, while non-insight problem solving was expected to heavily rely on the control system.

- (3) Working memory load varies across different stages of the problem solving process. We expected an increased control system load in non-insight problem solving and an increased storage systems load during the last stages of insight problem solving.

To test these hypotheses, we employed the $2 \times 2 \times 3$ factorial within-subject design. The first factor was primary problem-type with two levels: insight and non-insight. The second factor was a congruence of the primary problem format and the probe-task with two levels: congruent and non-congruent. The stage acted as a grouping variable with three levels: first, middle and last stage of the trial. The response time in the probe-task was measured.

Method

Participants

Participants in the experimental group were 32 volunteers (22 women; age $M = 21.03$; $SD = 3.01$). Participants in the control group were another 32 volunteers (21 women), aged 18–34 ($M = 21.5$; $SD = 4.86$). The majority of the sample consisted of undergraduate and graduate students at Yaroslavl State University. All participants were tested individually; participation was not monetarily compensated.

Stimuli

We modified the materials used in the original experiment, introducing two formats of the primary problem – involving visual images and text, as well as two formats of the probe-tasks: visual and text versions as well. These versions were meant to load the corresponding working memory storage system. The congruent condition always featured the problem and the probe-task of the same format (both visual or both text), while the opposite was true for the non-congruent condition.

The two types of the probe-tasks were as follows:

The Text Task

Participants were presented with two alternatives: open or closed syllables. They were instructed to respond with the right key every time they saw a closed syllable (e.g., “LON”) and with the left key every time they saw an open syllable (e.g., “PLE”). They were also instructed to perform the task as quickly and accurately as possible.

The Visual Task

Participants were presented with two alternatives: obtuse or acute angles. They were instructed to respond with the left key every time they saw an obtuse angle and with the right key every time they saw an acute angle. The instructions were to perform the task as quickly and accurately as possible.

Non-insight Text Problems

These problems have clear conditions, solution algorithms and logical answers. Participants know all important operators necessary to find the correct solution and to build the right condition representation. The problem solution is mainly based on the text code. An example of a non-insight text problem: “Three couples went to a party together. One woman was dressed in red, another one – in green and the third one – in blue. The men were also dressed in one of these colors. When all three couples danced, a man in red danced with the woman in blue. “Christina, it is funny, isn’t it? None of us danced with a partner dressed in the same color.” Think about the man dancing with the woman in red. What color is he wearing?”

Non-insight Visual Problems

These problems are similar to non-insight text problems, but the solution is mainly based on the visual code. An example of a non-insight visual problem is the following matchstick problem: “Turn inequality into equality by moving one match: $8 + 3 - 4 = 0$ ” (Figure 6).

Insight Text Problems

These problems are based on a representational change, but the participant is not aware of the new system of operators. Finding an answer occurs suddenly for solvers and is often accompanied by an emotional response. The solution is mainly based on the text code. An example of an insight text problem: “Sally Lu likes eucalyptus more than pine. She likes electric lighting and does not like to sit by candlelight. Eccentric people evoke more sympathy from her than balanced ones. What do you think is Sally’s profession - an economist or an accountant?”

Insight Visual Problems

These problems are similar to insight text problems, but the solution is mainly based on the visual code. An example of an insight visual problem: “Organize 6 identical pencils to get 4 equiangular triangles.”

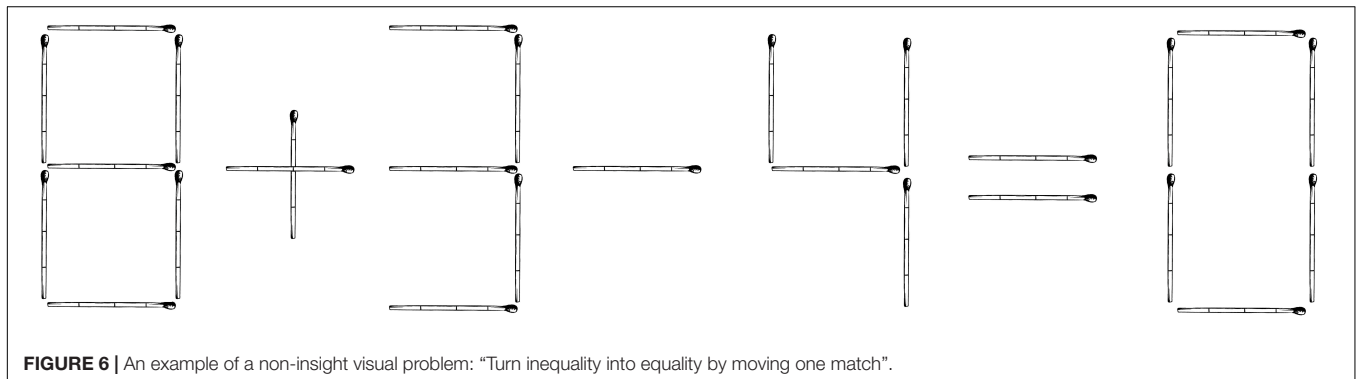


FIGURE 6 | An example of a non-insight visual problem: “Turn inequality into equality by moving one match”.

The problems with an average solution time between 70 to 185 s were selected for the experiment. Participants were not allowed to use notes or write any information down because this would conflict with the probe-task performance. The problems were solved aloud, and participants answered verbally. All the problems are presented in the **Supplementary Materials**.

The control group was included in this study to compare the solution times and solution rates of the problems solved in the dual-task conditions vs. the problems solved without any secondary task. Participants in the control group solved the same set of problems as in the experimental group but without any secondary task (4 insight and 4 non-insight problems).

The experiment was conducted using PsychoPy2 scripts (Version 1.81.02; Peirce, 2008) on the ASUS K55VD computer with a 15.6" screen.

Procedure

The procedure used in Experiment 2 was identical to the procedure of the Experiment 1. Each participant solved 8 problems total – one problem trial in each condition presented in random order. The problems were presented at the upper part of the screen; the probe-task stimuli were presented at its center.

The participants were solving problems while performing the probe-tasks continuously the whole time, except for when they were verbally reporting the solution to a problem they were solving. If their proposed solution was incorrect – they resumed performing the secondary probe-task as well as thinking about the problem solution. After the solution to the problem was found, participants had an option to take up to a 1 min break before proceeding to the next problem.

As in Experiment 1, the average response time for the probe-task served as a dependent variable of interest.

Preliminary Analysis

The data analysis was identical to that from Experiment 1. Thus, each of the 32 participants attempted to solve 8 problems (256 problems in total), but some problem solving trials were excluded: we excluded unsolved problems (took more than 5 min to solve) and problems that were solved in less than 30 s (due to possibility that participant already knew the answer). Besides this, extreme values for the probe-task reaction time above 3 IQR were identified as outliers. Trials with these outliers were excluded from further analysis. Overall, eleven insight problem trials and eighteen non-insight problem trials were excluded from the analysis for those reasons.

Identical to the experimental group, each of the 32 participants in the control group solved 8 problems – one trial in each condition. We used the same criteria for data exclusion. Overall, 51 insight problem trials and 25 non-insight problem trials were excluded from the analysis.

Each problem solving trial was preprocessed and its solution time was split into three equal time intervals as in the Experiment 1. The average reaction time for the probe-task in each of three stages was calculated.

Results

Obtained results indicated that participants typically solved the majority of the problems (the average solution rate is 70.3%). Similarly, the participants were successfully performing the probe-tasks (87.6% accuracy). This arguably shows that participants were actively engaged in the process and paid sufficient attention and effort to both the primary and secondary tasks.

The average probe-task reaction time in non-insight ($M = 1.55$; $SD = 0.33$) problem solving was greater than in insight problem solving ($M = 1.35$; $SD = 0.27$), $t(31) = 5.16$, $p < 0.001$, $r = 0.304$. Besides, the average probe-task reaction time in insight problems was significantly greater than when the probe-tasks were performed without problem solving ($M = 0.86$; $SD = 0.11$), $t(31) = 9.08$, $p < 0.001$, $r = 0.748$ (**Figure 7**).

We found that solution times in the experimental condition were greater both in insight [$t(62) = 2.61$, $p = 0.011$, $r = 0.315$] and non-insight [$t(62) = 4.51$, $p < 0.001$, $r = 0.497$] problems compared to the control condition. This supports the notion that modally specific probe-tasks affect the problem solving process, however, the probe-tasks were not destructive enough to meaningfully alter the solving process. The solution times of insight problems were significantly greater than that of non-insight problems [$t(31) = 2.29$, $p = 0.029$, $r = 0.269$] in the control group. However, there was no significant difference between insight and non-insight problems solution times in the experimental group [$t(31) = 1.97$, $p = 0.058$, $r = 0.185$]. These results revealed that insight problems were harder than we expected in the control condition, but probe-tasks involvement removed the difference between insight and non-insight problems. The solution rate data showed that insight problems were solved less often. A brief overview of these results can be found in **Table 2**. For a detailed analysis refer to the **Supplementary Table S4**.

Problem Type

A repeated measures ANOVA revealed a significant main effect of problem type. The probe-task was performed significantly slower during non-insight problem solving compared to insight problem solving, $F(1,30) = 37.75$, $p < 0.001$, $\eta_p^2 = 0.557$.

Modality Congruence

No significant main effect of modality congruence was revealed. The probe-task average reaction times were equal both in cases when the probe-task was of the same modality as the primary problem and in cases where they were different (e.g., visual problem and a text probe-task), $F(1,30) = 0.24$, $p = 0.631$, $\eta_p^2 = 0.008$.

Problem Stage

A repeated measures ANOVA with Greenhouse–Geisser correction revealed a significant main effect of problem stage, $F(1.68,50.26) = 19.59$, $p < 0.001$, $\eta_p^2 = 0.395$. A Holm–Bonferroni *post hoc* comparison revealed that the probe-task reaction time was significantly smaller in the first stage ($M = 1.34$, $SD = 0.04$) compared to the middle stage ($M = 1.42$, $SD = 0.05$), while the

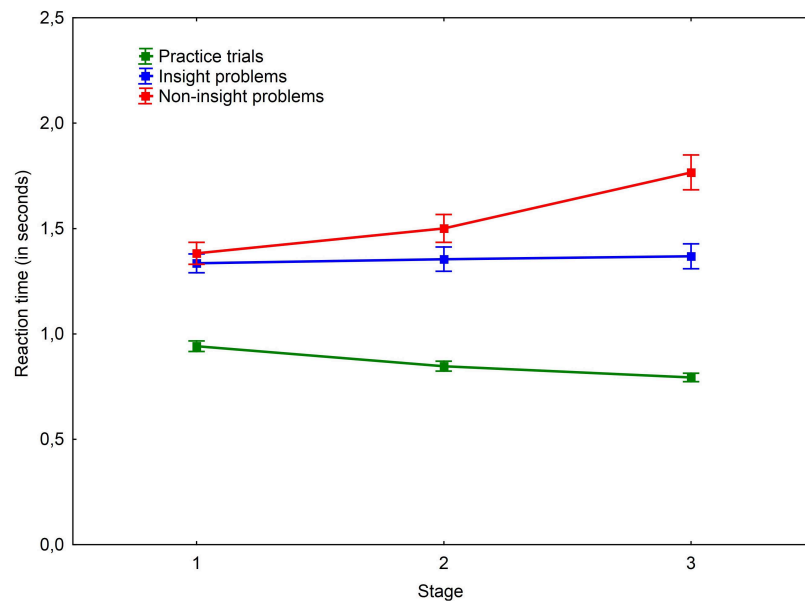


FIGURE 7 | Average dynamics of working memory load in all probe-tasks. Vertical bars denote standard errors.

TABLE 2 | The descriptive statistics of the solution time and the solution rate of the problems in Experiment 2.

	Control group				Experimental group			
	<30 s	>300 s	Solution rate	Solution time, sec (SD)	<30 s	>300 s	Solution rate	Solution time, sec (SD)
Insight problems	17 (13.28%)	34 (26.56%)	77 (60.16%)	141.26 (54.08)	4 (3.13%)	32 (25%)	92 (71.88%)	172.4 (40.33)
Non-insight problems	9 (7.03%)	16 (12.5%)	103 (80.47%)	114.44 (41.09)	0	18 (14.06%)	110 (85.94%)	158.01 (36)

<30 s, a number of problems solved in less than 30 s and excluded from the further analysis. >300 s, a number of problems solved in more than 5 min also excluded from the further analysis.

last stage featured the highest probe-task reaction time ($M = 1.59$, $SD = 0.07$).

Problem Type × Modality Congruence Interaction

An interaction effect of problem type and modality congruence was found, $F(1,30) = 8.63$, $p = 0.006$, $\eta_p^2 = 0.223$. A *post hoc* comparison revealed that if the probe-task modality was congruent to the problem modality, its performance became slower for insight problem solving, while it made no difference during non-insight problem solving. It is also notable that probe-task reaction time was significantly slower during non-insight problem solving, compared to insight problem solving only when the probe-task modality was non-congruent to the primary problem (Figure 8).

Modality Congruence × Problem Stage Interaction

No significant interaction of modality congruence × problem stage was found, $F(1.88,56.25) = 0.4$, $p = 0.657$, $\eta_p^2 = 0.01$. The probe-task temporal dynamic was approximately the same in both cases, when the problem modality was

congruent to the probe-task modality, and when it was not.

Problem Stage × Problem Type

A significant interaction effect of problem stage × problem type was found, $F(2,60) = 33.09$, $p < 0.001$, $\eta_p^2 = 0.524$. A *post hoc* comparison revealed that the probe-task reaction time was initially the same during the first stage for both insight and non-insight problems. However, in the middle stage the probe-task reaction time became significantly slower in non-insight problem solving. The magnitude of change further increased in the last stage. Each consecutive stage in non-insight (but not insight) problem solving featured a significant increase in probe-task reaction time (Figure 7).

No significant three-way interaction effect was found, $F(1.86,55.64) = 1.34$, $p = 0.269$, $\eta_p^2 = 0.043$.

Discussion

The results of the second experiment indicate that working memory systems are involved in insight and non-insight problem

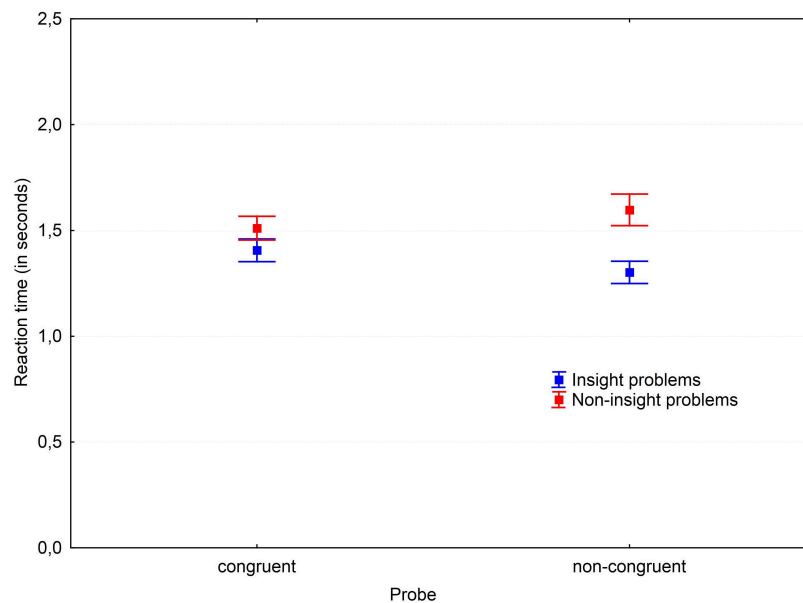


FIGURE 8 | Working memory load in congruent and non-congruent conditions. Vertical bars denote standard errors.

solving processes unequally. Whenever the probe-task and the primary problem were of the same modality, the resource demands were approximately the same (reflected by the same probe-task reaction time) in insight and non-insight problem solving processes. However, in cases when the probe-task and the primary problem were of different modalities – the probe-task during insight problem solving was performed faster than in non-insight problem solving. This leads to a conclusion that non-insight problem solving competes for general resources of working memory – the control system, since competing with the probe-task within the same storage system (phonological loop or visuo-spatial sketchpad) made no difference compared to when the primary problem and the probe-task were processed within separate storage systems. However, it made a substantial difference for insight problem solving – not having both the primary problem and the probe-task processed within the same system at the same time – significantly decreased the average reaction time, and, therefore, reflects better availability of resources in such cases. In other words, the general availability of the control system is more important for non-insight problem solving, while the availability of specific storage systems is more important for insight problem solving. The results suggest that the processing involved in a representation change in insight problem solving occurs on a level as low as the manipulations with the perceptual image of the visual information within the modal-specific storage systems. This falls in line with Duncker's (1945) ideas regarding insight mechanisms: the solver has to “re-see” the solution (to view the problem from a different angle). Similar findings regarding the importance of modal-specific components can be found in a number of studies which showed that insight problem solving relies on congruency with problem representation storage systems. For example, the nine-dots problem solving performance is positively associated with

visual working memory capacity (Chein et al., 2010); heavy visuo-spatial sketchpad load hinders the chess matches problem solving (Robbins et al., 1996); verbal insight problems are solved worse under the phonological loop load (Gilhooly and Murphy, 2005).

Within modality competition and cross-modality competition did not reveal different temporal dynamics over the course of the three stages of problem solving. It seems that although insight and non-insight problem solving processes are different in terms of what working memory components are more crucial for their processing; this difference is equally present during all the stages of the problem solving process. However, the stage-to-stage dynamics without regards to probe-task modality was different for insight and non-insight problem solving processes, replicating the results found in Experiment 1. We observed a gradual increase in the control system load in non-insight problem solving. This might represent the need to keep the results of the intermediate calculations in working memory, as well as the monitoring of the problem solving progress, and the necessity to hold rules and operators in memory. These factors are especially prevalent in non-insight problem solving, but are not as prominently present in insight problem solving because insight solutions mainly require a problem representation shift, which might be less working memory intensive because it does not require the accumulation of explicitly held pieces of information.

The temporal dynamics of working memory load across various stages of insight and non-insight problem solving processes were not affected by whether the probe-task and the primary problem were of the same modality or not. The first reason why this was the case lies in the homogeneity of the initial and final representations of the problem. The problems we used did not require participants to build a problem representation of a different modality in order to achieve the solution. The visual problems required participants to manipulate the visual problem

space, while verbal problems revolved around the semantics and the relation between the problem elements. Arguably, if in order to achieve the solution, participants had to switch the modality of the initial problem representation (e.g., verbal to visual), this would have been represented in the results; for example, the visual probe-task reaction time would increase after the initial verbal representation was changed to visual and vice versa. This hypothesis can be tested in future studies. For example, “symmetric problems” (Vladimirov et al., 2016) can be used to investigate this topic, since solving them requires participants to realize that the problem they are facing only appears to be a visual picture reconfiguration, while in reality the problem space represents signs and numbers. The methodological approach we developed (division of the problem into three equal time stages) would likely not be suitable to identify a singular event of the representation change since it is based on averaging a rather large portion of the problem solving session. We plan to supplement this approach by event-related measurements/grouping criteria as well. An impasse and an “aha” moment can serve as markers guiding our data analysis in the future. In particular, Jones (2003) proposed an eye-tracking procedure for identifying the impasse phase. They argue that the moment of the impasse gives way to a more than twofold increase in the fixation duration on certain elements of the problem compared to the average fixation duration prior to that. Identifying the moment of impasse would allow us to test whether the probe-task methodology is consistent with the eye-tracking data.

GENERAL DISCUSSION

In conclusion, we would like to note the technique we used to assess the dynamics of the solution. Despite the popular idea that an insight solution can be divided into various phases, empirical verification of this statement is hard to obtain. Our proposed technique allows one to uncover and probe different phases of the solution separate from each other. This approach lacks disadvantages commonly associated with participant self-reports or an individual differences approach such as: an inability to investigate the micro-dynamics of problem solving; invasiveness – alteration of the natural course of the problem solving process; as well as confound effects of metacognition and memory processes. The main disadvantages are the impossibility of recording the micro-dynamics of problem solving; invasiveness, i.e., influence on the course of the solving process; the low possibility of reflection; the general mechanics of the process; and the influence of metacognitive skills and memory processes in cases of self-reports. The probe-task can act as either a facilitator or a distractor of the problem solving process based on the experimental needs. Besides this, reaction time measurements typically provide a more robust and reliable effect that can benefit the research of working memory during the problem solving process.

It is worth noticing that the probe-task itself in Experiment 1 did not substantially increase the problem (both types) solution times. However, this was the case for Experiment 2 – both insight and non-insight problems were solved slower when performing

a dual-task. It is possible that this happened for the very same reason the effects obtained in Experiment 2 were more robust: the combined difficulty level of the problem and the probe-task were likely more appropriate (higher) in Experiment 2.

All in all, both experiments supported the notion that working memory is involved in insight problem solving. Every type of the probe-task used as the secondary task in insight problem solving revealed an increase of reaction time in the dual task condition compared to the single task performance, suggesting a fluctuating impact of the problem solving process on probe-task performance. Working memory in general is involved in both types of problem solving because they share some of the general activities involved in the solving process such as text comprehension, storage of problem elements, holding the interim calculations, attentional control of strategies, and heuristics. Both the control system and storage systems are involved in those general processes. However, the emphasis on either control system or storage systems is different in insight and non-insight problems. While non-insight problem solving is more demanding on the control system, insight problem solving seems to rely on the processing within the modal-specific storage systems to a greater extent. While working memory is typically viewed as a system involved in explicit processing, the fact that working memory (especially the storage systems) plays a role in insight problem solving (that features rather limited conscious self-awareness), supports the idea that working memory is crucial for implicit processing as well (Reber and Kotovsky, 1997; Baars and Franklin, 2003; Soto et al., 2011; Lebed and Korovkin, 2017). Overall, insight problem solving appears to be less demanding on working memory compared to non-insight problem solving, especially if the distinction between control system load and storage systems load is not accounted for.

In terms of the unique contribution of working memory systems, the results indicate that non-insight problems are more demanding on the control system. This could be the case because these problems typically involve more explicit processing, such as progress monitoring, implementation of heuristics, and operations within the problem space. Insight problem solving, on the contrary, involves rejection of the incorrect representations and ineffective rule-sets, which occurs only occasionally and does not require constant monitoring maintained by the control system. This differentiation between the working memory systems involvement was supported by the fact that the probe-task was performed more efficiently if it did not compete for same modality processing as the primary problem – but this was the case only for insight problem solving, not non-insight. Arguably, this notion supports the idea that insight restructuring relies on rather low-level processing that occurs within the working memory storage systems.

All the data regarding the temporal dynamics feature a similar pattern: gradual increase of working memory load in the non-insight problem solving process, but not in the insight problem solving process. This result is in line with our prediction that the solver exerts more and more effort associated with the control system as they progress toward the solution in non-insight problems. The insight problem solving dynamics results were somewhat ambiguous. Results obtained in Experiment 1 revealed

a significant increase in working memory load from phase to phase. The results on Experiment 2, however, reveal no such dynamics. Since the procedure in Experiment 2 was modified and participants were not required to perform the probe-task as they were verbally reporting their proposed solution is what might have caused these differences in the results. If this is the case, then the verbalization of the solution in insight problem solving might cooccur with some of the relevant processes contributing to the dynamics in Experiment 1. Such as when the verification of the proposed solution is pronounced verbally.

The lack of observable dynamics in insight problem solving does not speak in favor of the selective forgetting hypothesis (Simon, 1977; Ohlsson, 1992), according to which insight solution involves mere forgetting of the incorrect solutions; if that was the case, one might expect a decrease of working memory load after the incorrect solution was forgotten.

CONCLUSION

The proposed probe-tasks technique differs from the traditional distraction paradigm commonly employed in the field. This technique relies on the secondary probe-task reaction time over the course of problem solving, not the problem solution time itself. This paradigm is more suitable for research of working memory load in problem solving.

Insight problem solving is similar to non-insight analytical processing in terms of involvement of working memory resources. However, taking specific functions within working memory into consideration can reveal unique differences between the two problem solving types. Control systems and modal-specific storage systems play a rather different role in insight and non-insight problem solving processes. Insight problems appear to be less demanding on control systems while relying on the availability of modal-specific storage systems in working memory. The working memory demands seem to increase over the problem solving course for non-insight problems, but not for insight problems since they involve less cumulative explicit knowledge acquisition.

Even though identifying the key components involved in insight problem solving can tell us more about the nature of this phenomenon, the control system is crucial for the performance of almost every intellectual activity in humans, therefore, making it rather challenging to isolate its contribution to each problem type individually. Our claim of representational change in insight problem solving occurs within the modal storage systems, should and will be further tested in the future studies.

REFERENCES

- Ash, I. K., and Wiley, J. (2006). The nature of restructuring in insight: an individual-differences approach. *Psychon. Bull. Rev.* 13, 66–73. doi: 10.3758/BF03193814
- Baars, B. J., and Franklin, S. (2003). How conscious experience and working memory interact. *Trends Cogn. Sci.* 7, 166–172. doi: 10.1016/S1364-6613(03)00056-1
- Baddeley, A. D. (2002). “Fractionating the central executive,” in *Principles of Frontal Lobe Function*, eds D. T. Stuss and R. T. Knight (New York, NY: Oxford University Press), 246–260. doi: 10.1093/acprof:oso/9780195134971.003.0016
- 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. Reason.* 21, 76–96. doi: 10.1080/13546783.2014.934399
- Beilock, S. L., and DeCaro, M. S. (2007). From poor performance to success under stress: working memory, strategy selection, and mathematical problem solving under pressure. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 983–998. doi: 10.1037/0278-7393.33.6.983
- Chein, J. M., and Weisberg, R. W. (2014). Working memory and insight in verbal problems: analysis of compound remote

ETHICS STATEMENT

This study was approved by the Ethics Committee of the Psychology Department of the Yaroslavl State University. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

All authors together designed the experiments. AS conducted the first experiment. AC conducted the second experiment. SK and IV wrote the first draft of the manuscript and analyzed the data. All authors were critically involved in the interpretation of the results and in revising the manuscript.

FUNDING

This work was supported by the Russian Science Foundation 18-78-10103.

ACKNOWLEDGMENTS

The authors are grateful to Viktor Z. Gochiyayev for help in data processing. Alexandra Chistopolskaya is grateful to the Mikhail Prokhorov Fund for the opportunity to conduct this study in the Laboratory for Cognitive Studies at RANEPa within the postdoctoral fellowship under the Karamzin Scholarships program. Authors thank Anton Lebed for his significant help in manuscript preparation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01864/full#supplementary-material>

TABLE S1 | The list of insight and non-insight problems.

TABLE S2 | The results of Experiment 1.

TABLE S3 | The results of Experiment 2.

TABLE S4 | The solution time and solution rate data for both experiments.

- associates. *Mem. Nõgn.* 42, 67–83. doi: 10.3758/s13421-013-0343-4
- Chein, J. M., Weisberg, R. W., Streeter, N. L., and Kwok, S. (2010). Working memory and insight in the nine-dot problem. *Mem. Cogn.* 38, 883–892. doi: 10.3758/MC.38.7.883
- Chuderski, A. (2014). How well can storage capacity, executive control, and fluid reasoning explain insight problem solving. *Intelligence* 46, 258–270. doi: 10.1016/j.intell.2014.07.010
- Chuderski, A., and Jastrzębski, J. (2017). Working memory facilitates insight instead of hindering it: comment on DeCaro, Van Stockum, and Wieth (2016). *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 1993–2004. doi: 10.1037/xlm0000409
- Chuderski, A., and Jastrzębski, J. (2018). Much ado about aha!: insight problem solving is strongly related to working memory capacity and reasoning ability. *J. Exp. Psychol. Gen.* 147, 257–281. doi: 10.1037/xge0000378
- Cinan, S., and Doğan, A. (2013). Working memory, mental prospection, time orientation, and cognitive insight. *J. Individ. Differ.* 34, 159–169. doi: 10.1027/1614-0001/a000111
- Cowan, N. (2010). Multiple concurrent thoughts: the meaning and developmental neuropsychology of working memory. *Dev. Neuropsychol.* 35, 447–474. doi: 10.1080/87565641.2010.494985
- De Dreu, C. K. W., Nijstad, B. A., Baas, M., Wolsink, I., and Roskes, M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personal. Soc. Psychol. Bull.* 38, 656–669. doi: 10.1177/0146167211435795
- DeCaro, M. S., Van Stockum, C. A., and Wieth, M. B. (2016). When higher working memory capacity hinders insight. *J. Exp. Psychol. Learn. Mem. Cogn.* 42, 39–49. doi: 10.1037/xlm0000152
- DeCaro, M. S., Van Stockum, C. A., and Wieth, M. B. (2017). The relationship between working memory and insight depends on moderators: reply to Chuderski and Jastrzębski (2017). *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 2005–2010. doi: 10.1037/xlm0000460
- DeYoung, C. G., Flanders, J. L., and Peterson, J. B. (2008). Cognitive abilities involved in insight problem solving: an individual differences model. *Creat. Res. J.* 20, 278–290. doi: 10.1080/10400410802278719
- Duncker, K. (1945). On problem-solving. *Psychol. Monogr.* 58, i–113. doi: 10.1037/h0093599
- Ellis, J. J., Glaholt, M. G., and Reingold, E. M. (2011). Eye movements reveal solution knowledge prior to insight. *Conscious. Cogn.* 20, 768–776. doi: 10.1016/j.concog.2010.12.007
- Engle, R. W., Kane, M. J., and Tuholski, S. W. (1999). “Individual differences in working memory capacity and what they tell us about controlled attention, general fluid Intelligence, and functions of the prefrontal cortex,” in *Models of Working Memory*, eds A. Miyake and P. Shah (Cambridge: Cambridge University Press), 102–134. doi: 10.1017/CBO9781139174909.007
- Fleck, J. I. (2008). Working memory demands in insight versus analytic problem solving. *Eur. J. Cogn. Psychol.* 20, 139–176. doi: 10.1080/09541440601016954
- Gilhooly, K. J., and Fioratou, E. (2009). Executive functions in insight versus non-insight problem solving: an individual differences approach. *Think. Reason.* 15, 355–376. doi: 10.1080/13546780903178615
- Gilhooly, K. J., and Murphy, P. (2005). Differentiating insight from non-insight problems. *Think. Reason.* 11, 279–302. doi: 10.1080/13546780442000187
- Hambrick, D. Z., and Engle, R. W. (2003). “The role of working memory in problem solving,” in *The Psychology of Problem Solving*, eds J. E. Davidson and R. J. Sternberg (New York, NY: Cambridge University Press), 176–206. doi: 10.1017/CBO9780511615771.007
- Jarosz, A. F., Colflesh, G. J. H., and Wiley, J. (2012). Uncorking the muse: alcohol intoxication facilitates creative problem solving. *Conscious. Cogn.* 21, 487–493. doi: 10.1016/j.concog.2012.01.002
- Jones, G. (2003). Testing two cognitive theories of insight. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1017–1027. doi: 10.1037/0278-7393.29.5.1017
- Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Knoblich, G., Ohlsson, S., and Raney, G. E. (2001). An eye movement study of insight problem solving. *Mem. Cognit.* 29, 1000–1009. doi: 10.3758/BF03195762
- Korovkin, S., Vladimirov, I., and Savinova, A. (2014). The dynamics of working memory load in insight problem solving. *Russ. J. Cogn. Sci.* 1, 67–81.
- Lavric, A., Forstmeier, S., and Rippon, G. (2000). Differences in working memory involvement in analytical and creative tasks: an ERP study. *Cogn. Neurosci.* 11, 1613–1618. doi: 10.1097/00001756-200006050-00004
- Lebed, A., and Korovkin, S. (2017). The unconscious nature of insight: a dual-task paradigm investigation. *Psychol. Russ. State Art* 10, 107–119. doi: 10.11621/pir.2017.0307
- Lv, K. (2015). The involvement of working memory and inhibition functions in the different phases of insight problem solving. *Mem. Cognit.* 43, 709–722. doi: 10.3758/s13421-014-0498-7
- Murray, M. A., and Byrne, R. M. J. (2005). “Attention and working memory in insight problem-solving,” in *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, eds B. G. Bara, L. Barsalou, and M. Bucciarelli (Mahwah, NJ: Lawrence Erlbaum Associates), 1571–1575.
- Necka, E., Żak, P., and Gruszka, A. (2016). Insightful imagery is related to working memory updating. *Front. Psychol.* 7:137. doi: 10.3389/fpsyg.2016.00137
- Ohlsson, S. (1992). “Information-processing explanations of insight and related phenomena,” in *Advances in the psychology of thinking*, eds M. T. Keane and K. J. Gilhooly (New York, NY: Harvester-Wheatsheaf), 1–44.
- Ohlsson, S. (2011). *Deep Learning. How the Mind Overrides Experience*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511780295
- 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
- Pearce, J. W. (2008). Generating stimuli for neuroscience using psychopy. *Front. Neuroinform.* 2:10. doi: 10.3389/neuro.11.010.2008
- Reber, P. J., and Kotovsky, K. (1997). Implicit learning in problem solving: the role of working memory capacity. *J. Exp. Psychol. Gen.* 126, 178–203. doi: 10.1037/0096-3445.126.2.178
- Reverberi, C., Toraldo, A., D’Agostini, S., and Skrap, M. (2005). Better without (lateral) frontal cortex? Insight problems solved by frontal patients. *Brain* 128, 2882–2890. doi: 10.1093/brain/awh577
- Robbins, T. W., Anderson, E. J., Barker, D. R., Bradley, A. C., Fearnlyhough, C., Henson, R., et al. (1996). Working memory in chess. *Mem. Cognit.* 24, 83–93. doi: 10.3758/BF03197274
- Simon, H. A. (1977). “Scientific discovery and the psychology of problem solving,” in *Models of Discovery. And Other Topics in the Methods of Science*, eds R. Cohen and M. Wartofsky (Berlin: Springer), 286–303.
- Soto, D., Mäntylä, T., and Silvanto, J. (2011). Working memory without consciousness. *Curr. Biol.* 21, R912–R913. doi: 10.1016/j.cub.2011.09.049
- Suzuki, H., Fukuda, H., Miyata, H., and Tsuchiya, K. (2014). “Exploring the unconscious nature of insight using continuous flash suppression and a dual task,” in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, eds P. Bello, M. Guarini, M. McShane, and B. Scassellati (Québec: Cognitive Science Society), 2955–2960.
- Vladimirov, I., Chistopolskaia, A., Sekurtseva, J., and Lebed, A. (2016). “Representation change as a mechanism of insight problem solving. Study on ‘symmetrical problems,’” in *Proceedings of the 57th Annual Meeting of the Psychonomic Society* (Boston: Psychonomic Society), 120.
- Wieth, M. B., and Burns, B. D. (2014). Rewarding multitasking: negative effects of an incentive on problem solving under divided attention. *J. Probl. Solving* 7, 60–72. doi: 10.7771/1932-6246.1163
- Wiley, J., and Jarosz, A. F. (2012). “How working memory capacity affects problem solving,” in *Psychology of Learning and Motivation*, ed. B. H. Ross (San Diego, CA: Elsevier Inc), 185–227. doi: 10.1016/B978-0-12-394393-4.00006-6
- Yeh, Y., Tsai, J.-L., Hsu, W.-C., and Lin, C. F. (2014). A model of how working memory capacity influences insight problem solving in situations with multiple visual representations: an eye tracking analysis. *Think. Skills Creat.* 13, 153–167. doi: 10.1016/j.tsc.2014.04.003

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 Korovkin, Vladimirov, Chistopolskaya and Savinova. 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.



Paving the Way to Eureka—Introducing “Dira” as an Experimental Paradigm to Observe the Process of Creative Problem Solving

Frank Loesche^{1,2*}, Jeremy Goslin³ and Guido Bugmann²

¹ CogNovo, Cognition Institute, Plymouth University, Plymouth, United Kingdom, ² School of Computing, Electronics and Mathematics, Plymouth University, Plymouth, United Kingdom, ³ School of Psychology, Plymouth University, Plymouth, United Kingdom

OPEN ACCESS

Edited by:

Philip A. Fine,
University of Buckingham,
United Kingdom

Reviewed by:

Linden John Ball,
University of Central Lancashire,
United Kingdom
Darya L. Zabelina,
University of Arkansas, United States

*Correspondence:

Frank Loesche
frank.loesche@cognovo.eu

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 30 March 2018

Accepted: 03 September 2018

Published: 02 October 2018

Citation:

Loesche F, Goslin J and Bugmann G
(2018) Paving the Way to
Eureka—Introducing “Dira” as an
Experimental Paradigm to Observe
the Process of Creative Problem
Solving. *Front. Psychol.* 9:1773.
doi: 10.3389/fpsyg.2018.01773

“Dira” is a novel experimental paradigm to record combinations of behavioral and metacognitive measures for the creative process. This task allows assessing chronological and chronometric aspects of the creative process directly and without a detour through creative products or proxy phenomena. In a study with 124 participants we show that (a) people spend more time attending to selected vs. rejected potential solutions, (b) there is a clear connection between behavioral patterns and self-reported measures, (c) the reported intensity of Eureka experiences is a function of interaction time with potential solutions, and (d) experiences of emerging solutions can happen immediately after engaging with a problem, before participants explore all potential solutions. The conducted study exemplifies how “Dira” can be used as an instrument to narrow down the moment when solutions emerge. We conclude that the “Dira” experiment is paving the way to study the process, as opposed to the product, of creative problem solving.

Keywords: creative problem solving, divergent thinking, convergent thinking, behavioral experimental paradigm, chronometric temporal measures, insight, chronology

1. INTRODUCTION

Creativity (Runco and Acar, 2012), innovation (Amabile, 1988), and problem solving (Newell and Simon, 1972) have shaped human history, culture, and technology. Valued by today’s society for their contributions to education, recruiting, and employment (Cropley, 2016) they are also likely to play an essential role in our future society. Moreover, creativity, innovation, and problem solving are required to address the increasingly complex problems we are facing. A commonality between these phenomena is the aim of identifying novel and useful answers to more or less well-defined and ill-defined questions (Simon, 1973; Weisberg, 2006). Based on observations and reports from eminent scientists such as Helmholtz and Poincaré, Wallas (1926) famously suggested that the process of generating answers or creative products consists of several consecutive phases. Since then the exact structure and number of these stages are being debated (Amabile, 1983; Finke, 1996; Csikszentmihalyi, 2009; Amabile and Pratt, 2016), but arguably, the moment when a solution emerges lies at the heart of the matter. This “illumination” phase often follows and

precedes other stages (Howard et al., 2008): Before finding the solution, the problem solver needs to “prepare” for the problem at hand, for example by understanding the question, potentially within the larger context. If people do not solve the problem in this phase, they might enter a stage of “incubation.” In this stage, they are thought to unconsciously keep processing the problem while they consciously attend to other tasks. The feeling of manifesting associations or fringe consciousness coined as “intimation” is the next stage in this model (Sadler-Smith, 2015). Following this, the problem solvers experience a phase of “illumination” when they suddenly have an idea that answers the question. Afterwards, during the “verification” stage, this solution is tested. Certain models consider additional stages to communicate and implement a found solution as part of the process. Csikszentmihalyi (2009), for example, calls it the “elaboration” stage. To sum up, within existing case studies of creativity, innovation, and problem solving and the theories behind them, the moment when solutions emerge is part of a longer “creative process.” However, most studies focused on the outcome of these three phenomena, without considering the various processes behind them.

Previous studies identify the moment when solutions emerge through a range of different phenomena (Kounios and Beeman, 2014), for example restructuring the problem representation (Knoblich et al., 1999; Fleck and Weisberg, 2004), an alteration of mood (Baas et al., 2008; Subramaniam et al., 2009), and the suddenness of changes (Topolinski and Reber, 2010a). Reports of these potentially associated phenomena have been used as markers of “insights,” “Aha! moments,” and “Eureka experiences.” However, some of these phenomena might only be weak proxies. Danek et al. (2016) have shown that not every solved problem relies on restructuring. In a follow-up study, Danek and Wiley (2017) revealed that not every experience of insight results in a solved problem. Even if a link between observed phenomenon and “Eureka experience” is well established as for the mood change, the chronology or even causality remains unclear: Does insight increase mood (Akbari Chermahini and Hommel, 2012), does a stimulated positive mood cause “Aha! moments” (Isen et al., 1987; Ritter and Ferguson, 2017), or are they both results of another process? Therefore, there is a need to detect emerging solutions directly and not via proxy phenomena. Moreover, most studies on insight assume Eureka experiences are dichotomous, “Aha! moments” either suddenly happen or not (Bowden and Jung-Beeman, 2003; Gilhooly and Murphy, 2005; Subramaniam et al., 2009; Hedne et al., 2016). Possibly the phenomenon benefits from a more differential view, theoretically and empirically.

In this paper, we introduce “Dira” as a novel experimental paradigm to narrow down the moments of emerging solutions within the creative process. In each of the forty “Dira” tasks, participants are asked to find a solution. A solution is the image they consider to correspond best with a one-line text. On a computer display, the on-screen text and images appear blurred by default and can only be seen clearly when the mouse hovers above them (see **Figure 1**). Tracing the mouse movement and the hover time on each image allows to measure the time participants spend processing an image during task execution

and before they report a solution. After each task participants provide metacognitive self-reports, such as the intensity of their Eureka experience that accompanies emerging solutions (Cushen and Wiley, 2012; Danek et al., 2014). We hypothesize that the combination of behavioral measures of the process and self-reports can be used to identify distinctive behaviors when solutions emerge and localize the solutions’ emergence in time. Further, we hypothesize that feedback on the participants’ choice moderates the behavior and the reported Eureka experience thereafter.

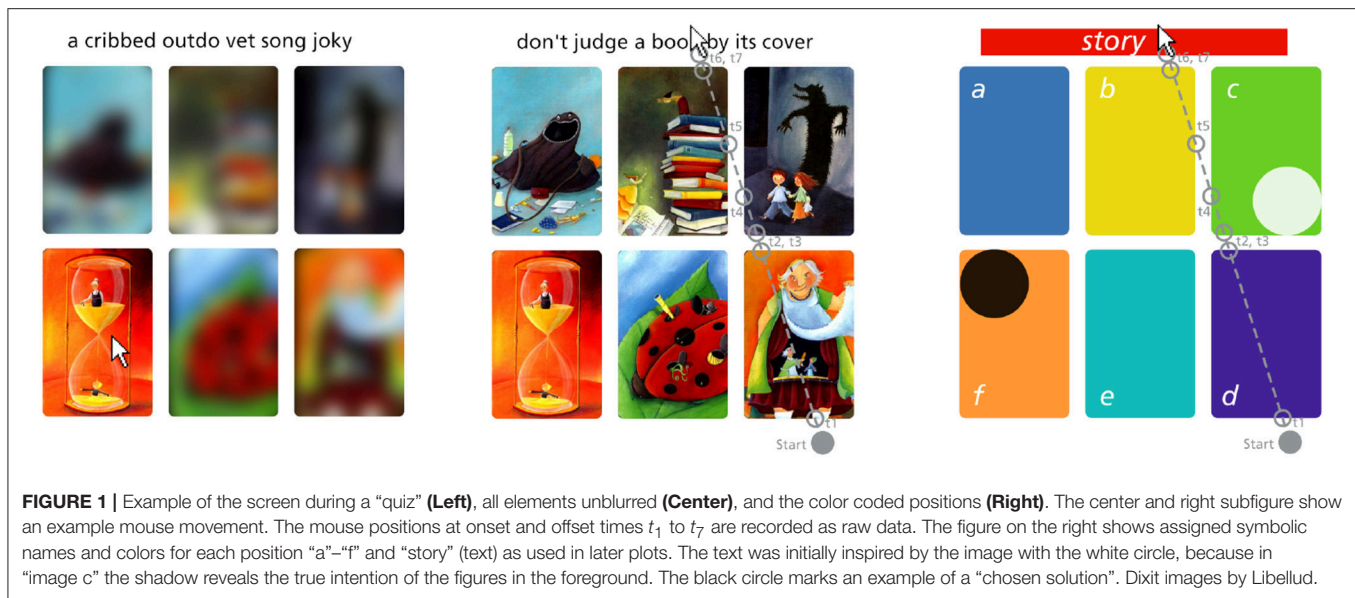
2. RATIONALE

In this section, we summarize existing tasks that have been used to observe the moment solutions emerge during creative problem solving and we provide an argument for a novel experimental paradigm. We describe the origin of “Dira” and how we acquired the problems participants are asked to solve. Finally, we argue for the mouse-tracking method to trace people’s problem solving process.

2.1. Existing Tasks Related to Emerging Solutions

Different types of tasks have traditionally been associated with the creative process and emerging solutions, namely insight tasks, divergent thinking tasks, and convergent thinking tasks.

From a historical perspective, insight tasks (Maier, 1930; Duncker, 1963; Gardner, 1978; MacGregor et al., 2001) are the oldest of these types of tasks. They predate the distinction between divergent and convergent production as introduced by Guilford (1967) and were consequently developed without a direct reference to one of these processes. These insight tasks often take the form of riddles or visual puzzles and are built around the assumption that the task itself requires restructuring (Knoblich et al., 1999; Fleck and Weisberg, 2004). The overlap between insight tasks and convergent thinking tasks seem particularly strong: for example, Bowden and Jung-Beeman (2003) argue, that convergent thinking problems like the Remote Associate Task share properties with insight tasks. Nevertheless, convergent thinking tasks can either be solved via insight or without. Similarly, classical insight problems are often thought to converge to a single solution, even though examples for the nine-dot problem show that more than one solution is possible (Maier, 1930; Sarcone, 2014). Furthermore, and as Bowden et al. (2005) and Danek et al. (2016) demonstrate, finding solutions to insight tasks does not require insight or an Aha experience. While timing has been discussed since the earliest studies on insight tasks, often it only relates to the time when a solution is found. These type of tasks are not repeatable and allow only between-subject comparisons. Even more, having solved similar problems in the past seems to influence the process (Lung and Dominowski, 1985), and it is difficult to identify the similarity between problems as well as to control for previous exposure. Consequently, the classic insight problems are not considered for this study.



Divergent thinking tasks (Torrance, 1966; Guilford, 1967; Runco et al., 2016), in which people are asked to generate several potential solutions to a question, are associated with individual creative processes. Nevertheless, the measurement of originality is usually assessed within the cohort of the experiment and not for an isolated individual. Consider a “Brick Uses” task (Wilson et al., 1954; Guilford, 1967, p. 143) in which participants are asked for alternative uses of a brick. An answer to use the brick’s pigments to paint might be unique within an experiment, but the participant might just have reported an instance from memory (Gilhooly et al., 2007; Hass, 2017). Hence this solution, although original within the experiment, did not require creative problem solving from this particular individual. Furthermore, before assessing the originality, raters decide if answers are considered for the scoring. For the answer “to paint” in a “Brick Uses” task, which is similar to the previous example, some would consider it an “impossible answer” and consequently remove the answer before scoring originality. Time measurements are often provided by a minimum or maximum task time and through fluency measures, and recently the moments of the production of a solution have received more attention (Forthmann et al., 2017). Divergent thinking tasks are in general repeatable, but the difficulty in scoring, and the unknown origin of the solution, either from memory or as a novel product, disqualify these types of tasks for our purpose.

Finally, Convergent thinking tasks (Mednick, 1962; Knoblich et al., 1999; Bowden and Jung-Beeman, 2003), require participants to come up with a single solution. These tasks are based on the difficulty to search a large problem space, produce interim solutions, and verify these results. Some of these tasks, such as the Compound Remote Associates test, were developed to specifically address the shortcomings of the classical insight tasks (Bowden and Jung-Beeman, 2003). Convergent thinking tasks typically provide a large number of stimuli for repeated measures. For word-based convergent

thinking problems, language fluency affects the ability to solve the problem (Hommel et al., 2011).

In our study, we intended to observe behavior during the creative process, but for problems with three verbal stimuli such as the Compound Remote Associate task, prospective problem solvers might not exhibit much observable behavior. The low number of word-based stimuli within a single task (typically three) are easy to memorize, and participants can operate entirely on their working memory. There is little incentive to reread the words or exhibit other behavioral cues through which the internal thought process could be traced. The timing of the solution and the success within a given time are central measurements in this type of task. For example, Salvi et al. (2016) ask their participants to press a button as soon as they found a solution. This timing relates only to the whole process but does not allow the identification of the involved sub-stages. Therefore we decided not to use convergent thinking tasks to trace the emerging solution within the creative problem solving process.

2.2. Development of “Dira”

“Dira” has been developed out of the necessity to collect fine-grained measurements of the creative process. As an experimental paradigm to observe the moment when solutions emerge, “Dira” needs to address one fundamental requirement: the solution should not be known from the beginning. In this sense, a solution could either be the answer itself or an algorithm how to arrive at the answer. If either was known at the moment the task was given, “Dira” would merely provide measures related to other processes, for example processing fluency and memory retrieval.

“Dira” is inspired by “Dixit,” a commercially available and internationally acclaimed card game. The word “Dixit” is Latin for “he or she said,” chosen by the French developers of the game, supposedly to highlight the story-telling aspect. We use the French word “Dira” for “he or she will point out” as a reference

to the process throughout the task as well as the origin of the inspiring game. The 84 unique images of a “Dixit” card deck are described as “artwork”¹ and “dreamlike”² and have previously been used in teaching a foreign language (Cimermanová, 2014), in research on imaginative design narratives (Berger and Pain, 2017), and observing conformity and trust between humans and robots (Salomons et al., 2018). The cards have also inspired interventions to foster creativity (Liapis et al., 2015), and are suggested as “an additional source of inspiration” (Wetzel et al., 2017, p. 206) for an ideation method.

The task “Dira” we developed uses elements and data from the game “Dixit.” Therefore, we briefly introduce some relevant aspects of the game. Three to six players can participate in the “Dixit” game, which is played in several rounds. At the beginning of a round, one of the players is appointed as the storyteller. From the deck of 84 unique cards with beautifully drawn images, each player receives six cards in their hand. Based on the drawing on one of the cards, the storyteller invents a short text and tells it to the other players. Related to this text, all other players select one card from their hand. The selected cards are shuffled and played on the table. Now all players except the storyteller have to guess which of the images originally inspired the text. Based on their choice, the storyteller and all other players receive points. Hereby the scoring system penalizes storyteller that produce descriptive texts and associations that are easy to find. Furthermore it encourages the others to play cards with a similar non-obvious connection to the text. Moreover, and based on the different associations the players formed, each image has some connection to the text. At the end of a round, a group of players has produced a combination of a short text and as many associated images as there are players. Nevertheless, and as the example in **Figure 1** illustrates, it would defy the purpose of the game if the other players would immediately understand any of these connections.

In each “Dira” task we ask people to find a connection between a short text and one of six images sampled from past “Dixit” games with six players. As argued before, people are unlikely to identify the image that inspired the text immediately. Instead, they might find a connection between the text and one of the six potential solutions through controlled processes in creative cognition (Beatty and Silvia, 2012; Silvia et al., 2013) or unconscious associations (Mednick, 1962; Kenett et al., 2014). In the first case, participants generate several metaphors or potential solutions from available information and select one of them as the best fit at a specific time. In the second case, existing associations are mediated through similarities of common elements before one of them is identified as the best match. In both cases, the solution emerges at a distinct moment before participants select one image by a mouse click. Participants in the “Dira” task are forced to make a choice, but which of the six possible solutions they choose depends on their prior knowledge and their subjective understanding of the task at hand. These differences

in problem difficulty are described for other problems as well. Often, the correctness of a task solution is considered vital to the measures and consequently needs to be controlled for, as Öllinger et al. (2014) demonstrate for a well know 9-dot problem. “Dira” does not have one objectively correct solution and we are not interested in the exact timing of finding the subjectively correct solution. Instead, we assess the behavior during the process through the interaction times with text and images.

For the developed task we assume that two different modalities for the stimuli are advantageous to isolate remote conceptual associations. If the two stimuli that were to be matched used the same modality, matches could be found for aspects of these stimuli that are outside the interest of this study. For example matches between two visual stimuli could not only be based on the depicted content, but also on colors, forms, and dynamics of the image. For two verbal stimuli the constructing syllables, cultural connotations, and language fluency of the problem solver would play a decisive role in the selection of an answer. By asking people to match content from different modalities, we hope to circumvent the issues above.

2.3. Dataset

The experience of an emerging solution relies on the inherent quality of the task; in the case of “Dira” on the text as well as on each of the potentially associated images. Instead of constructing a synthetic dataset, we crowdsourced the combination of a single text and six accompanying images from a community of experienced “Dixit” players. Usually, the card game “Dixit” is played locally around a table. For groups not sharing the same space, Boite-a-jeux³ provides an online gaming platform to play this game across distances and with other players of a similar skill level. In August 2014 we accessed the publicly available recorded game data of 115,213 rounds of “Dixit.” We filtered this initial dataset for English rounds with six players. After stopword removal (such as “the,” “is,” “at”) and word stemming, we removed the rounds with stories containing the most frequent words from the 90th percentile. Looking at the text and images, candidate sets for the “Dira” task were selected from the remaining 1,000 rounds of recorded “Dixit” games. The authors of this paper, two of which are experienced “Dixit” players, chose 40 combinations of text and images. Afterwards, we identified between one and three contexts of associated knowledge to control for participants’ domain-specific knowledge in a later analysis. For example, the sentence “Standing on the shoulders of giants” is meaningful in different domains like the scientific community exposed to life and work of Newton, but also for fans of the Britpop group “Oasis,” who released an album with the same name. The identified contexts were then grouped into the following eight clusters (with the number of associated stories in brackets): Literature (8), music (6), film (7), science (7), popular culture (12), and high culture (7) as well as word games (11), and literal interpretations of visual cues (10). These contexts allow to control for required knowledge to solve the tasks. Finally, the order of the tasks within the “Dira” experiment was initially

¹ Dixit publisher’s website <http://en.libellud.com/games/dixit>, last access: 2018-02-23.

² Wikipedia: Dixit (card game) [https://en.wikipedia.org/w/index.php?title=Dixit_\(card_game\)&oldid=823435686](https://en.wikipedia.org/w/index.php?title=Dixit_(card_game)&oldid=823435686), last access: 2018-04-05.

³ <http://boiteajeux.net>; last access 2017-11-15.

chosen at random but kept the same throughout all conditions reported in this paper.

2.4. Mouse-Tracking as Process-Tracing

“Dira” is based on the fundamental assumption that psychological processes can be traced through observable behavior (Skinner, 1984). Of particular interest to the emerging solutions is the participants’ behavior during the task when they are engaged in a creative problem solving process. At the beginning of each task, participants do not know the text or the images. To solve the problem, they have to acquire information from these elements and find associations between the text and the images. For “Dira” the process of information acquisition is related to the order and timing of interactions with each of the elements on the “quiz” screen. Different methods are commonly used to trace these chronology and chronometric measures of processes, for example through verbal protocols (Newell and Simon, 1972), eye-tracking (Thomas and Lleras, 2007), and mouse-tracking (Freeman and Ambady, 2010).

Verbal and think-aloud protocols have been used in insight tasks (Fleck and Weisberg, 2004), divergent thinking tasks (Gilhooly et al., 2007), convergent thinking tasks (Cranford and Moss, 2012), and also in real-world problem solving (Newell and Simon, 1972; Kozbelt et al., 2015). While Schooler et al. (1993) identified an overshadowing effect for insight problem solving, Gilhooly et al. (2007) did not find any effect on fluency and novelty production in a divergent thinking task. In a meta-study, Fox et al. (2011) did not see an effect of verbalization on the results of tasks, but they noted an increase in the time required. These results suggest that think-aloud protocols might or might not change the solutions provided for a task, but they most certainly change the process. With our interest in narrowing down the time of emerging solutions within a process, verbal protocols seemed too invasive and were disregarded.

In a direct comparison between eye-tracking and mouse-tracking, Lohse and Johnson (1996, p. 37) conclude that mouse interactions “predispose people to use a more systematic search and process more information than they normally would.” Similar to the technique described by Ullrich et al. (2003), elements in the “quiz” of “Dira” that are not directly under the mouse pointer are blurred. These indistinct images prevent participants from accessing this information without moving the mouse pointer to an element. A notable difference to the method developed by Ullrich et al. (2003) is that elements in “Dira” do not fade over time; elements are visible for the whole time the mouse pointer hovers over them. Uncovered images imply that information acquisition and information processing is possible throughout the whole hover time. Indeed, participants will not necessarily direct their full attention to the currently unblurred text or image. While this appears as a disadvantage of mouse-tracking, Ferreira et al. (2008) have observed the same issue for eye-tracking. People are also known to not always perceive visual input when generating ideas (Walcher et al., 2017). Furthermore, other processes such as memory access are related to eye movements as well (Johansson and Johansson, 2013; Scholz et al., 2015). Nevertheless, Freeman and Ambady (2010) have shown that mouse-tracking provides reliable insight

into mental processes and while it provides more robust measures than eye-tracking, it is also easier to administer. Mouse-tracking was chosen as the process-tracing method for the “Dira” task, also because it allows running several studies in parallel in a non-invasive setup using standard hardware participants are familiar with.

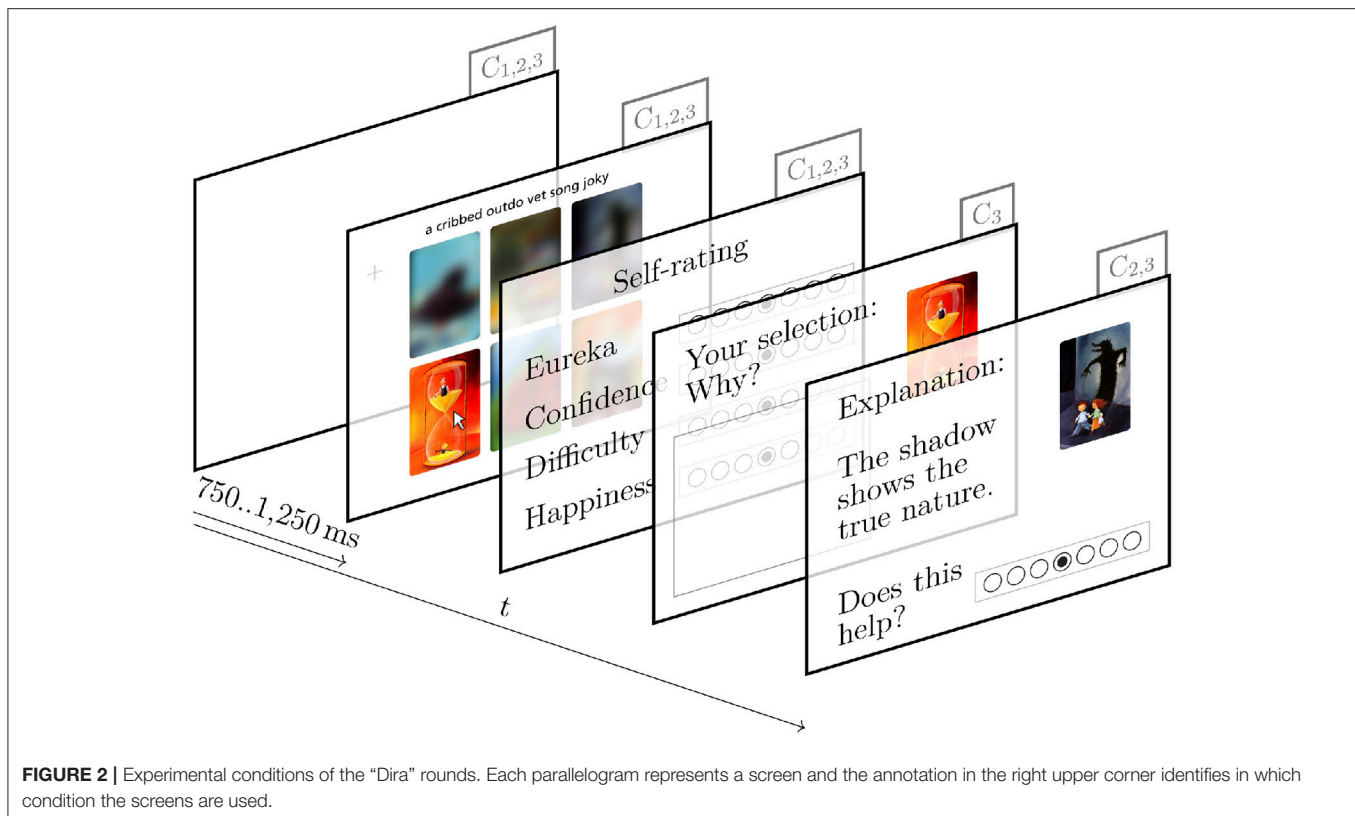
3. METHODS

3.1. Experimental Design and Conditions

The computer-based experiment “Dira” is programmed as a series of different screens. From the participants’ perspective, “Dira” combines perceived freedom to explore the task with aesthetically pleasant stimuli. Participants interact with the text and images of the task by hovering the mouse pointer over these elements. The order and duration of these interactions are up to the prospective problem solvers. The images are taken from the “Dixit” card game which has been praised for its artistic and beautiful drawings. Moreover, the whole experiment is designed like a game. These design choices are intended to make the “Dira” tasks “inherently interesting or enjoyable,” one of the critical elements that are known to increase intrinsic motivation in participants (Ryan and Deci, 2000, p. 55). In turn, Baas et al. (2008) and da Costa et al. (2015) have shown positive correlations between intrinsic motivation and performance in creative problem solving tasks.

For the current study, “Dira” was administered in three different between-subject conditions. In condition 1 “Dira” does not provide any feedback and participants have no reference to evaluate their answers and performance in the task. In condition 2 we added a potential solution to trigger extrinsic insights. Given that tasks are often perceived as difficult, this demonstrates a possible solution to the participants and hence is thought to increase the motivation to solve the next problem. Furthermore, these solutions have the potential of triggering extrinsic insights, which are a special type of insight following the recent argument by Rothmaler et al. (2017). Given the correlation between mood and insight (Subramaniam et al., 2009; Akbari Chermahini and Hommel, 2012) a triggered Eureka experience could have a positive effect on the intrinsic motivation and metacognition. In condition 2 we want to explore if this leads to a change in the reported experience and observed behavior. In condition 3 we ask participants to elaborate on their reported solution. We expect this verbalization of an answer to increase the metacognitive awareness during task execution (Hedne et al., 2016) and hence an effect on “quiz time” and reported Eureka experience. Condition 1 was the first to be run and all participants at the time followed the same protocol. Subsequent participants at a later time were randomly assigned to either condition 2 or condition 3.

In condition 2 the additional screen “explanation” is added to each round as illustrated in **Figure 2**. Appended after the “rating,” it is the last screen before the start of the next round. The “explanation” screen shows the “intended solution,” the image that initially inspired the storyteller to invent the text. We also show a short explanation on how the intended solution and text are connected. The short sentence is based on a text taken from



the stimulus dataset and is designed to help the participants: One method to solve a “Dira” task is to empathize with the storyteller and find the intended solution that initially inspired the text. To assess the success of this help, we then ask the participants to rate “How much does the Explanation help [you] to understand the association between image and text?” Their answer ranges from “not at all” to “very much” on a seven-point Likert item. Submitting the answer starts the next round of condition 2 with a “fixation cross.”

In condition 3 an “elaboration” screen is placed between the “rating” and the “explanation” screen as shown in **Figure 2**. In this screen, participants see the given text and their selected image, and they are asked to elaborate on their decision. Afterwards, they see the same “explanation” screen as described above. Once they have completed these additional screens, participants restart the next “round” of condition 3 with a “fixation cross.”

3.2. Procedure

Any “Dira” experiment starts with an opening sequence consisting of a “welcome” screen, a “questionnaire,” and a “description” of the task. This initial series is followed by 40 rounds containing a “fixation cross,” “quiz,” “rating,” and optional “explanation” or “elaboration” screens. The experiment concludes with an on-screen “debrief.”

A “welcome” screen explains the basic idea of the study as well as potential risks and the right to withdraw data. The study only continues if participants understand and agree to the minimum

requirements that have been cleared by the Faculty of Health and Human Sciences Ethics Committee at Plymouth University. Once participants have given their consent, they are shown the “questionnaire.”

During the “questionnaire” participants are asked to specify their age, gender and primary language and if they have participated in the study “Dira” before. They are also asked to rate their fluency in understanding written English and familiarity with the card game “Dixit” on a seven-point Likert item. Participants are also asked to rate themselves in 14 additional seven-point Likert item questions, four of which belong to the Subjective Happiness Scale (SHS) developed by Lyubomirsky and Lepper (1999) and ten more of the Curiosity and Exploration Inventory II (CEI-II) as published by Kashdan et al. (2009). The scales were chosen because emotional states (Baas et al., 2008), openness to experience, and intrinsic motivation (Eccles and Wigfield, 2002) are known to influence problem solving (Beaty et al., 2014). These results are not discussed here since the interaction between individual differences and the performance in the “Dira” task are beyond the scope of the current article.

Once participants have completed the questionnaire, the procedure of the experiment is explained to them in detail in a “description” screen. This screen also holds a minimal and neo-Gestalt inspired definition of the “Eureka moment” as “the common human experience of suddenly understanding a previously incomprehensible problem or concept,” for accessibility reasons taken from Wikipedia (2016). Afterwards, the 40 “rounds” of the experiment begin.

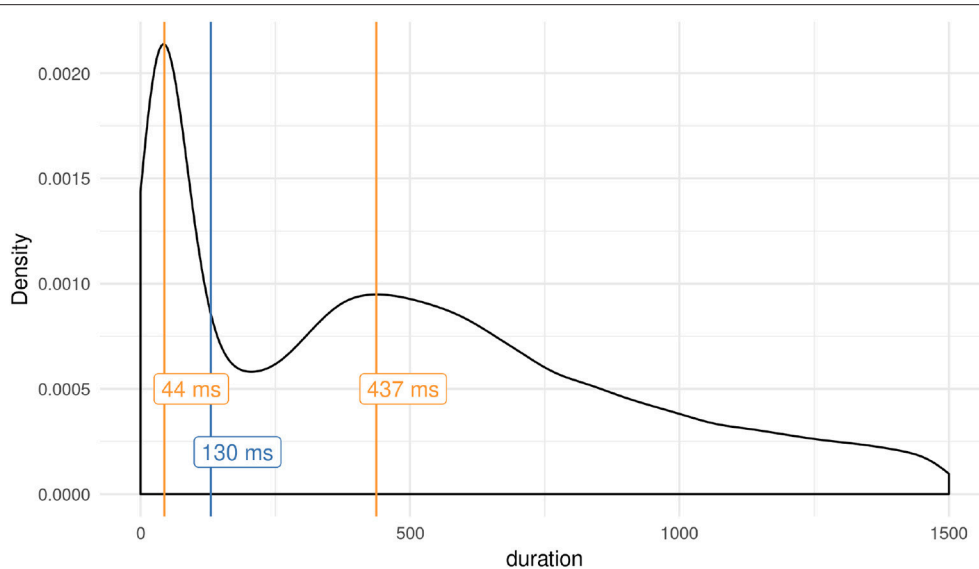


FIGURE 3 | Distribution of hovering times on elements during the “quiz.” The modes of the bimodal distribution are marked with red lines. The cutoff time between the two distributions, a result of the classification described in the text, is shown in blue.

Each “round” starts with a “fixation cross” which is shown at the center of the screen for a randomized time between 750 and 1,250 ms. Afterwards text and images appear on the “quiz” screen as illustrated in **Figure 1**: one text on top and six images in a grid of two rows by three columns. Unless the participants hover the mouse on top of these elements, the letters of the text are shown in a randomized order, and the images are strongly blurred. An example can be seen in the second screen of **Figure 2** which shows the text “Don’t judge a book by its cover” with the letters in a randomized order and images blurred except for “image f” over which the mouse pointer hovers. The recording of hover times during the “quiz” allows to track when participants pay attention to each of the elements and for how long (Navalpakkam and Churchill, 2012). On this screen, participants attempt to find the image that they think is most likely associated with the text and select it through a single click. There is no time limit for completing this task. Once participants have chosen a solution, they advance to the “rating” screen.

During the “rating” screen, participants are asked to rate their performance in the “quiz.” They are asked the following four questions, with the range of possible responses on seven-point Likert items in brackets: “How confident are you that the solution is right?” (not confident—very confident), “How hard was it for you to come up with the solution?” (not hard—very hard), “How strong did you experience a Eureka moment?” (not at all—very strong), and “How happy are you with your answer?” (very unhappy—very happy). After submitting the answers, the next round starts with a “fixation cross.”

Participants who have completed the 40 rounds conclude their participation with the “debrief” screen. Here they are informed that the study intended to measure the timing of their behavior during the “quiz.” Participants are encouraged to give additional feedback concerning the experiment, and they have the option

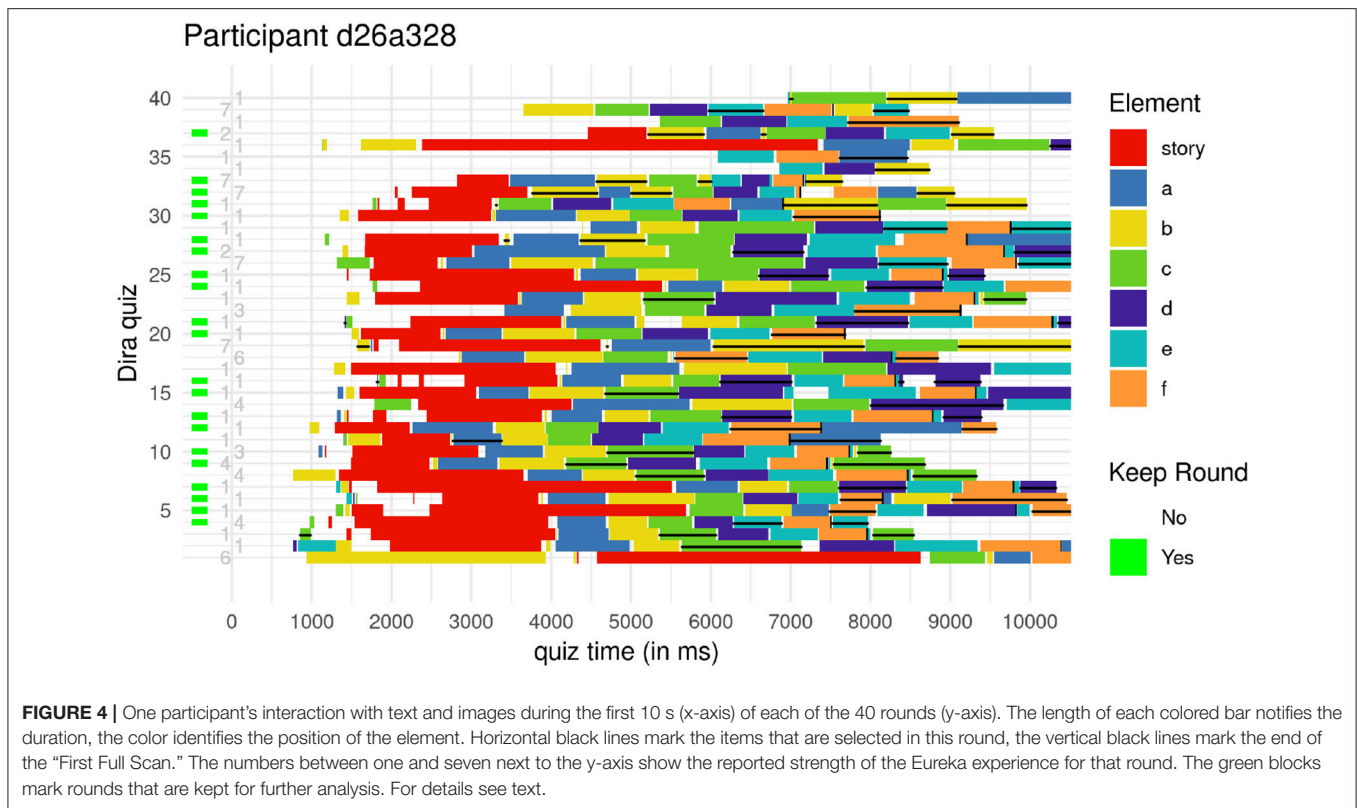
to leave an email address in case they want to be informed of the results of the study. This on-screen debrief was followed by a short unstructured personal discussion relating to their experience in the Dira experiment.

3.3. Task Administration

The controlled study “Dira” was designed as a computer-based task administered in a laboratory setup. The task was delivered through a custom developed web application delivered through a full-screen web browser. The same type of computer mouse with an optical sensor and the same type of 22 inch LCD screen with $1,920 \times 1,080$ pixel resolution were used for the whole experiment. Participants are most likely familiar with the setup as it is the same hardware available to students in library and public computing spaces across campus. The experiment was delivered in a dedicated room with no more than five participants at the same time who were asked to stay silent during the experiment. Welcome and debrief was performed outside the room to keep any distraction to a minimum. Informed consent was collected from participants; then they were accommodated at a computer showing a “welcome” screen.

3.4. Participants

One hundred and twenty-four participants between the age of 18 and 56 ($\bar{age} = 22.6$, $sd = 6.99$) were recruited from a local pool of pre-registered psychology students and a second pool that was open to students of other courses and members of the public. While two of the participants chose not to report their gender, 83 identified as female and 39 as male. Psychology students received course credits and points for running their studies. Participants from the second pool received monetary compensation. The overall sample appears similar to the one described by Henrich et al. (2010).



3.5. Data Pre-processing

The data collected during the "quiz" of the "Dira" task are intended to trace the participants' thought process through their behavior. The recorded dataset includes chronological information concerning the order in which participants engage with elements, as well as the duration of the interactions.

The chronology or order in which participants engage with elements shows that they do not interact with all elements in each round. If participants do not look at the text, this has implications on their ability to solve the problem: Participants who have not seen the text will not be able to find an association between the text and one of the images for this particular round. On the other hand, if they have seen the text but not all images, they are still able to find a solution. Rounds in which participants did not look at the text were therefore excluded from further analysis, whereas rounds with missing interactions for some images were still analyzed. Furthermore, cognitive processes deployed in rounds that start with the text might differ from the ones starting with one of the images. To control for these different modalities, we focus in this paper on the rounds starting with text and remove all others.

The duration of interactions with text and images is assumed to relate to the amount of acquired and processed information. However, the data also include quick movements that do not contribute to acquiring information, as illustrated in **Figure 3**. If people want to look at an element not adjacent to the current mouse position, they need to move the pointer across one or more elements. In this case, the distance of the mouse pointer from the target image is between 1.5 times and 4.3 times the

size of the target. According to Fitts' law, the task of moving to a distant image has an index of difficulty between 1.3 and 2.4. Applying the extreme values for throughput suggested in Soukoreff and MacKenzie (2004), participants are estimated to require between 260 and 640 ms for the whole distance and therefore between 150 and 170 ms to cross an image between the starting position and the target image. During this movement, the element is briefly unblurred on screen. **Figure 4** shows examples of this movement at the beginning of rounds 4–7. The density of the duration of interactions in **Figure 3** shows how often participants interact with elements for certain durations. The bimodal distribution suggests that there are at least two different types of behavior recorded. Shorter interactions, in **Figure 3** marked as the local maxima around 44 ms, are distinctly different from longer hover times peaking around 437 ms. A cluster model fitted to the log-transformed duration using two components (Scrucca et al., 2016) classifies 17,849 interactions as short and 63,452 as long, divided at 130 ms. The predicted movement time according to Fitts' law and the identified time dividing the bimodal distribution of hover times suggest that the shorter engagements with elements might be movements across the element, targeting another one. If participants follow the mouse movement and see the intermediately unblurred image on screen during the shorter engagement, the following unblurred target image acts as a backward mask. Previous research does not provide evidence for perceptual discrimination between visual stimuli shown for less than 100 ms (VanRullen and Thorpe, 2001; Zoefel and VanRullen, 2017). Furthermore, Salti et al. (2015) argue for a required exposure of more than 250 ms

necessary to consciously perceive a stimulus. Assuming that specific information from a higher conceptual level is required to identify remote associations in the “Dira” task, these activations would require additional time, as Quiroga et al. (2008) have shown in single neuron recordings. For the “Dira” experiment we are interested in interactions for which participants can distinguish between different images. Concluding the different cited streams of research we assume that shorter interactions from the bimodal distribution shown in **Figure 3** have no or little influence on the process “Dira” intends to capture. In accordance with Fitts’ law, we assume that the shorter observed behavior represents mouse movements across elements moving for a different target without cognitive processing of the image. Consequently, element interactions below the identified 130 ms are excluded from further analysis.

4. RESULTS

We first report on the type of raw behavioral data collected during the “quiz” and derived measures such as the chronology of information acquisition. Secondly we present the self-reported measures collected during the “rating” screen. We then show that the number of interactions with elements relates to the reported strength of the Eureka experience. Finally, we report results of the length of different interactions in comparison to the reported strength of reported Eureka experience. For the statistical tests we adopted a critical α level of 0.01 as originally put forward by Melton (1962) and Trafimow et al. (2018). For each test where the estimated amount of false discoveries surpasses this threshold, we transparently report this value as suggested by Lakens et al. (2018). We adopt this practice for our study and the chosen traditional threshold, in particular since the discussion on statistical testing is far from over (Benjamin et al., 2017; Trafimow et al., 2018).

4.1. Available Process-Tracing Measures

Participants’ interaction with elements on the “quiz” screen is a metric for tracing their problem solving process. The time to produce solutions has previously been used in convergent

thinking tasks (Salvi et al., 2016) and divergent thinking tasks (Forthmann et al., 2017), a measure that is similar to the “quiz time” in this paper. “Dira” employs a novel method by collecting behavioral data, namely the interaction times with the stimuli, throughout the creative process. This is a novel approach by shifting the focus from measuring the duration to produce a “creative product” to providing chronological measures of the process itself. While the current paper focuses on the moment solutions emerge, the experimental paradigm could be used to trace other aspects of the creative process such as preparing for the task or the verification of solutions. Since the extracted behavioral measures are vital for understanding the subsequent writing, we elaborate on the raw data and their derived measures in this section.

To illustrate the kind of data collected in “Dira,” we will now discuss in detail **Figure 4**. The duration of interaction with each element is the difference between offset and onset time which is the raw data recorded during the task. **Figure 4** shows the example of one participant’s interaction within the first 10 seconds of each of the 40 rounds. Each of the colored bars represents a timespan during which the mouse pointer hovers on top of an element. The length represents the duration, and the color signifies with which element the participants interact. For example, in the first round on the bottom of **Figure 4**, this particular participant spent a long time on “image b” (for color and naming scheme see **Figure 1**). The second round instead starts with three short interactions with “image d,” “image e,” and “image b” followed by a short time without any element interaction before hovering on top of the “text” for almost two seconds. Some rounds, like the third one, are finished within the ten second period shown in **Figure 4**, others like the first two continued for a more extended period.

Figure 4 also shows additional data that is available in “Dira.” We refer to the moment participants select their solution as the “quiz time” since it ends the current “quiz.” This measure is similar to existing measures in other tasks, such as the total time to solve convergent thinking tasks as reported by Salvi et al. (2016) or to produce utterances for divergent thinking tasks (Forthmann et al., 2017). The example participant selects

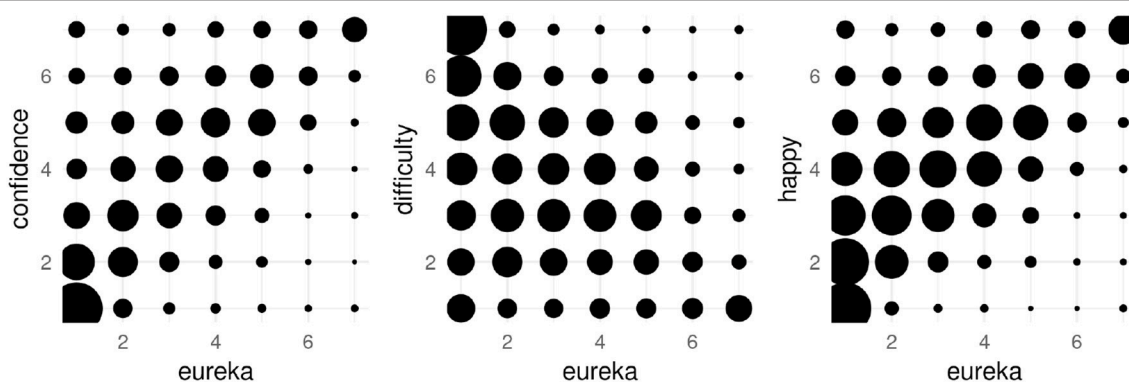


FIGURE 5 | Confidence, perceived task difficulty, and happiness related to the reported strength of the Eureka experience. The size of the circle represents the number of rounds in which the combination was reported, larger circles representing more answers.

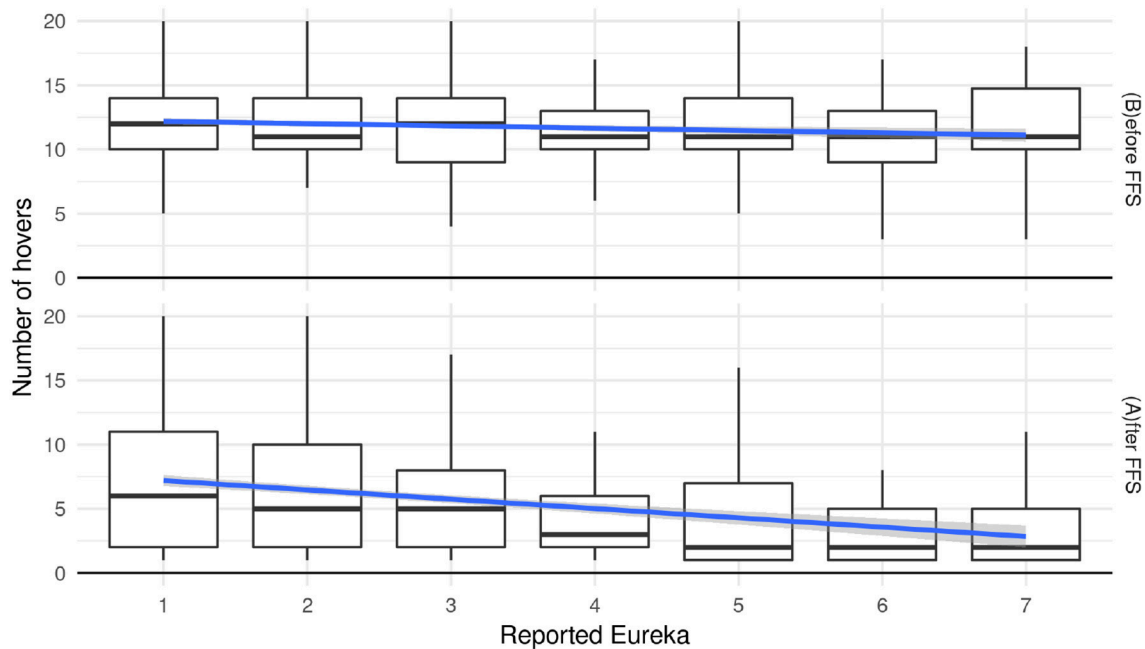


FIGURE 6 | Number of hovers before (Top row) and after (Bottom row) “First Full Scan” over the reported strength of Eureka experience.

the solution for round 3 at around 8,500 ms and round 4 at around 8,000 ms. The selected solution, for example, “image c” for round 3, is also indicated as a horizontal black line for the rounds in **Figure 4**. The vertical black line marks the end of what we call the “First Full Scan,” the end of the interaction with the seventh unique element. Participants have interacted with each element at least once at the end of the “First Full Scan.” The number next to the vertical axis in **Figure 4** represents the strength of the Eureka moment participants indicate during the “rating” screen. The example participant had no Eureka experience in round 2 and 3, but a strong one in round 19 and 26. Finally, the green box next to the vertical axis indicates rounds that are part of the analysis and not filtered out for one of the reasons explicated previously.

We administered “Dira” in three different conditions with a between-subject design as introduced in section 3.1. Based on the previously provided argument we hypothesized a longer interaction time for conditions 2 and 3. To test this, we built two linear mixed-effects models. Firstly we used the length of the First Full Scan as a dependent variable with the participant and round of the experiment as a random effect. We found no evidence for a difference between the three conditions ($\chi^2(2) = 2.4$, $p = 0.3$). In a second model, we used the quiz time as the dependent variable as it is most similar to the task time used in other tasks (Salvi et al., 2016; Forthmann et al., 2017). With participant and round of the experiment as random effects, we found no evidence that would support an effect of the experimental condition on time to report a solution ($\chi^2(2) = 0.87$, $p = 0.65$). Without support for the effect of the experimental conditions, there is no argument to distinguish between the three conditions regarding behavioral data.

4.2. Available Self-Reported Measures

Participants in the “Dira” task are required to provide self-reported measures in addition to the implicit behavioral data collected during the “quiz.” During the “reporting” screen they are asked to account for the strength of their just encountered Eureka experience, their confidence in the given solution, the perceived difficulty of the task, and their current happiness on seven-point Likert items respectively. Besides, participants in condition 2 and 3 are also asked to rate how well they understand the connection between the text and a potential solution. In condition 3 they are furthermore asked to write down how their solution is associated with the text. These measures are collected during each of the 40 rounds. In section 3.1 we hypothesized an increase in the reported Eureka experience for condition 3. Nevertheless, this is not supported by the collected data ($\chi^2(2) = 4.81$, $p = 0.09$). Consequently, we cannot maintain a separate analysis for the self-reports in the three conditions.

As illustrated in **Figure 5**, for rounds in which participants report a strong Eureka experience they are also confident regarding their solution. Rounds with weaker or no Eureka experience are reported across the whole spectrum of confidence, but with a tendency toward low confidence as well. Instead, rounds with strong Eureka experiences are rarely rated as low confidence. This asymmetry leads to an overall Spearman’s rank correlation of $\rho = 0.62$, $p < 0.01$. In contrast, rounds with strong reported Eureka experiences are rarely rated as “hard to come up with a solution.” Rounds with a low or no Eureka experience are perceived with varying difficulty. The overall correlation between the reported Eureka experience and stated task difficulty is $\rho = -0.41$, $p < 0.01$. Finally, for weak or no perceived Eureka, participants express a range of

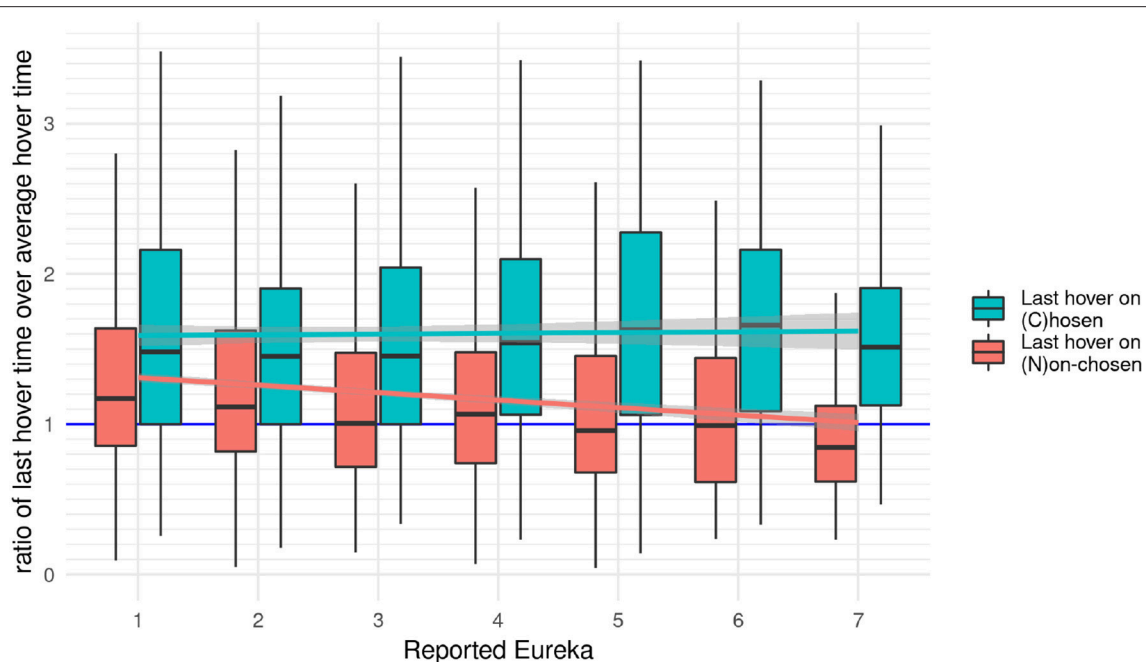


FIGURE 7 | The ratio between the last hover time within the First Full Scan and the average of all other hover times in the “First Full Scan,” separated by rounds in which participants hover over the chosen image last vs. the ones they look at another picture. A value of one means that they are equal, lower than one means the last scan is shorter than the previous ones. In addition to the box-whisker plot (showing the median and distribution), the lines show a linear model fitted to the mean ratio and surrounded by the 95% confidence interval in light gray.

different happiness, but only high happiness for strong Eureka experiences. Reported Eureka and happiness are correlated by $\rho = 0.6$, $p < 0.01$. The reliability of the rating is either good for reported Eureka ($\alpha = 0.86$) and difficulty ($\alpha = 0.87$), or acceptable for happiness ($\alpha = 0.78$) and confidence ($\alpha = 0.77$) based on Cronbach's alpha. Conceptually these four measures are linked by the literature review of Topolinski and Reber (2010a), who discuss the relationship between ease, positive affect, and confidence to insight. This link is reflected by the data collected in “Dira” with good reliability suggested by Cronbach's $\alpha = 0.86$ across the four measures. Consequently, these findings confirm our second hypothesis that participants can report their experience on more than a binary scale.

4.3. Number of Interactions

In this section, we take a first look at the relationship between the self-reported intensity of the Eureka experience and the chronology extracted from the behavioral data. For example, when participants acquire information during the “quiz” and they find a solution, they might stop looking at more images. Therefore we hypothesize that the Eureka experience is stronger for rounds with fewer interactions. **Figure 6** shows how many elements a participant interacts with during each of the 40 rounds of the “Dira” experiment. The sub-figure on the top shows the number of interactions during the “First Full Scan” before participants have seen each element at least once. An average of ten to twelve interactions means that participants tend to go back and forth between elements even before they have seen all seven elements. More specifically, if participants look at elements in a

certain order, looking back at one element and then continuing with the round can result in two additional interactions. To give an example: one participant has looked at “image a” and “image b” and then goes back to “image a” before continuing with “image b,” “image c,” and “image d.” In this case, the participant had interacted twice with “image a” and “image b” during the “First Full Scan.” This particular round would have accounted for at least nine interactions before the end of the “First Full Scan.” To arrive at the numbers shown in **Figure 6**, this seems to happen twice in a typical “First Full Scan.”

To test the above hypothesis, we built an ordinal mixed-effects model (Christensen, 2015) with reported Eureka as a dependent variable. The number of interactions, the classification into before and after “First Full Scan,” and the experimental conditions were used as predictors. The rounds of the experiment as well as participants were considered as random effects. Results from this model indicate that there is a significant negative effect (estimate = -0.06 , $z = -6.27$, $p < 0.01$) of numbers of hovers on the reported Eureka before the end of the “First Full Scan.” The model also shows a significant negative effect (estimate = -0.35 , $z = -3.68$, $p < 0.01$) for the number of interactions after the end of the “First Full Scan.” This confirms our hypothesis for the interactions during and after the “First Full Scan.” On the other hand, there is no evidence that condition 2 or 3 have an effect compared to participants in condition 1 (estimates = $[-0.12, -0.28]$, $z = [-0.35, -0.88]$, $p = [0.73, 0.38]$).

During the “First Full Scan,” the above model shows a significant effect of the number of interactions with elements on the strength of the Eureka experience. Across all conditions, this

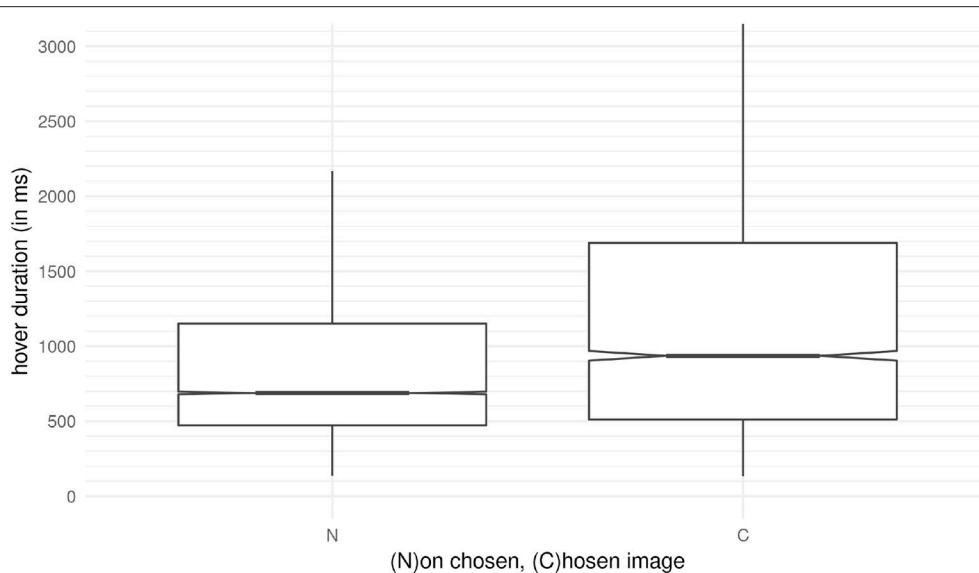


FIGURE 8 | The hover duration on the images within the First Full Scan. The time on the (C)hosen picture is longer the time on the five other images that are (N)ot chosen.

difference is between 12.61 interactions for no or low Eureka experiences and 11.38 interactions for strong reported Eureka. After the “First Full Scan” participants do not interact with all the images and text, again. The significant effect of the number of interactions on the reported strength of Eureka is higher this time and more pronounced in **Figure 6**: the difference is between 9.65 interactions for no experience of a Eureka and 4.24 interactions for a strong one. There is no evidence for an effect of the experimental condition on these results. Considering that the behavior of participants with different Eureka experiences seems to change before the end of the “First Full Scan,” it is of interest to examine the behavior during the “First Full Scan” in more detail. Hereafter we will examine whether the duration of hovering over elements provides additional information.

4.4. Last Hover During First Full Scan

Here we report the results for the hover duration on the seventh unique element. It is the last image during the “First Full Scan” and the first time participants interact with this specific element. Following up on the previous finding of an interesting difference between interactions during and after the “First Full Scan,” we want to narrow down the time of emerging solutions by exploring this specific hover time. More specifically we show the ratio of the duration on the last image compared to the mean of previous interactions. The chronometrical measure of hover time is illustrated in **Figure 7**. To correct for individual differences in processing speed, we plot the ratio of the hover time on the last image and the average hover times on all other images during the “First Full Scan.” **Figure 7** plots separately the ratio of rounds in which this element is the one (C)hosen later in the experiment and rounds which end on a (N)on-chosen one.

Figure 7 shows two effects: Firstly, for the “First Full Scans” ending on a chosen image, the median of the hover time is

roughly 50% higher on that element than for non-chosen ones (1,323 vs. 855.9 ms). Secondly, less time seems to be spent on the last non-chosen image than on the previous ones for stronger Eureka experiences, whereas more time is spent on the last image for low Eureka values. To quantify these effects we built an ordinal mixed-effect regression model with the strength of the reported Eureka experience as a dependent variable and the ratio, the type of element for the last hover, and the experimental condition as predictors. The round of the experiment and the participant were used as random effects. This model shows a significant effect of the ratio on the strength of the reported Eureka (estimate = -0.24 , $z = -6.1$, $p < 0.01$). It further shows a significant effect for rounds in which the last element is the chosen one on the strength of the reported Eureka (estimate = 0.2 , $z = 2.71$, $p < 0.01$). There is no evidence for the ratio in condition 2 or 3 affecting the reported Eureka intensity (estimate = $[-0.09, -0.55]$, $z = [-0.32, -0.32]$, $p = [0.75, 0.05]$).

The negative slope of the ratio over the strength of Eureka, in **Figure 7** particularly evident for the last hover on the non-chosen image, suggests that a solution has emerged before the end of the “First Full Scan.” The change of the ratio is either the result of a decrease of the numerator, an increase of the denominator, or a combination of both. The numerator decreases if participants spent less time on the last image when having a stronger Eureka experience. The denominator represents the average time spent on all previous images. It increases if participants spend more time on at least one of the previous images. If participants had Eureka experiences while looking at the image they are going to choose later, and this would be associated with them looking longer at that image, this would increase the denominator in the rounds which end on the non-chosen images. The observed increase would also explain the difference between rounds that

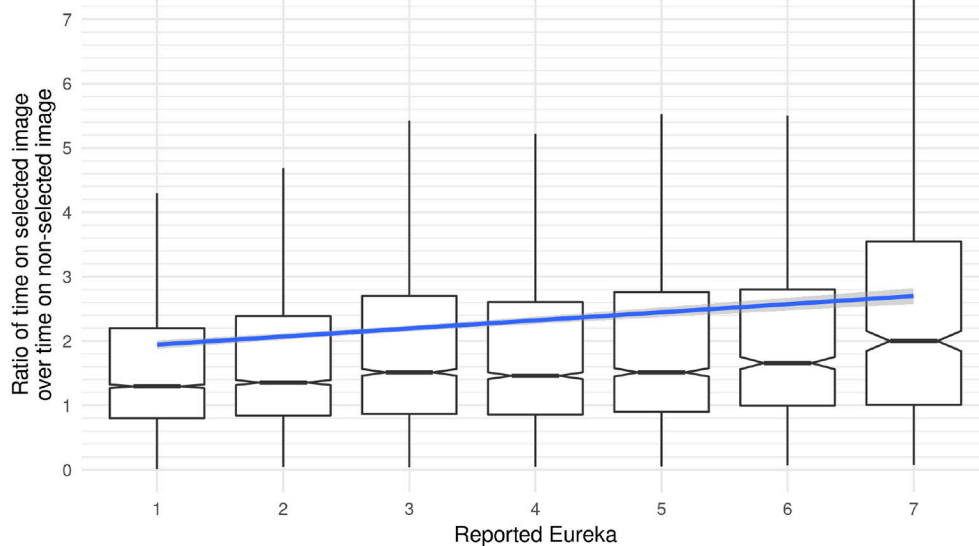


FIGURE 9 | The ratio of time spent on the chosen image over non-chosen images as a box-whisker plot with a linear model fit to the mean. The median for the ratio depicted in the box-whisker plot shows that participants spend nearly 1.3 times as much time on the chosen image compared to the others for a low Eureka, but about twice the time for a strong Eureka. The difference between mean (linear model) and median (box-whisker) results from outliers in the data.

end on chosen and non-chosen images. If participants spent less time on subsequent images, for example after a Eureka experience, this would decrease the numerator for the rounds ending on non-chosen images, but not for the ones ending on the chosen images. This interpretation of the observations suggests that the measured ratio is a compound of chronological effects and hover duration. Therefore we focus now on the duration spent on the chosen image and its relation to the strength of Eureka.

4.5. Chosen Images and Length of Interactions

The observation of the ratio of interaction times during the “First Full Scan” suggests that the interaction times between chosen and non-chosen images differ. Instead of a compound measure, we purely show the duration of hover times during the “First Full Scan” on (C)hosen and (N)on-chosen images in **Figure 8**. A Mann-Whitney test indicates that the duration of viewing chosen images (duration = 935.9 ms) is significantly longer than for non-chosen pictures (duration = 687.8 ms), $U = 20,873,370$, $p < 0.01$). Furthermore, there is a significant difference between the three conditions regarding the hover duration on non-chosen images ($H = 42.07$), $p < 0.01$ $Md_{\text{Condition 1}} = 663.2$, $Md_{\text{Condition 2}} = 679.7$, $Md_{\text{Condition 3}} = 727.9$), according to a Kruskal-Wallis test. Furthermore, there is a difference between conditions for the chosen images ($H = 9.18$, $p = 0.01$, $Md_{\text{Condition 1}} = 879.8$, $Md_{\text{Condition 2}} = 915.9$, $Md_{\text{Condition 3}} = 1,048$). Participants spend a significantly longer time on the chosen image in the third condition than in the other two conditions, and more time in the second condition compared to the first one.

We now look at the link between hover duration and reported Eureka experience in more detail. We built an ordinal regression model with the reported strength of the Eureka experience as the dependent variable. With the hovering time on the chosen images as a predictor, we failed to find evidence for a link between the strength of the Eureka and interaction time (estimate = 0.01, $z = 0.21$, $p = 0.83$). This is not unexpected since the raw data include slower and faster participants. Instead, if an ordinal mixed-effects model considers the participant as a random effect, the evidence supporting the link between hover duration and Eureka experience surpasses the threshold (estimate = 0.14, $z = 3.16$, $p < 0.01$). From this example we conclude that the recorded raw hover durations with text and images have little validity in connection with the self-reported measures collected during the “rating” screen. To address this, we remove the influence of participants and the task by considering the ratio between the time spent on chosen and non-chosen images calculated separately for each round. This suggested ratio between interaction times for a single round and with a single participant does not include chronological components related to the order of interactions; it is between measured times only.

Figure 9 shows the ratio between the hover duration on the chosen image and the average time spent on the other images. This ratio is higher for rounds in which participants report a stronger Eureka experience. An ordinal mixed-effects model fitted to the data supports this observation. The model uses the strength of the reported Eureka experience as a dependent variable and the ratio between the time spent on the selected image compared to the average duration on all other images as well as the experimental condition as a predictor. The round of the “Dira” task and the participant are used as random variables.

This model confirms that an increase in the ratio corresponds to a stronger Eureka experience (estimate = 0.02, $z = 5.65$, $p < 0.01$). With a ratio of 1.3 for no Eureka and 2 for a strong Eureka, participants seem to spend approximately 50% more time on the chosen image in rounds when they report a strong Eureka experience. However, the model does not provide evidence for an influence of condition 2 or 3 on the reported Eureka (estimates = $[-0.1, -0.58]$, $z = [-0.33, -2]$, $p = [0.74, 0.05]$).

Here we have presented two main findings. Firstly, the observations of the length of interaction with elements show that participants spend more time on the images they will select later in the task. Secondly, for rounds with a strong reported Eureka experience, the time spent on the chosen image is significantly longer than in rounds with a weaker or no Eureka experience.

5. DISCUSSION

The moment when a solution to a problem emerges is an extraordinary experience. It causes people to cry out “Eureka” (Pollio, 1914), “Aha” (Bühler, 1908), or “Uh-oh” (Hill and Kemp, 2016) and often their mood increases. In this paper, we suggest “Dira” as a novel experimental paradigm to observe these moments as part of the creative process. Many previous studies rely on the judgement of creative products, persons, or press (Rhodes, 1961)—or use proxy phenomena to assess the process contributing to creativity, innovation, and problem solving. In this study, we tested 124 people who participated in a controlled lab experiment designed to study the emergence of solutions. “Dira” records behavioral data during each task to observe the creative process directly. Specifically, we determine the chronology and chronometric measures of participants’ interaction with potential solutions. After each task, we ask the participants to self-report their experience on four different items. Here we discuss the implications of combined behavioral and metacognitive measures in the “Dira” task.

5.1. Eureka Experiences in “Dira”

Results from the behavioral data within the “First Full Scan” of “Dira” show that participants spend longer times on images they are going to select as their solution. Moreover, the length of the interaction on these chosen images is linked to the strength of the reported Eureka experience, with longer hover durations associated with stronger Eureka experiences. As shown in section 4.4, the median interaction time on the chosen image is about 50% longer than on the non-chosen ones. Another result related to the strength of Eureka is reported in section 4.5. For rounds that evoke a strong Eureka experience, participants spend about 50% more time hovering on the chosen image as compared to rounds with no or low reported Eureka. The current analysis does not allow drawing any conclusions regarding causality. Future studies could test if more extended engagement yields stronger Eureka experiences or if stronger Eureka experiences lead to longer hover durations.

After participants have interacted with the chosen image, they are less likely to continue looking for more elements

according to the results in section 4.3. Supposedly participants continuously scan the elements on the screen for a solution. If they find an association, the number of elements they interact with afterwards is related to the strength of the Eureka experience reported later. The significant effect can be observed as early as during the “First Full Scan” and the initial interaction with the images. These results suggest that something distinctive might already be happening during the initial engagement with the images.

With support from the ordinal mixed-effects model considering behavioral and self-reported measures, we confirm our first hypothesis that behavior happening during the “quiz” results in the reported intensity of Eureka. It would seem natural that the Eureka experience also happens during this time. However, it is not impossible that the Eureka experience is the result of a post-event evaluation. In any case, due to the short quiz time, these experiences would qualify as immediate insights according to Cranford and Moss (2012). In their study of convergent thinking, they found a difference between solutions found through a “classical insight” sequence and “immediate insights.” The immediate insights only consisted of an “Aha!” or Eureka experience and were considerably faster. This quick insight is also in line with the idea of intrapersonal creativity or mini-c introduced by Beghetto and Kaufman (2007). It would be interesting to design a modified version of “Dira” to elicit non-immediate insights as well, for example by tapping into the thought suppression as used in the delayed incubation paradigm (Gilhooly et al., 2014) or more generally in “little-c” type of tasks. We leave this speculation for future studies.

5.2. Subjective Experience

In more detail, the strong Eureka experience in rounds with high confidence is consistent with previous findings, for example by Hedne et al. (2016). In their study on magic tricks, problems solved via insight were rated with higher confidence than problems solved without insight. Previously Danek et al. (2014) had assessed a higher confidence rating for insight solutions as well, but they had used confidence in the definition of insight given to the participants, so this could have been a potential confound in their results. Hedne et al. (2016) also explicitly link confidence with the correctness of the solution, and Steele et al. (2018) highlight that confidence predicts a creative outcome. Further support comes from Topolinski and Reber (2010b) and Salvi et al. (2016) who identified a higher probability to be correct for insight solutions in convergent thinking tasks.

Happiness and, more generally, a positive mood is strongly linked to insights and Eureka experiences in the existing literature. In the “Dira” task participants experiencing a strong Eureka seldom report low happiness, but instead are consistently happier than with weaker or no Eureka experiences. The meta-review of Baas et al. (2008) provides a comprehensive overview of the relationship between mood and insight. More recently Shen et al. (2015) explore 98 different emotional states and their relationship to “Aha!” experiences. Results from their studies 2 and 3 suggest a link between insight and happiness—along with a list of other positive emotional states.

The mapping of states in two dimensions affords that other emotions could mask happiness for weaker Eureka experiences. While Abdel-Khalek (2006) finds single-item measurements of happiness sufficient to assess related positive affects and emotions, the fine-grained exploration of the emotional space associated with emerging solutions could be a topic for future research.

Our results for the relationship between difficulty and Eureka show that “Dira” tasks with a strong Eureka experience are rarely perceived as difficult. This finding seems counter-intuitive from the perspective of the classical “insight sequence” (Ohlsson, 1992) in which a complicated impasse has to be navigated. However, perceived difficulty can change in hindsight. Even if the task appears to be problematic while working on it, Topolinski and Reber (2010a) have shown that having an insight can change this. In a review of the literature, they identify a change of processing fluency as a result of having an insight. After having found the solution, they conclude, the problem appears to be easier than it was during the attempt to solve it. Alternatively, yet another interpretation is that the participants experience insights in tasks that are not difficult for them.

5.3. Differences Between Conditions and Personalities

In section 3.1 we provide a theoretical argument for administering “Dira” in the three different conditions. In particular, we hypothesized providing a potential solution would result in an increased interaction time. The collected data do not support this hypothesis as the results in section 4.1 show. We had further assumed that the additional task of elaborating on the chosen solution would increase the interaction time and change the self-report. As section 4.2 demonstrates, the data do not provide evidence for this effect. This could either mean that the theoretical argument is not sound and additional variables would influence the measurements to an extent that masks the hypothesized effect. Furthermore, the introduced interventions might tap into different effects than expected. Assuming that the theoretical argument is valid, the effect size could be too small or “Dira” as an instrument not sensitive enough to measure the effect within the sample. In summary, there is no evidence that supports a difference between the behavioral or self-reported measures among the three conditions.

In a trial-by-trial comparison, we reveal a link between fewer interactions and stronger Eureka experiences. In section 4.3 we compare the differences in the number of interactions observed between Eureka intensities, separately during and after the “First Full Scan.” We observe a significantly larger variance between no and strong Eureka experiences after the “First Full Scan.” This difference implies that the experience is influenced by element interactions and not by the participants’ distinctive approach to the task. On the other hand, individual variability might moderate the experience and performance in the “Dira” experiment. Future research could expand the method we suggest to address the relationship with personality traits.

Specifically, “Dira” could be used to test if traits known to correlate with creative production (Batey et al., 2010) predict eureka experiences.

5.4. Experimental Control

The participants’ freedom to choose the order and duration of stimulus interaction is supposed to increase task engagement, but it does not come without costs. The flexibility to look at elements in any order allows participants in the “Dira” experiment to not look at elements necessary to solve the problem. For example, some participants choose not to look at the text before selecting one of the images. Furthermore, participants who start with the text and try to find a matching image afterwards might use a different approach to solve the problem than others who engage with images first and interact with the text later during the task. In the first case, they only need to store the text itself or a derived concept in working memory to match it against each of the images they look at. In the second case instead, they need to remember up to six images and related concepts to match each of them with the text. In the current study, we filtered for rounds in which participants started with the text and removed all others. Future studies could eliminate the second case by specifying the chronology, for example by showing the text first.

As discussed earlier, the bimodal distribution of hover durations suggests that participants unblur elements for at least two different reasons. As discussed in section 3.5, participants might either intend to move the mouse pointer across by targeting elements on the other side or consciously engage with the text and images. In the current study, we assumed interactions shorter than 130 ms to represent mouse movement across elements. While these interactions were removed *post-hoc* from the current study, avoiding short unblurring could be implemented in the experimental design. The elements could only be shown clearly if the hover time exceeds the movement time predicted by Fitts’ law (Soukoreff and MacKenzie, 2004).

6. CONCLUSION

In the “Dira” task, we estimate the moment of the emerging solution based on the participants’ behavior and self-reports without relying on additional indicators. Like in many design and engineering problems, more than one solution is correct for this task. For “Dira” we demonstrate how behavioral data and meta-cognitive monitoring are integrated by this instrument to identify sub-processes of the creative process.

The results suggest that participants can distinguish between Eureka experiences of different strengths. Thus, our results suggest that Eureka experiences are not limited to having or not having an insight, but that the perception of this experience can have different intensity levels. Future studies should keep this in mind when assessing Eureka experiences.

Looking at the whole process of finding a solution to an ill-defined problem, people experience something early in the problem solving process that they relate to the Eureka experience. While the exact timing remains unclear, observations in “Dira” help narrowing down insight and other sub-processes. For example, before seeing all the elements in the “Dira” task,

participants in our study exhibit distinctive behavior related to the strength of their reported Eureka experience. Our results suggest that immediate insights exist and can be reported by people who experience them.

The creative process is often studied indirectly through the creative product, person, or press. We propose “Dira” as an experimental platform to record behavior as Eureka experiences are happening. This instrument and future studies applying the same underlying principle can bring us another step closer to understanding the creative process.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Plymouth University Research Ethics Policy, University Research Ethics and Integrity Committee. The protocol was approved by the Faculty Psychology Research

Ethics Committee. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

FL: design of experiment, implementation of the experiment, data collection, data analysis and interpretation, write up; JG: design of experiment; GB: design of experiment, data analysis and interpretation, paper structure and edit.

ACKNOWLEDGMENTS

The authors want to thank Dr. Katharine Willis for her insight that creative tasks, for example in Architecture, have more than one solution. This work was supported by CogNovo, a project funded by the University of Plymouth and the EU Marie Skłodowska-Curie programme (FP7-PEOPLE-2013-ITN-604764).

REFERENCES

- Abdel-Khalek, A. M. (2006). Measuring happiness with a single-item scale. *Soc. Behav. Pers.* 34, 139–150. doi: 10.2224/sbp.2006.34.2.139
- Akbari Chermahini, S., and Hommel, B. (2012). Creative mood swings: divergent and convergent thinking affect mood in opposite ways. *Psychol. Res.* 76, 634–640. doi: 10.1007/s00426-011-0358-z
- Amabile, T. M. (1983). The social psychology of creativity: a componential conceptualization. *J. Pers. Soc. Psychol.* 45, 357–376. doi: 10.1037/0022-3514.45.2.357
- Amabile, T. M. (1988). *A Model of Creativity and Innovation in Organizations*. (Berkeley, CA: Elsevier) 123–167. Available online at: <https://www.elsevier.com/books/research-in-organizational-behavior/staw/978-0-89232-748-5>
- Amabile, T. M., and Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: making progress, making meaning. *Res. Organ. Behav.* 36, 157–183. doi: 10.1016/j.riob.2016.10.001
- Baas, M., Dreu, C. K. W. D., and Nijstad, B. A. (2008). A meta-analysis of 25 years of mood–creativity research: hedonic tone, activation, or regulatory focus? *Psychol. Bull.* 134, 779–806. doi: 10.1037/a0012815
- Batey, M., Chamorro-Premuzic, T., and Furnham, A. (2010). Individual differences in ideational behavior: can the big five and psychometric intelligence predict creativity scores? *Creat. Res. J.* 22:90–97. doi: 10.1080/10400410903579627
- Beaty, R. E., Nusbaum, E. C., and Silvia, P. J. (2014). Does insight problem solving predict real-world creativity? *Psychol. Aesthet. Creat. Arts* 8, 287–292. doi: 10.1037/a0035727
- Beaty, R. E., and Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychol. Aesthet. Creat. Arts* 6, 309–319. doi: 10.1037/a0029171
- Beghetto, R. A., and Kaufman, J. C. (2007). Toward a broader conception of creativity: a case for “mini-c” creativity. *Psychol. Aesthet. Creat. Arts* 1, 73–79. doi: 10.1037/1931-3896.1.2.73
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., et al. (2017). Redefine statistical significance. *Nat. Hum. Behav.* 2, 6–10. doi: 10.1038/s41562-017-0189-z
- Berger, E., and Pain, F. (2017). Model and mobilise imaginary for innovative experience design. *Design J.* 20(Suppl. 1):S4690–S4696. doi: 10.1080/14606925.2017.1352967
- Bowden, E. M., and Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behav. Res. Methods Instrum. Comput.* 35, 634–639. doi: 10.3758/BF03195543
- 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
- Bühler, K. (1908). Tatsachen und probleme zu einer psychologie der denkvorgänge. *Arch. Psychol.* 12, 1–23.
- Christensen, R. H. B. (2015). *Ordinal—Regression Models for Ordinal Data*. R package version 2015.6-28. Available online at: <https://cran.r-project.org/package=ordinal>
- Cimermanová, I. (2014). Graphic novels in foreign language teaching. *J. Lang. Cult. Educ.* 2, 85–94. Available online at: <https://www.slovakedu.com/issues/>
- Cranford, E. A., and Moss, J. (2012). Is insight always the same? a protocol analysis of insight in compound remote associate problems. *J. Problem Solving* 4, 128–153. doi: 10.7771/1932-6246.1129
- Cropley, D. H. (2016). “Nurturing creativity in the engineering classroom,” in *Nurturing Creativity in the Classroom*, chapter Creativity is Vital to Engineering, eds R. A. Beghetto and J. C. Kaufman (Cambridge: Cambridge University Press), 212–226.
- Csikszentmihalyi, M. (2009). *Creativity: Flow and the Psychology of Discovery and Invention*. Harper Perennial Modern Classics. New York, NY: HarperCollins e-books.
- Cushen, P. J., and Wiley, J. (2012). Cues to solution, restructuring patterns, and reports of insight in creative problem solving. *Conscious. Cogn.* 21, 1166–1175. doi: 10.1016/j.concog.2012.03.013
- da Costa, S., Páez, D., Sánchez, F., Garaigordobil, M., and Gondim, S. (2015). Personal factors of creativity: a second order meta-analysis. *J. Work Organ. Psychol.* 31, 165–173. doi: 10.1016/j.rpto.2015.06.002
- Danek, A. H., Fraps, T., von Müller, A., Grothe, B., and Öllinger, M. (2014). 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. (2014). 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
- Danek, A. H., Wiley, J., and Öllinger, M. (2016). Solving classical insight problems without aha! experience: 9 dot, 8 coin, and matchstick arithmetic problems. *J. Prob. Solving* 9, 47–57. doi: 10.7771/1932-6246.1183
- Duncker, K. (1963). *Zur Psychologie des Produktiven Denkens, 2nd Edn*. Berlin; Heidelberg: Springer.

- Eccles, J. S., and Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annu. Rev. Psychol.* 53, 109–132. doi: 10.1146/annurev.psych.53.100901.135153
- Ferreira, F., Apel, J., and Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends Cogn. Sci.* 12, 405–410. doi: 10.1016/j.tics.2008.07.007
- Finke, R. A. (1996). Imagery, creativity, and emergent structure. *Conscious. Cogn.* 5, 381–393. doi: 10.1006/ccog.1996.0024
- Fleck, J. I., and Weisberg, R. W. (2004). The use of verbal protocols as data: an analysis of insight in the candle problem. *Mem. Cogn.* 32, 990–1006. doi: 10.3758/BF03196876
- Forthmann, B., Holling, H., Çelik, P., Storme, M., and Lubart, T. (2017). Typing speed as a confounding variable and the measurement of quality in divergent thinking. *Creat. Res. J.* 29, 257–269. doi: 10.1080/10400419.2017.1360059
- Fox, M. C., Ericsson, K. A., and Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychol. Bull.* 137, 316–344. doi: 10.1037/a0021663
- Freeman, J. B., and Ambady, N. (2010). MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behav. Res. Methods* 42, 226–241. doi: 10.3758/BRM.42.1.226
- Gardner, M. (1978). *Aha! Insight*. New York, NY; San Francisco, CA: Scientific American.
- Gilhooly, K. J., Fioratou, E., Anthony, S. H., and Wynn, V. (2007). Divergent thinking: strategies and executive involvement in generating novel uses for familiar objects. *Brit. J. Psychol.* 98, 611–625. doi: 10.1111/j.2044-8295.2007.tb00467.x
- Gilhooly, K. J., Georgiou, G. J., Sirota, M., and Paphiti-Galeano, A. (2014). Incubation and suppression processes in creative problem solving. *Think. Reason.* 21, 130–146. doi: 10.1080/13546783.2014.953581
- Gilhooly, K. J., and Murphy, P. (2005). Differentiating insight from non-insight problems. *Think. Reason.* 11, 279–302. doi: 10.1080/13546780442000187
- Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York, NY: McGraw-Hill.
- Hass, R. W. (2017). Semantic search during divergent thinking. *Cognition* 166, 344–357. doi: 10.1016/j.cognition.2017.05.039
- Hedne, M. R., Norman, E., and Metcalfe, J. (2016). Intuitive feelings of warmth and confidence in insight and noninsight problem solving of magic tricks. *Front. Psychol.* 7:1314. doi: 10.3389/fpsyg.2016.01314
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X
- Hill, G., and Kemp, S. M. (2016). Uh-oh! what have we missed? a qualitative investigation into everyday insight experience. *J. Creat. Behav.* 52, 201–211. doi: 10.1002/jocb.142
- Hommel, B., Colzato, L. S., Fischer, R., and Christoffels, I. K. (2011). Bilingualism and creativity: benefits in convergent thinking come with losses in divergent thinking. *Front. Psychol.* 2:273. doi: 10.3389/fpsyg.2011.00273
- Howard, T. J., Culley, S. J., and Dekoninck, E. (2008). Describing the creative design process by the integration of engineering design and cognitive psychology literature. *Design Stud.* 29, 279–302. doi: 10.1016/j.destud.2008.01.001
- Isen, A. M., Daubman, K. A., and Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *J. Pers. Soc. Psychol.* 52, 1122–1131. doi: 10.1037/0022-3514.52.6.1122
- Johansson, R., and Johansson, M. (2013). Look here, eye movements play a functional role in memory retrieval. *Psychol. Sci.* 25, 236–242. doi: 10.1177/0956797613498260
- Kashdan, T. B., Gallagher, M. W., Silvia, P. J., Winterstein, B. P., Breen, W. E., Terhar, D., et al. (2009). The curiosity and exploration inventory-II: development, factor structure, and psychometrics. *J. Res. Pers.* 43, 987–998. doi: 10.1016/j.jrp.2009.04.011
- Kenett, Y. N., Anaki, D., and Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Front. Hum. Neurosci.* 8:407. doi: 10.3389/fnhum.2014.00407
- Knoblich, G., Ohlsson, S., Haider, H., and Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *J. Exp. Psychol.* 25, 1534–1555. doi: 10.1037/0278-7393.25.6.1534
- Kounios, J., and Beeman, M. (2014). The cognitive neuroscience of insight. *Ann. Rev. Psychol.* 65, 71–93. doi: 10.1146/annurev-psych-010213-115154
- Kozbelt, A., Dexter, S., Dolese, M., Meredith, D., and Ostrofsky, J. (2015). Regressive imagery in creative problem-solving: comparing verbal protocols of expert and novice visual artists and computer programmers. *J. Creat. Behav.* 49, 263–278. doi: 10.1002/jocb.64
- Lakens, D., Adolfs, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., et al. (2018). Justify your alpha. *Nat. Hum. Behav.* 2, 168–171. doi: 10.1038/s41562-018-0311-x
- Liapis, A., Hoover, A. K., Yannakakis, G. N., Alexopoulos, C., and Dimaraki, E. V. (2015). “Motivating Visual Interpretations in Iconoscope: designing a game for fostering creativity,” in *Proceedings of the 10th International Conference on the Foundations of Digital Games* (Pacific Grove, CA).
- Lohse, G. L., and Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organ. Behav. Hum. Decis. Process.* 68, 28–43. doi: 10.1006/obhd.1996.0087
- Lung, C.-T., and Dominowski, R. L. (1985). Effects of strategy instructions and practice on nine-dot problem solving. *J. Exp. Psychol.* 11, 804–811. doi: 10.1037/0278-7393.11.1-4.804
- Lyubomirsky, S., and Lepper, H. S. (1999). A measure of subjective happiness: preliminary reliability and construct validation. *Soc. Indic. Res.* 46, 137–155.
- MacGregor, J. N., Ormerod, T. C., and Chronicle, E. P. (2001). Information processing and insight: a process model of performance on the nine-dot and related problems. *J. Exp. Psychol.* 27, 176–201. doi: 10.1037/0278-7393.27.1.176
- Maier, N. R. F. (1930). Reasoning in humans. I. On direction. *J. Comp. Psychol.* 10, 115–143. doi: 10.1037/h0073232
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* 69, 220–232. doi: 10.1037/h0048850
- Melton, A. W. (1962). Editorial. *J. Exp. Psychol.* 64, 553–557.
- Navalpakkam, V., and Churchill, E. F. (2012). “Mouse tracking: measuring and predicting users’ experience of web-based content,” in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*. Austin, TX: Association for Computing Machinery.
- Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Eaglewood Cliffs, NJ: Prentice-Hall International, Inc.
- Ohlsson, S. (1992). *Information-Processing Explanation of Insight and Related Phenomena*, vol. 1. London: Prentice-Hall.
- Öllinger, M., Jones, G., and Knoblich, G. (2014). The dynamics of search, impasse, and representational change provide a coherent explanation of difficulty in the nine-dot problem. *Psychol. Res.* 78, 266–275. doi: 10.1007/s00426-013-0494-8
- Pollio, M. V. (1914). *The Ten Books on Architecture*. London: Harvard University.
- Quiroga, R. Q., Mukamel, R., Isham, E. A., Malach, R., and Fried, I. (2008). Human single-neuron responses at the threshold of conscious recognition. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3599–3604. doi: 10.1073/pnas.0707043105
- Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan* 42. Available online at: <http://www.jstor.org/stable/20342603>
- Ritter, S. M., and Ferguson, S. (2017). Happy creativity: listening to happy music facilitates divergent thinking. *PLoS ONE* 12:e0182210. doi: 10.1371/journal.pone.0182210
- Rothmaler, K., Nigbur, R., and Ivanova, G. (2017). New insights into insight: neurophysiological correlates of the difference between the intrinsic “aha” and the extrinsic “oh yes” moment. *Neuropsychologia* 95, 204–214. doi: 10.1016/j.neuropsychologia.2016.12.017
- Runco, M. A., Abdulla, A. M., Paek, S. H., Al-Jasim, F. A., and Alsuwaidi, H. N. (2016). Which test of divergent thinking is best? *Creativity* 3, 4–18. doi: 10.1515/ctra-2016-0001
- Runco, M. A., and Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creat. Res. J.* 24, 66–75. doi: 10.1080/10400419.2012.652929
- Ryan, R. M., and Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemp. Educ. Psychol.* 25, 54–67. doi: 10.1006/ceps.1999.1020
- Sadler-Smith, E. (2015). Wallas’ four-stage model of the creative process: more than meets the eye? *Creat. Res. J.* 27, 342–352. doi: 10.1080/10400419.2015.1087277
- Salomons, N., van der Linden, M., Sebo, S. S., and Scassellati, B. (2018). “Humans conform to robots,” in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI’18*. Chicago, IL: ACM Press.

- Salti, M., Monto, S., Charles, L., King, J.-R., Parkkonen, L., and Dehaene, S. (2015). Distinct cortical codes and temporal dynamics for conscious and unconscious percepts. *eLife* 4, 1–19. doi: 10.7554/eLife.05652
- Salvi, C., Bricolo, E., Kounios, J., Bowden, E., and Beeman, M. (2016). Insight solutions are correct more often than analytic solutions. *Think. Reason.* 22, 1–18. doi: 10.1080/13546783.2016.1141798
- Sarcone, G. A. (2014). *Most Wanted Puzzle Solutions - The Nine Dot Puzzle*. Archived by WebCite*. Available online at: <http://www.webcitation.org/70XKaevlf>
- Scholz, A., von Helversen, B., and Rieskamp, J. (2015). Eye movements reveal memory processes during similarity- and rule-based decision making. *Cognition* 136, 228–246. doi: 10.1016/j.cognition.2014.11.019
- Schooler, J. W., Ohlsson, S., and Brooks, K. (1993). Thoughts beyond words: when language overshadows insight. *J. Exp. Psychol.* 122, 166–183. doi: 10.1037/0096-3445.122.2.166
- Scrucca, L., Fop, M., Murphy, T. B., and Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J.* 8, 205–233.
- Shen, W., Yuan, Y., Liu, C., and Luo, J. (2015). In search of the ‘aha’ experience: elucidating the emotionality of insight problem-solving. *Brit. J. Psychol.* 107, 281–298. doi: 10.1111/bjop.12142
- Silvia, P. J., Beaty, R. E., and Nusbaum, E. C. (2013). Verbal fluency and creativity: general and specific contributions of broad retrieval ability (gr) factors to divergent thinking. *Intelligence* 41, 328–340. doi: 10.1016/j.intell.2013.05.004
- Simon, H. A. (1973). The structure of ill-structured problems. *Artif. Intell.* 4, 181–201. doi: 10.1016/0004-3702(73)90011-8
- Skinner, B. F. (1984). Behaviorism at fifty. *Behav. Brain Sci.* 7:615. doi: 10.1017/S0140525X00027618
- Soukoreff, R. W., and MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of fitts’ law research in HCI. *Int. J. Hum. Comput. Stud.* 61, 751–789. doi: 10.1016/j.ijhcs.2004.09.001
- Steele, L. M., Johnson, G., and Medeiros, K. E. (2018). Looking beyond the generation of creative ideas: confidence in evaluating ideas predicts creative outcomes. *Pers. Individ. Diff.* 125, 21–29. doi: 10.1016/j.paid.2017.12.028
- Subramaniam, K., Kounios, J., Parrish, T. B., and Jung-Beeman, M. (2009). A brain mechanism for facilitation of insight by positive affect. *J. Cogn. Neurosci.* 21, 415–432. doi: 10.1162/jocn.2009.21057
- Thomas, L. E., and Lleras, A. (2007). Moving eyes and moving thought: on the spatial compatibility between eye movements and cognition. *Psychon. Bull. Rev.* 14, 663–668. doi: 10.3758/BF03196818
- Topolinski, S., and Reber, R. (2010a). Gaining insight into the “aha” experience. *Curr. Direct. Psychol. Sci.* 19, 402–405. doi: 10.1177/0963721410388803
- Topolinski, S., and Reber, R. (2010b). Immediate truth – temporal contiguity between a cognitive problem and its solution determines experienced veracity of the solution. *Cognition* 114, 117–122. doi: 10.1016/j.cognition.2009.09.009
- Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking-Norms: Technical Manual Research Edition-Verbal Tests, Forms A and B- Figural Tests, Forms A and B*. Princeton, NJ: Personnel Press. Inc.
- Trafimow, D., Amrhein, V., Areshenkoff, C. N., Barrera-Causil, C. J., Beh, E. J., Bilgiç, Y. K., et al. (2018). Manipulating the alpha level cannot cure significance testing. *Front. Psychol.* 9:699. doi: 10.3389/fpsyg.2018.00699
- Ullrich, C., Wallach, D., and Melis, E. (2003). “What is poor man’s eye tracking good for?” in *17th Annual Human-Computer Interaction Conference 2003* (Zürich).
- VanRullen, R., and Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* 13, 454–461. doi: 10.1162/089989290152001880
- Walcher, S., Körner, C., and Benedek, M. (2017). Looking for ideas: eye behavior during goal-directed internally focused cognition. *Conscious. Cogn.* 53, 165–175. doi: 10.1016/j.concog.2017.06.009
- Wallas, G. (1926). *The Art of Thought*. London: Jonathan Cape.
- Weisberg, R. W. (2006). *Creativity: Understanding Innovation in Problem Solving, Science, Invention, and the Arts*. Hoboken, NJ: Wiley.
- Wetzel, R., Rodden, T., and Benford, S. (2017). Developing ideation cards for mixed reality game design. *Trans. Digit. Games Res. Assoc.* 3, 175–211. doi: 10.26503/todigra.v3i2.73
- Wikipedia (2016). *Eureka Effect—Wikipedia, the Free Encyclopedia*. San Francisco, CA, (Online; accessed December 27, 2016).
- Wilson, R. C., Guilford, J. P., Christensen, P. R., and Lewis, D. J. (1954). A factor-analytic study of creative-thinking abilities. *Psychometrika* 19, 297–311. doi: 10.1007/BF02289230
- Zoefel, B., and VanRullen, R. (2017). Oscillatory mechanisms of stimulus processing and selection in the visual and auditory systems: state-of-the-art, speculations and suggestions. *Front. Neurosci.* 11:296. doi: 10.3389/fnins.2017.00296

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 Loesche, Goslin and Bugmann. 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.



The Role of Motor Activity in Insight Problem Solving (the Case of the Nine-Dot Problem)

Vladimir Spiridonov^{1,2*}, Nikita Loginov^{1,3}, Ivan Ivanchei¹ and Andrei V. Kurgansky^{1,4}

¹ Laboratory for Cognitive Research, The Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia, ² Laboratory for Cognitive Research, National Research University Higher School of Economics, Moscow, Russia, ³ Laboratory for the Cognitive Psychology of Digital Interface Users, National Research University Higher School of Economics, Moscow, Russia, ⁴ Laboratory of Neurophysiology of Cognitive Processes, Institute of Developmental Physiology, Russian Academy of Education, Moscow, Russia

OPEN ACCESS

Edited by:

Ian Hocking,
Canterbury Christ Church University,
United Kingdom

Reviewed by:

Rory MacLean,
Edinburgh Napier University,
United Kingdom
Pinar Oztop,
Liverpool Hope University,
United Kingdom

*Correspondence:

Vladimir Spiridonov
vspiridonov@yandex.ru

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 29 April 2018

Accepted: 03 January 2019

Published: 23 January 2019

Citation:

Spiridonov V, Loginov N, Ivanchei I and Kurgansky AV (2019) The Role of Motor Activity in Insight Problem Solving (the Case of the Nine-Dot Problem). *Front. Psychol.* 10:2. doi: 10.3389/fpsyg.2019.00002

Attempts to estimate the contribution made by motor activity to insight problem solving is hindered by a lack of detailed description of motor behavior. The goal of this study was to develop and put to the test a novel method for studying the dynamics of insight problem solving based on a quantitative analysis of ongoing motor activity. As a proper problem model, we chose the nine-dot problem (Maier, 1930), in which solvers had to draw a sequence of connected line segments. Instead of using the traditional pen-and-paper way of solving the nine-dot problem we asked participants to use their index finger to draw line segments on the surface of a tablet computer. We are arguing that successful studying of the role of motor activity during problem solving requires the distinction between its instrumental and functional role. We considered the functional role on the motor activity as closely related to the on-line mode of motor planning. The goal of Experiment 1 was to explore the potential power of the method and, at the same time, to assay the patterns of motor activity related to on-line and off-line modes of motor planning. Experiments 2 and 3 were designed to uncover the potential impact of preliminary motor training on the motor output of successful and unsuccessful problem solvers. In these experiments, we tested hypotheses on how preliminary motor training, which presumably played a functional role in Experiment 2 and an instrumental role in Experiment 3, affects the motor activity of a problem solver and hence their effectiveness in solving the problem. The three experiments showed consistent results. They suggest that successful solving of the nine-dot problem relies upon the functional role of motor activity and requires both off-line and on-line modes of motor planning, with the latter helping to overcome the perceptual constraints imposed by a spatial arrangement of the nine dots. The method that we applied allows for systematic comparison between successful and unsuccessful problem solvers based on the quantitative parameters of their motor activity. Through it, we found new specific patterns of motor activity that differentiate successful and unsuccessful solvers.

Keywords: problem solving, insight, nine-dot problem, motor planning, preliminary motor training

INTRODUCTION

The concept of insight has remained in focus of researchers since its introduction in 1917 by Köhler (1921). An insight can be defined as the moment of sudden comprehension of a problem solution often accompanied by an aha experience (Öllinger et al., 2013, 2017). Since then, a considerable number of theoretical models have been suggested to explain insight (insight solution) in terms of various mental mechanisms: for example, heuristic search (Kaplan and Simon, 1990; Ormerod et al., 2002) or representational change (Ohlsson, 1984; Knoblich et al., 1999; Öllinger et al., 2013).

The most popular theoretical models usually do not consider the solver's own motor activities which emerge while solving insight problems as a factor contributing to their solutions (Ohlsson, 1984; Kaplan and Simon, 1990; Knoblich et al., 1999; Ormerod et al., 2002). At odds with this view, data accumulated through a number of studies have shown that the motor activity of the solver is intimately woven into the fabric of the solving process. The solving process can be speeded up or delayed if preceded (Weisberg and Alba, 1981; Lung and Dominowski, 1985; Kershaw and Ohlsson, 2004) or accompanied (Thomas and Lleras, 2009) by the motor activity of the solver. The solver's movements can even play a decisive role in choosing among possible solutions of the problem at hands (Werner and Raab, 2013). In the study by Werner and Raab (2013), participants were asked to solve a modified version of the Maier's two-string problem. This version of the problem has two possible solutions: participants can either turn one of the strings into a pendulum by securing a weight to it (swing-like solution) or gain a higher position by stepping on the desk and connect the strings (step-like solution). Two groups of solvers participated in the experiment (Werner and Raab, 2013, Experiment 1). Prior to the test session, participants belonging to the first group were asked to swing their arms back and forth, while participants belonging to the second group had to step up onto and down off a chair. This experiment showed that participants from the 1st group more frequently chose the swing-like solution, while participants from the 2nd group preferred step-like solution. These and similar results are clearly not in line with existing models of insight and beg for an explanation.

Any attempt to estimate the contribution made by overt motor activity to a person's success (or failure) in finding an insight problem solution is hindered by the lack of variables quantifying motor behavior. A common practice among researchers is to use variables such as the number of trials along with the overall time needed to solve the problem and the percentage of correct responses. Unfortunately, using these variables results in averaging out any potential temporal dynamics in ongoing motor activity and, therefore, brings about an inability to differentiate between successful and unsuccessful problem solvers based on the patterns of those dynamics.

In this work, our first priority was to develop and put to the test a novel method for studying the dynamics of insight problem solving based on a quantitative analysis of ongoing motor activity. As a proper problem model, we chose one of the most studied insight problems, the nine-dot problem (Maier,

1930) (see **Figure 1A**). This problem is traditionally considered insightful because it provokes the emergence of an inadequate initial representation, which hinders the solution: in the initial stages, the subjects connect dots with lines, without going beyond the limits of the square. To solve the problem, a radical change (restructuring) of the initial representation is required. It is this change of the initial representation, which is associated with insight (Scheerer, 1963). For a detailed analysis and criticism, see (Weisberg, 1995).

In the nine-dot problem, motor activity takes the form of sequential movements executed in order to draw a proper spatial trajectory – a sequence of connected line segments. Instead of using the traditional pen-and-paper way of solving the nine-dot problem we asked participants to use their index finger to draw line segments on the surface of a tablet computer. This allows for using variables that characterize the temporal structure of the graphical movements executed by problem solvers. Since the whole experiment is arranged as a block of trials (i.e., successive attempts to solve the problem), the sequence of parameters could be used to discover characteristic patterns of motor activity and to see if and how these patterns change across the series of trials.

Our second priority was to try to describe what patterns of motor activity distinguish between successful and unsuccessful nine-dot problem solvers.

THE ROLE OF MOTOR ACTIVITY IN SOLVING THE NINE-DOT PROBLEM

There are two roles that motor activity might play in solving insight problems: instrumental and functional. When taken in its instrumental role, the motor activity does not influence the nature of the solution but merely implements the solution already found with some other cognitive processes. For example, in case of the nine-dot problem, the instrumental role of motor activity would be limited by drawing a correct sequence of connected line segments (similar to the one shown in **Figure 1B**), which had been prepared in advance. The instrumental role of motor activity in solving other insight [e.g., 6-coin (Chronicle et al., 2004), 8-coin (Ormerod et al., 2002), 6 matches (Scheerer, 1963), etc.] and non-insight [e.g., 5 rings Tower of Hanoi (Anzai and Simon, 1979)] problems is also the implementation of the sequence of movements leading to the correct solution, which was previously constructed in the mind. The examples that illustrate the instrumental role of motor activities for relatively simple motor tasks are in: (Tessari and Rumiati, 2004; Tessari et al., 2006).

When playing a functional role, motor activity lays the very ground for the solution being sought, i.e., the motor activity directly affects the process of problem-solving and the outcome of that process. This view has received some experimental support (Grant and Spivey, 2003; Thomas and Lleras, 2009; Werner and Raab, 2013). Thus, in a study by Werner and Raab (2013), in experiment 2, the modified water-jar problem (Luchins, 1942) was used. This problem could be solved either by (1) subtracting the amount of water held by one of the smaller jars twice from the biggest one or (2) by adding the amount of water held by

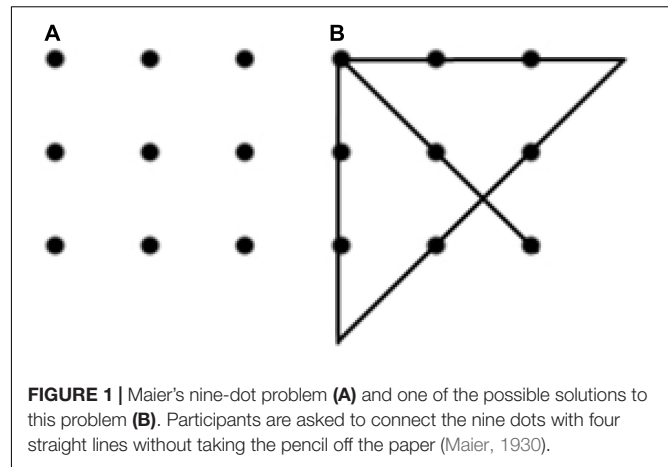
one smaller jar twice to the other smaller jar. As a prime for the subtraction solution (group 1), a 30-s preliminary procedure was used to move marble balls from the middle jar into two outer jars, while the priming for addition solution (group 2), a similar procedure of moving the same balls from the outer jars to the middle one. It was found that subjects of group 1 more often used the subtraction solution while their group 2 counterparts more often relied on the addition solution.

However, there are few such studies, and they are vulnerable to criticism. In most cases, it remains unknown whether the reported results are truly related to the functional role of motor activities rather than reflecting the form of some abstract idea hinted at by these activities. For example, in two similar studies (Thomas and Lleras, 2009; Werner and Raab, 2013) the arm swinging preceding the test session not only directly points to the movement pattern critical for solving the two-string problem but also indirectly prompts the abstract idea of a pendulum and similar ideas. Thus, experimental studies that have been conducted so far leave unanswered the question of how the motor activity relates to the process of solving insight problems. In particular, the question of whether motor activity plays a functional role also remains largely unanswered.

We assumed that in the case of the nine-dot problem, it may be related to a certain mode of motor planning. According to Wilson's definition, two kinds of cognition have to be distinguished: "on-line" (or "situated") cognition and "off-line" cognition (Wilson, 2002, p. 626). On-line cognition critically depends on the particular conditions (including spatial ones) in which they take place. It is linked to the properties of the surroundings and makes use of the latter in order to reduce the cognitive processing burden, is sensitive to different kinds of affordances which automatically trigger specific motor programs, etc. In contrast, off-line cognition takes place in the mental domain without any apparent influence of the surrounding environment.

The off-line vs. on-line distinction fully applies to a motor activity which includes two major phases, known as the motor preparation phase and the motor execution phase. It is often assumed that the most important cognitive processes take place during the first preparatory phase and that taken together constitute what is known as motor planning. In other words, the term "motor planning" refers to those cognitive processes that are related to a movement and precede it (Stanford, 2013).

One might think that motor planning is an off-line process by definition. However, studies of movements toward a spatial goal in the condition of the uncertainty of its position (Scott, 2012; Wong et al., 2015; Wong and Haith, 2017) and the data on the role of sensory feedback and its prediction (Scott et al., 2015) show that planning can be an on-line process. When relying on off-line planning, a problem solver prepares an entire movement sequence (or a substantial fraction thereof) ahead of time and then executes it uninterrupted. In this mode, the only opportunity to estimate the surrounding environment and to select the appropriate movements is prior to the sequence execution. Similarly, the opportunity to estimate the results of the movement execution exists after the sequence has been executed. Therefore, one may say that off-line planning has a long but



narrow horizon. In the case of the nine-dot problem, this mode of planning is akin to the notion of a "mental lookahead" (Ohlsson, 1984; MacGregor et al., 2001). Mental lookahead directs the heuristic search in the course of the problem solution due to the anticipation of new states within the "problem space." Its range is limited (Ohlsson, 1984). In the course of solving the nine-dot problem, it can vary in horizon by representing from one to four straight lines (MacGregor et al., 2001). Regardless of the depth of the mental lookahead, the off-line planning is completed before any movement has occurred (drawing lines connecting dots).

In contrast, on-line planning goes hand by hand with movement execution. This mode of planning allows for a continuous re-evaluation of the surrounding conditions while taking into account the solution being searched for and the results of the already executed movements. Thus, when compared to off-line planning, on-line planning has a wide but short horizon. It opens different options to continue with the already started movement or movement sequence.

A major difficulty that the problem solver faces while attempting to find a solution to the nine-dot problem is incompleteness of the mental representation of the task, i.e., a lack of constituents (perceptual and abstract entities) which are critical for constructing a correct solution. Such incompleteness manifests itself in a limited repertoire of movements and results in an inability to solve the problem. An attempt to solve the problem usually begins with drawing straight lines along the outer sides of the nine-dot square which points to a rather narrow repertoire of movements.

We assumed that in the case of the nine-dot problem, relying exclusively on off-line planning is insufficient in order to overcome this narrow repertoire of movements. In the study by MacGregor et al. (2001), a theoretical model was developed to explain heuristic search in the course of the solving of the nine-dot problem through the exploit of maximization and progress-monitoring heuristics with a variable lookahead depth ranging from 1 to 4 consecutive line segments. This model has gained empirical support from the experiments involving the problems similar to but way more simple than the nine-dot problem (MacGregor et al., 2001, Experiments 1, 2, 3). Thus, in experiment 1 of the cited paper, a percentage of participants who

successfully solved the problems varied from 80 to 93% while no one solved the nine-dot problem. It seems that unlike the original nine-dot problem, the simplified problems (see **Figure 2** in the cited paper) provide the stronger hints for the initial line segments which are the part of a correct solution. It helps solvers to rely on a shorter mental lookahead. However, the model does not explain the evolution of the line segments that are drawn by solvers. What begs for explanation is how the participants manage to go beyond the square area defined by the nine dots, i.e., to start and end the line segments outside this area. It is at this point that the on-line planning reveals its significant role.

The advantages of this mode of planning are as follows. First, within a single attempt to solve the problem (i.e., to draw a proper sequence of four connected line segments), on-line planning gives more opportunities to build a proper solution than does the off-line mode. This is because in the former case, the construction process goes on all the time, and it is not limited to the period of time prior to the sequence execution. Second, the evaluation of the intermediate results of movements makes it more probable to get an idea that a trajectory vertex (its joint or turning point) may not necessarily coincide with one of the nine visible dots. Finally, a permanent monitoring of motion, i.e., keeping track of an index fingertip position and its velocity, might bring into focus the idea of motion direction, whose spatial trajectory is a straight-line segment with off-dot margins. Under these circumstances, a problem solver may discover with a greater probability that a line segment does not necessarily begin or end with one of the visible dots and that the angle between two consecutive lines is not necessarily a right angle.

It is required by the nature of the nine-dot problem that the spatial trajectory (path) corresponding to its correct solution has to take a form of piecewise linear curve containing 4-line segments and connecting (passing through) all 9 points. However, these requirements do not impose any constraints on whether or not this trajectory is pre-planned as a whole ahead of its execution or on the timing of the fingertip movement along this path. The trajectories produced by solvers of the nine-dot problem showed multiple stops between positions of visible dots sometimes very long (up to several seconds). Because of that, we do not have any reason to think that off-line planning takes place during pauses in the spatial trajectory vertices. Instead, we made two assumptions. We assume that (A1) the off-line planning contribution is proportional to the average stop duration (inter-movement pause duration) and (A2) the contribution of on-line planning is proportional to the average movement duration (i.e., inversely proportional to the average movement velocity). These assumptions are supported by the following. First, longer movement sequence is characterized by a longer latency time and a longer execution time of its units (for a review, see Rhodes et al., 2004). Second, planning complex trajectories takes longer than simple reaching movements to a certain spatial position (Wong et al., 2015). Finally, relying on on-line planning leads to a reduction in movement latency time (Orban de Xivry et al., 2017)

and therefore results in shorter pauses between consecutive movements.

PRELIMINARY MOTOR TRAINING AND ITS IMPACT ON SOLVING THE NINE-DOT PROBLEM

We conducted three experiments. The primary goal of Experiment 1 was to assess the method's potential explanatory power and, at the same time, to assay the patterns of motor activity related to on-line and off-line modes of motor planning. The second and third experiments were designed to uncover the potential impact of preliminary motor training on the motor output of the successful and unsuccessful problem solvers.

A known way to boost the probability of the correct solving of the nine-dot problem is to ask participants to precede their attempts to solve the problem by motor training – by drawing those line segments that are part of the correct solution (Weisberg and Alba, 1981, Experiment 2; Lung and Dominowski, 1985; Chronicle et al., 2001, Experiment 3). Using preliminary motor training allows us to uncover the movements (and combinations of thereof) that play an important role in problem-solving and to shed light on both the nature and the sources of the difficulties the problem solvers met (Kershaw and Ohlsson, 2004). In particular, we believe that using motor training also allows for studying the contribution made by the two modes of motor planning mentioned above.

The traditional variant of preliminary motor training does not distinguish between the instrumental and functional role of motor activity. For example, Kershaw and Ohlsson (2004, Experiment 1) varied two factors that were related to preliminary motor training. These factors were (i) the presence/absence of non-dot turns, i.e., actual abrupt changes in movement direction taking place outside the nine dots area and (ii) the presence/absence of perceptual cues for non-dot turns. In order to accomplish the task, a solver has to arrange the required movements while keeping in mind the verbal instructions (“connect the dots by straight lines”). This mode of motor training involves both kinds of motor planning (on-line and off-line) as well as instrumental aspects of a motor activity.

In order to discriminate between the instrumental and functional roles that preliminary motor training might play, we studied the impact of the training on the solving process in each of the following two conditions: in the “no task” condition (movements played a predominantly instrumental role) and in the context of a task in which movements played both an instrumental and functional role. In our Experiment 2, we used traditional preliminary motor training in which participants practiced drawing pairs of consecutive segments with their connection point (vertex) situated out of the nine-dot display [usually referred to as “non-dot turns” (Kershaw and Ohlsson, 2004)]. These line drawings are known to be the crucial elements of the correct solution for the nine-dot problem. This kind of training involved both off-line and on-line planning modes. In Experiment 2, participants were asked to connect dots by two connected straight-line segments. These line segments were

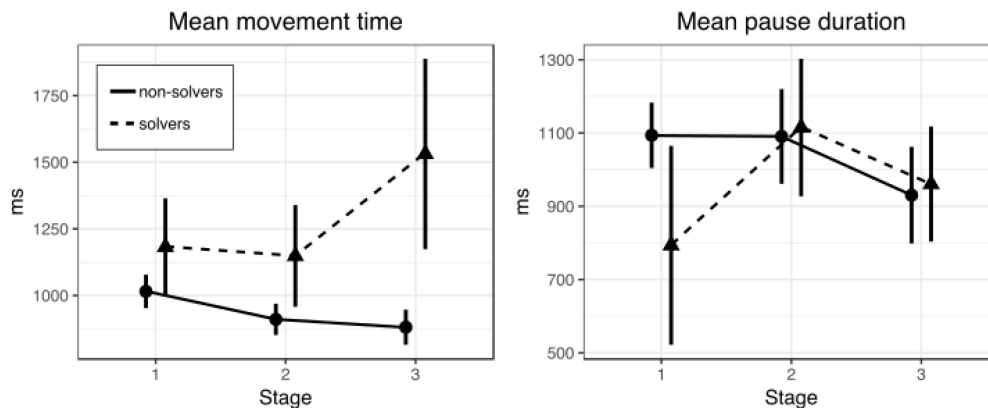


FIGURE 2 | Mean movement (left) and pause (right) time in three stages of the nine-dot problem solving (Experiment 1). Bars represent within-subject 95% confidence intervals.

oriented at an angle that could take two different values. Here the preliminary motor training was explicit and took place in the context of a task that was relevant to the upcoming problem. In our Experiment 3, the preliminary motor training was implicit and proceeded in the context of a task that was seemingly irrelevant to the nine-dot problem. In this Experiment, we used a modified version of the implicit learning paradigm, in which participants remained unaware of either the results of the learning or the learning itself (Nissen and Bullemer, 1987; Cleeremans et al., 1998). Applying this experimental technique makes it possible to estimate the effect of specific movements on how efficient solvers are in finding the solution to the problem. In Experiments 2 and 3, we tested hypotheses on how preliminary motor training, which presumably played a functional role in Experiment 2 and an instrumental role in Experiment 3, affects the motor activity of a problem solver and hence the effectivity of solving the problem. In sum, the goal of the present study was to identify movement sequences executed during attempts to solve the nine-dot problem. To this end, the experimental procedure was modified so that it allowed for the recording of the motor activity with a tablet computer and for the extraction of informative parameters of this activity such as the times taken for drawing line segments and the duration of pauses between successive movements.

EXPERIMENT 1

In the first study, we attempted to identify the differences between successful and unsuccessful nine-dot problem solvers by using several variables that characterized the motor activity of solvers.

Methods

Participants

Forty-five volunteers (35 women; 18–21 years old, $M = 19.32$; $SD = 0.59$) from Moscow universities (RANEPA, NRU HSE) participated in the experiment in return for course credits. Six participants were excluded from the further analysis because in the post-experimental survey they reported that the nine-dot

problem was familiar to them. Three participants solved the nine-dot problem unconventionally (angles were not equal to 45 degrees) and were excluded from the analysis too.

This study was carried out in accordance with the recommendations of institutional guidelines of the ethics committee of the Department of Psychology of RANEPA (Russian Academy of National Economy and Public Administration). The protocol was approved by the ethics committee of the Department of Psychology of RANEPA. All participants gave written informed consent in accordance with the Declaration of Helsinki.

Apparatus and Stimuli

Conducting experiments, we used a custom program in Delphi language on an Asus tablet (10.1-inches screen diagonal; 1280 pixels \times 800 pixels, PPI = 143; Intel Atom X5-Z8500 quad-core processor clocked at 1.44 GHz; operating system Windows 10). The software presented the nine-dot problem and recorded the motor activity of participants trying to solve the problem. The participants used the tip of their index finger to draw line segments on the screen of the tablet. All movements left visible traces on the tablet screen.

At the beginning of the experiment, the program recorded the age, sex, and participant identification number. Then it presented the instructions and an image of nine dots. Nine black dots were presented in the form of a “square” in the center of a tablet’s screen. Each dot was 10 mm in diameter. The distance between neighboring dots was 15 mm vertically and horizontally.

Design and Procedure

Participants solved the nine-dot problem while sitting at a table. The tablet was on the table in front of them. Participants were asked to solve the nine-dot problem. First, they were presented with on-screen instructions (in Russian): “Please connect all 9 dots by drawing four straight lines with the tip of your index finger without taking your index finger off the screen of the tablet.”

No standard home position for the index finger was used so participants were free to start from any point on the screen. As

soon as participants began drawing lines, the program collected raw data of their motor activity (coordinates of all points in drawn lines in pixels and the processor time corresponding to each coordinate value in milliseconds). In the upper left corner of the screen were two buttons: “Save” and “Next trial.” If participants succeeded in solving the problem, they pressed “Save.” However, if they failed to solve the problem, they pressed “Next trial” and tried to solve it again. The experiment was limited to 100 trials, and if a participant did not solve the nine-dot problem within this number of trials, he or she was considered unsuccessful. In addition to the parameters of motor activity, the solution time, solution rate and a number of used trials were also recorded. The experiment was carried out individually. At the end of the experiment, participants were asked whether they were familiar with the nine-dot problem. If they responded positively, they were excluded from further analysis.

The duration of pauses between lines in milliseconds and the duration of one line drawing in milliseconds were the dependent variables. The grouping variables were the solution rate and the stage of the nine-dot problem-solving. The stages of problem-solving were set by dividing the total number of trials of each participant into three equal parts (first, second, and third). A similar way of analyzing data was used in studies of oculomotor activity during the insight solution (Knoblich et al., 2001).

Data Analysis

We used Octave/Matlab custom software to analyze movement recordings. The analysis proceeded through several successive stages (Korneev and Kurgansky, 2013). In the first stage, we used the linear interpolation technique to convert the original time series into the time series $x(n)$ and $y(n)$ equally spaced in time (here n stands for a discrete time). In the second stage, the $x(n)$ and $y(n)$ series were smoothed with the 2nd order Butterworth low-pass filter with cut-off frequency of 5 Hz. A forward and reverse filtering was applied to the time series to preserve the original phase spectrum. The resultant smooth planar trajectory $\{x(n), y(n)\}$ was used to compute instantaneous tangential velocity $v(n)$. In the third stage, the entire movement recording was broken into a sequence of successive submovements. To that end, all the local peaks in $v(n)$ time series were found. In order to reduce the noise caused by physiological tremor and small corrective submovements, any peak whose height was less than 10% of the height of the tallest peak was excluded from further analysis. For each of the valid tangential velocity peaks its margins were determined. It was assumed that $v(n)$ is a monotonically increasing function of the discrete time n on the left-hand side of a peak corresponding to a submovement while it is a monotonically decreasing function of n on the right slope of the peak. Therefore, the leftmost time point of increasing slope and the rightmost time point of the decreasing slope were taken as the beginning and the end of the peak. As a result of the above procedure, all movement recordings were broken into a sequence of peaks (corresponding to non-overlapping fractions of submovements).

In the final stage, all extracted submovements were assigned to a certain line segment. For each extracted peak, a vector pointing from the starting to the end position was computed.

Any pair of adjacent vectors were considered as belonging to the same line segment if the angle between these two vectors did not exceed a chosen critical angle (usually 30 degrees). Potentially, the sequence of extracted peaks and their assignment to a particular trajectory segment can be used in order to compute a number of variables that constitute very detailed multidimensional characteristics of a motor activity of a problem solver. In the present work, we used two variables which are referred to throughout the paper as “movement time” and the “pause duration.” The movement time variable corresponds to the mean time across all segments required to draw a single line segment. This value does not include the time of staying motionless (or moving very slowly with a velocity below some predefined threshold) in the joints of the trajectory. The latter time is characterized by the second variable, pause duration. This variable is computed by averaging all the particular pauses detected during drawing a sequence of line segments. The reason why we limited our scope to these two variables is that they are presumably related to the on-line and off-line motor planning modes, correspondingly.

Results

Movement Time

The first question is whether solvers and non-solvers differed in the movement time during line drawing at different stages (first, second, and third) of the solution. The overall solution rate was 52.8% (19 solvers and 17 non-solvers). A 2×3 repeated measures ANOVA with SUCCESS (solvers and non-solvers) as a between-subjects factor and STAGE (first, second, and third) as a within-subjects factor revealed no significant main effects ($p = 0.09$ and $p = 0.27$, respectively). However, there was a significant interaction between factors of SUCCESS and STAGE [$F(2,68) = 3.3$, $p = 0.044$, $\eta_p^2 = 0.09$]. **Figure 2** shows mean movement time for solvers and non-solvers in the three stages of the nine-dot problem solving.

A series of t -tests for independent samples were conducted to clarify at which stages solvers and non-solvers differ (**Table 1**). There were no differences between solvers and non-solvers at the first ($p = 0.63$) and the second stages ($p = 0.38$). But we found that solvers drew lines significantly more slowly than non-solvers at the third stage [$t(21) = 2.39$, $p = 0.03$, $d = 0.78$]. We used Welch's t -test because variances were unequal.

Pause Duration

The next question is whether solvers and non-solvers differed in the pause duration between line drawing at different stages (first, second, and third) of the solution (**Table 2**). A 2×3 repeated measures ANOVA with SUCCESS (solvers and non-solvers) as a between-subjects factor and STAGE (first, second, and third) as a within-subjects factor revealed no significant main effect of SUCCESS ($p = 0.55$), STAGE ($p = 0.18$) or interaction of SUCCESS and STAGE ($p = 0.13$).

Discussion

These results show that the way off-line planning mode was used did not change across successive stages of the process of solving the nine-dot problem either in successful or in unsuccessful

problem solvers. This conclusion is supported by the absence of significant changes in pause duration across successive stages of the solving process. However, we observed a significant difference between successful and unsuccessful problem solvers in the movement time parameter at the final third stage of the solving process (supported by the presence of a statistically significant SUCCESS and STAGE interaction). This finding suggests that successful solvers relied more on on-line planning than their unsuccessful peers. The results of this experiment show that analyzing actual movement patterns is capable of providing new information on the processes underlying the solving of insight problems.

Similar differences between successful and unsuccessful problem solvers were found using eye tracking during the final stage of the problem solving (Knoblich et al., 2001). They found that it was the third stage of the solving process in which the average duration of long fixations spent on crucial elements in matchstick arithmetic problems was significantly longer in successful than in unsuccessful problem solvers. The explanation suggested by Knoblich et al. (2001) involved a re-structuring the inner representation of the problem, which in turn caused a re-distribution of attention from irrelevant to relevant task conditions. Thus, one may say that they considered the motor activity of problem solvers from the instrumental perspective, i.e., as something caused by the functioning of mental mechanisms. However, our data showed that successful nine-dot problem solvers mostly rely on on-line planning, thus pointing to the functional role of motor activity. The experiments that follow are designed to study the functional role of motor activity.

EXPERIMENT 2

Experiment 1 suggests a link between the success in solving the nine-dot problem and the on-line mode of motor planning. Experiment 2 is aimed at verifying whether the on-line planning can causally influence successfulness of the nine-dot problem solving. In order to elucidate the role of motor activity in the

successful solving of the nine-dot problem we used a well-known method – preliminary motor training, i.e., practicing isolated constituents of a correct solution of a problem. If such preliminary training has a positive impact on finding the problem solution (Weisberg and Alba, 1981; Lung and Dominowski, 1985) then this method allows studying not only the instrumental role of motor activity but also its functional role. We expected that the functional role of motor activity would be most noticeable in the case of practicing non-dot turn, which is one of the key elements of the correct nine-dot problem solution. A non-dot turn is a turn made by the pen tip outside the square area that contains all nine dots. This element of the solution was considered by Kershaw and Ohlsson (2004). The purpose of the second experiment was to study how two factors that characterize the preceding motor activity, practicing dot vs. non-dot turns and practicing turns with the solution-relevant (45 degrees) vs. solution-irrelevant (26.6 degrees) angles, influence solving of the nine-dot problem.

Methods

Participants

A total of 74 volunteers (65 women; 17–28 years old, $M = 19.0$; $SD = 0.59$) from Moscow universities (RANEPA, RSUH) participated in the experiment in return for course credits. Five participants were excluded from the further analysis because their solution to the nine-dot problem, although correct, was geometrically unconventional (angles were not equal to 45 degrees). Five participants were removed from the further analysis because they solved the nine-dot problem in less than three trials. One participant who had mean values of pauses duration more than 3 standard deviations from the average was excluded too.

This study was carried out in accordance with the recommendations of institutional guidelines of the ethics committee of the Department of Psychology of RANEPA. The protocol was approved by the ethics committee of the Department of Psychology of RANEPA. All participants gave written informed consent in accordance with the Declaration of Helsinki.

TABLE 1 | Mean and standard deviation of movement time in the three stages of the nine-dot problem solving.

	Solvers			Non-solvers			
	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	<i>M</i> _{stages}
First stage	1142.3	642.16		1059.73	287.89		1103.31
Second stage	1108.06	625.32	1252.11	955.13	337.95	974.58	1035.84
Third stage	1505.96	1052.35		908.87	271.79		1223.99

TABLE 2 | Mean and standard deviation of pause duration in the three stages of the nine-dot problem solving.

	Solvers			Non-solvers			
	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	<i>M</i> _{stages}
First stage	749.6	450.38		1114.04	654.85		931.82
Second stage	1118.07	944.16	924.93	1089.08	912.91	1065.43	1103.58
Third stage	907.12	896.49		993.16	611.79		950.14

Apparatus and Stimuli

The nine-dot problem was administered the same way as in Experiment 1 with the only exception of the tablet computer model. In this experiment, we used an HP tablet (10.1-inches screen diagonal; 1280 pixels × 800 pixels, PPI = 143; the Intel Atom Z3735G quad-core processor clocked at 1.33 GHz; operating system Windows 10).

Before solving the nine-dot problem, participants solved a series of motor training tasks. These were presented on the tablet using the same software as in Experiment 1. Participants were presented with four or five dots, which were arranged so that two straight lines with a turn of 45 or 26.6 degrees could connect them (see **Figure 3**). Each motor training task was repeated 4 times with the angle vertex pointing to different directions (angle up, angle down, angle to the right, and angle to the left).

Design and Procedure

Participants were asked to solve several motor training tasks. In the first four tasks, it was necessary to connect dots with two lines, without lifting the index finger from the screen of the tablet. In the upper left corner of the screen, there were two buttons: “Done” and “Next trial.” If participants succeeded in solving a task, they pressed “Done”. However, if they failed to solve a task, they pressed “Next trial” and tried to solve it again. The number of trials for these motor training tasks was unlimited.

Participants were randomly distributed into four groups. In Group 1, motor training tasks required participants to perform a non-dot-turn with 26.6 degrees (see **Figure 3**). In Group 2, motor training tasks required participants to perform a non-dot-turn with 45 degrees. In Group 3, motor training tasks required participants to perform a dot-turn with 26.6 degrees. And in Group 4, motor training tasks required participants to perform a dot-turn with 45 degrees. Within the groups, the sequence of presentation of tasks was random. After these tasks, participants proceeded to the nine-dot problem. The procedure for solving the nine-dot problem was the same as in Experiment 1. At the end of the experiment, participants were asked whether they used their experience in solving the first four tasks during the nine-dot problem-solving and whether they were familiar with the nine-dot problem.

Results

To control whether the non-dot-turn training and correct angle of turn training affected problem solving performance, we compared two groups of participants in terms of solution rate.

Impact of the Preliminary Motor Training on the Performance: Non-dot Turn vs. Dot Turn

The overall solution rate of the nine-dot problem was 57.8%; 37.5% in the dot-turn training group and 78.1% in the non-dot-turn training group (see **Table 3**). According to Chi-square test, the association between training type (non-dot turn vs. dot turn) and solution rate was statistically significant [$\chi^2(1, N = 64) = 10.83, p = 0.001$].

Impact of the Preliminary Motor Training on the Performance: Correct vs. Incorrect Angle

There were 51.6% of successful solutions to the nine-dot problem in the incorrect angle (26.6 degrees) training group and 63.6% in the correct angle (45 degrees) training group (**Table 4**). According to Chi-square test, the association between training type (correct vs. incorrect angle of turn) and solution rate was not significant: $\chi^2(1, N = 64) = 0.95, p = 0.33$.

Movement Time

As in Experiment 1, we tried to find similar differences between solvers and non-solvers in the movement time during line drawing at different stages (first, second, and third) of the solution. Movement times were subjected to a 2×3 repeated measures ANOVA with SUCCESS (solvers and non-solvers) as a between-subjects factor and STAGE (first, second, and third) as a within-subjects factor. This analysis revealed a significant main effect of SUCCESS [$F(1,61) = 18.38, p = 0.001, \eta_p^2 = 0.23$] and a significant interaction between factors of SUCCESS and STAGE [$F(2,122) = 6.23, p = 0.003, \eta_p^2 = 0.09$]. There was no significant main effect of STAGE ($p = 0.96$). **Figure 4** shows mean movement time for solvers and non-solvers in all three stages of the nine-dot problem solving after motor training.

A series of *t*-tests for independent samples were conducted to clarify at which stages solvers and non-solvers differ (**Table 5**). We used Welch’s *t*-test as variances are unequal. There were no differences between solvers and non-solvers at the first stages ($p = 0.08$). But we found that at the second [$t(54) = 3.39, p = 0.03, d = 0.98$] and at the third [$t(50) = 6.07, p < 0.001, d = 1.49$] stages solvers drew lines significantly more slowly than non-solvers.

Impact of Motor Training on Movement Time

Movement time was subjected to a 2×2 ANOVA with NON-DOT TURN (non-dot turn training and dot turn training) and ANGLE (correct angle training and incorrect angle training) as a between-subjects factors. This analysis revealed a significant main effect of NON-DOT TURN [$F(1,61) = 7.8, p = 0.007, \eta_p^2 = 0.12$], but no significant main effect of ANGLE ($p = 0.22$) and no interaction of NON-DOT TURN and ANGLE ($p = 0.76$) were found (**Table 6**).

TABLE 3 | Solution rate in two experimental groups with and without non-dot-turn at the motor training.

	Non-solvers	Solvers	Total
Dot-turn training	20	12	32
Non-dot-turn training	7	25	32
Total	27	37	64

TABLE 4 | Solution rate in two experimental groups with and without the correct angle of turn at the motor training.

	Non-solvers	Solvers	Total
Correct angle training	12	21	33
Incorrect angle training	15	16	31
Total	27	37	64

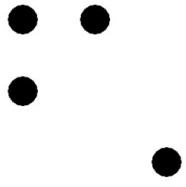
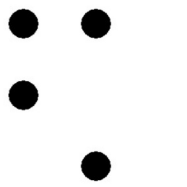
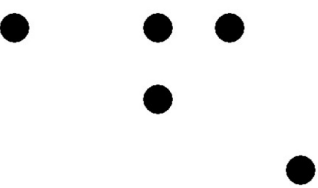
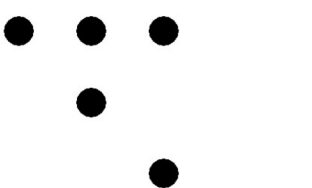
	Incorrect angle of turn (26.6 degrees)	Correct angle of turn (45 degrees)
Non-dot turn		
Dot turn		

FIGURE 3 | Types of motor training tasks. Motor training tasks were divided into four groups: non-dot turn and incorrect angle of turn (Group 1); non-dot turn and correct angle of turn (Group 2); dot-turn and incorrect angle of turn (Group 3); and dot-turn and correct angle of turn (Group 4).

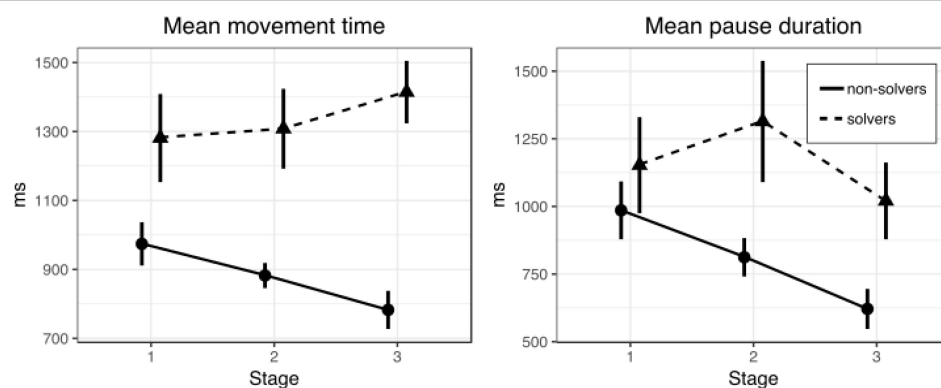


FIGURE 4 | Mean movement (left) and pause (right) time in three stages of the nine-dot problem solving (Experiment 2). Bars represent within-subject 95% confidence intervals.

TABLE 5 | Mean and standard deviation of movement time in the three stages of the nine-dot problem solving after motor training.

	Solvers			Non-solvers			M_{stages}
	M	SD	M_{total}	M	SD	M_{total}	
First stage	1288.13	656.49	1376.08	1035.28	323.82	928.39	1180.36
Second stage	1346.97	572.98		927.93	302.30		1168.36
Third stage	1493.15	578.97		821.97	261.85		1207.07

Pause Duration

Also, as in Experiment 1, we tried to find similar differences between solvers and non-solvers in the pause duration between lines drawing at different stages of the solution. A 2×3 repeated measures ANOVA with SUCCESS (solvers and non-solvers) as a between-subjects factor and STAGE (first, second, and third) as a within-subjects factors revealed significant main effects of SUCCESS $F(1,61) = 4.49$, $p = 0.038$, $\eta_p^2 = 0.07$, and

STAGE $F(2,122) = 3.18$, $p = 0.045$, $\eta_p^2 = 0.05$. The interaction between factors of SUCCESS and STAGE was also significant, $F(2,122) = 4.32$, $p = 0.015$, $\eta_p^2 = 0.07$. **Figure 4** shows means for pause duration for solvers and non-solvers in all three stages of the nine-dot problem solving after motor training.

A series of t -tests for independent samples were conducted to clarify at which stages solvers and non-solvers differ (**Table 7**). We used Welch's t -test, as variances are unequal. There were no

TABLE 6 | Movement time in four experimental groups with different types of motor training.

	Dot-turn training		Non-dot-turn training		Total	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Correct angle training	888.32	279.54	1261.08	410.59	1081.13	395.61
Incorrect angle training	1128.33	323.45	1412.24	859.76	1270.29	655.05
Total	1016.33	322.45	1339.10	673.92		

TABLE 7 | Mean and standard deviation of pause duration in the three stages of the nine-dot problem solving after motor training.

	Solvers			Non-solvers			
	<i>M</i>	<i>SD</i>	<i>M_{total}</i>	<i>M</i>	<i>SD</i>	<i>M_{total}</i>	<i>M_{stages}</i>
First stage	1067.9	642.2		1137.9	702.6		1097.7
Second stage	1390.2	1064.4	1164.6	825.3	395.7	885.4	1149.4
Third stage	1035.6	754.3		693.0	414.6		889.6

differences between solvers and non-solvers in the first stages of the solution ($p = 0.69$). But we found that at the second [$t(46) = 2.88$, $p = 0.01$, $d = 0.5$] and at the third [$t(55) = 2.27$, $p = 0.03$, $d = 0.56$] stages solvers make significantly longer pauses between drawing lines than non-solvers.

Impact of the Motor Training on Pause Duration

Pause duration was subjected to a 2×2 ANOVA with NON-DOT TURN (non-dot turn training and dot turn training) and ANGLE (correct angle training and incorrect angle training) as a between-subjects factors. This analysis revealed no significant main effect of NON-DOT TURN ($p = 0.08$), ANGLE ($p = 0.23$) and interaction of NON-DOT TURN and ANGLE ($p = 0.41$) (Table 8).

Discussion

Experiment 2 showed that preliminary motor training involving non-dot turns resulted in more success in finding a correct solution as compared to the training that did not involve these turns. Practicing a task-relevant turn of 45 degrees was no better than practicing a task-irrelevant turn of 26.6 degrees. Although the latter finding is in line with previous studies (Kershaw and Ohlsson, 2004), it does not support our hypothesis of the superiority of a task-relevant angle of 45 degrees. It may well be that the direction of the upcoming movement is an essential part of the motor plan since it helps to transcend the perimeter of the visible nine dot display whereas angles between two successive segments are not parts of the movement plan.

When comparing the results of the present experiment with those of Experiment 1, one can notice that motor training caused the difference between successful and unsuccessful problem solvers in parameters quantifying on-line planning not only at the final stage of the solution but also at the second stage. This finding suggests that being affected by preliminary motor training, successful problem solvers tended to invoke an on-line mode of movement planning at earlier stages of the process of solving the nine-dot problem. Besides, in this experiment, we found a difference between successful and non-successful problem solvers in pause duration during the second and the third stages of the problem-solving process. The latter finding suggests that successful solvers rely to a greater extent on off-line planning than their unsuccessful peers.

Results of this experiment suggest that processes underpinning motor planning make a substantial contribution to the successful solving of the nine-dot problem. We found that successful problem solvers showed greater movement time (associated with on-line planning) as well as greater pause duration (associated with off-line planning) than their unsuccessful counterparts. This finding is in accordance with the view that both kinds of planning contribute to the successful solving of the nine-dot problem.

A slowing down of drawing lines which is found in successful problem solvers suggests that they spend progressively more time preparing the rest of the ongoing and upcoming line segment amidst executing a current movement. It should be noted that the on-line planning mode leads to resource re-distribution favoring the remaining part of the movement being executed. Since the

TABLE 8 | Pause duration in four experimental groups with different types of motor training.

	Dot-turn training		Non-dot-turn training		Total	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Correct angle training	830.02	478.01	1048.81	455.88	943.18	471.57
Incorrect angle training	1335.57	905.26	1139.95	647.48	1237.76	780.55
Total	1099.64	769.49	1095.85	555.65		

movement's starting point and line direction are set by already executed movement(s), i.e., by the already completed fraction of the line being drawn, it is the choice of a final position that becomes the focus of the planning process. In its turn, the final position becomes the starting position for the next line segment. Therefore, planning a final position for a current line segment might be accompanied by the planning of a specific angle for the next turn if a direction of the next line is also chosen.

We also observed a progressive growth in pause duration along the solution process for successful nine-dot problem solvers. This observation suggests that apart from on-line planning activity, these solvers also used off-line planning in multiple attempts to arrange sequences of line segments required for the nine-dot problem solution in the mental space. One might think that on-line and off-line modes of planning are mutually exclusive. Our results showed that this is not the case. Instead, solvers seem to rely on both modes of planning, with the heaviest use of both modes being observed at the late stages of the solution process. One might hypothesize then that using on-line mode of planning lays the ground for the successful use of off-line planning. Early inadequate representations of the nine-dot problem are constrained by certain perceptual templates (e.g., arrangement of nine dots inside the square area) which are used for off-line planning of line segments. Using on-line planning allows for relaxing these constraints and, after a while, it allows for lifting them altogether, thus clearing the way for the adequate off-line planning correct solution of the nine-dot problem.

EXPERIMENT 3

In the third experiment, we aimed to test the effectiveness of implicit hints on the solution of the nine-dot problem. Before the nine-dot problem, participants performed a preparatory task,

which included exact movements making up one of the possible solutions of the target problem. The preparatory task involved a serial reaction time task which masked target movements with intervening irrelevant movements, making the hint implicit. This task was widely used to study implicit motor learning (Nissen and Bullemer, 1987; Cleeremans et al., 1998). The typical paradigm usually includes several locations presented to a participant. In each trial, participants are asked to press as fast as possible a button corresponding to the location where the target stimulus appeared. If a sequence of target locations follows some complex regularity, participants demonstrate sensitivity to it (i.e., faster responses to regular vs. irregular target locations) but fail to report the regularity or even do not notice that there was some regularity at all. We expected that participants would implicitly learn the sequence, which in turn would lead to a higher probability of successful problem solving since the learned sequence constitutes the correct solution for the nine-dot problem.

Methods

Participants

Fifty-eight volunteers (47 women, 17–20 years old, $M = 18.0$, $SD = 0.71$) took part in the experiment. All of them were RANEP students and participated for a part of course credit.

This study was carried out in accordance with the recommendations of institutional guidelines of the ethics committee of the Department of Psychology of RANEP. The protocol was approved by the ethics committee of the Department of Psychology of RANEP. All participants gave written informed consent in accordance with the Declaration of Helsinki.

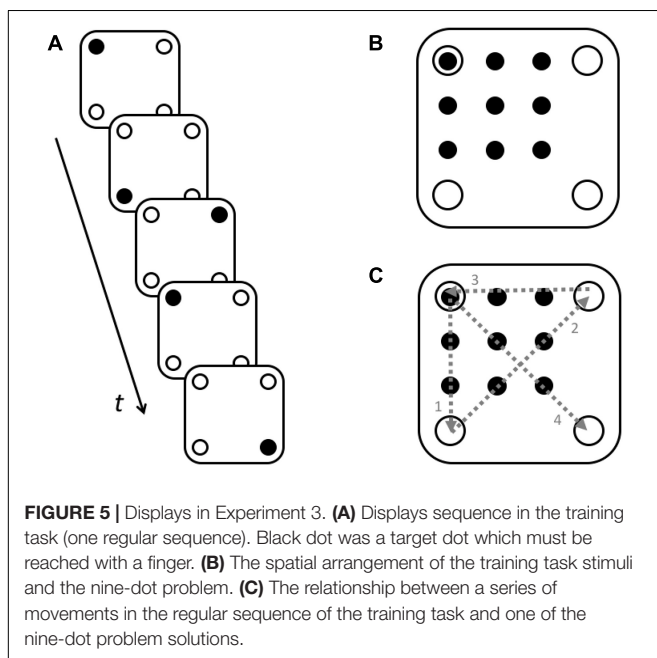
Apparatus and Stimuli

The nine-dot problem was administered the same way as in Experiments 1 and 2. However, it was preceded by an additional task. The setup was presented on a tablet using the same software as the abovementioned experiments. Participants were presented with a series of displays with four dots, three of which were empty and one – black. Participants had to trace a black dot moving their finger on the tablet's screen from old black dot position to a new black dot position (see **Figure 5**). The upper left dot was placed in the same position as the upper left dot in nine-dot problem. The other three dots were placed outside of the nine-dot square, but in those positions, which must be crossed in correct nine-dot problem solution.

Design and Procedure

Participants were told that they were going to solve several tasks. The first task was to catch the black dot among white dots with an index finger of the dominant arm. When the task was launched, participants were presented with the first display and had to start the task. When they touched the black dot, a new display appeared with the new position of the black dot. Participants were instructed to move the finger toward the new dot without lifting the finger from the screen.

Unbeknownst to participants, this task consisted of 60 series of 5 displays in each. Thirty series were regular (repeating the



same sequence of black dot positions) and thirty series were random (five displays presented the random position of black dot). Random and regular series followed one by one. The first series was random, then regular, then random and so on. The sequence of displays was programmed such that black dot did not appear in the same dot place twice in a row. The random series contained the same number of every position for the black dot as the regular series, for example if the regular series was 1-3-2-1-4, in the random series, black dot had to appear once in the first, third and fourth positions and twice in the first position. Thus, this task may be seen as a variant of a serial reaction time task (Nissen and Bullemer, 1987).

Participants were randomly distributed in two groups. In the first group ($N = 29$), regular series required participants to perform exactly the same movements that are needed for one of the successful solutions of nine-dot problem, thus we will refer to this group as “Relevant training” group (see **Figure 5**). The second group ($N = 29$) was divided into two subgroups ($N = 15$ and 14) with different regular series. In both subgroups, regular series contained another combination of movement which were useless in nine-dot problem solution, thus “Irrelevant training” group.

After that task, participants proceeded to the nine-dot problem. The procedure of the nine-dot problem solving was the same as in Experiment 1. In the end of the experiment, participants were asked whether they noticed any regularities in the first task. If they responded positively, they were asked to explain what sequence of dots they noticed.

Results

Learning

To evaluate learning, we deleted 1.5% of fastest and 1.5% of slowest responses for every participant. All the trials were averaged by blocks of 10 trials (2 series: random + regular) for every participant. The first block was deleted from the analysis as participants were very slow on the first trials. The first five trials were always random. Learning was assessed by fitting a linear regression with the number of blocks as predictor and RT as the dependent variable. The quadratic model indicated a better fit than a linear one ($F = 28.43$, $p < 0.001$), indicating the non-linear decrease of RTs with practice. The learning of regular sequence was examined by paired t -test (regular vs. random sequences), $t(57) = 7.15$, $p < 0.001$, indicating faster movements for regular sequences ($M = 566$ ms, $SD = 63$) in comparison to random sequences ($M = 584$ ms, $SD = 66$). None of the participants correctly reported the sequence of regular displays when asked.

The Effect of Training

The number of successful solutions in the Relevant training group was 7 (24.1%), and in the Irrelevant training group it was 14 (48.3%) (**Table 9**). The difference in the proportion of successful solutions in two groups did not reach significance according to Chi-squared test with Yates continuity correction, $\chi^2(1) = 2.69$, $p = 0.101$. To assess a non-specific effect of training, we compared solution rates in each group with the solution rates from Experiment 1 (52.8% successful solutions). The Relevant training group had significantly lower proportion of successful

solutions [$\chi^2(1) = 9.56$, $p = 0.002$], whereas the Irrelevant training group did not differ from the group of participants in Experiment 1 [$\chi^2(1) = 0.24$, $p = 0.626$].

Movement Time

As in previous experiments, we analyzed movement patterns in the nine-dot problem solution. First, we analyzed the difference in movement times between solvers and non-solvers. A 2×3 repeated measures ANOVA with SUCCESS (solvers and non-solvers) as a between-subjects factor and STAGE (first, second, and third) as a within-subjects factor revealed significant main effects of SUCCESS, $F(1,56) = 4.85$, $p = 0.032$, $\eta_p^2 = 0.08$, and STAGE, $F(2,112) = 6.60$, $p = 0.002$, $\eta_p^2 = 0.11$. The interaction between SUCCESS and STAGE was also significant, $F(2,112) = 15.48$, $p < 0.001$, $\eta_p^2 = 0.22$, indicating different dynamics in movement time in solvers and non-solvers across three stages. Pairwise comparisons using t -test revealed that there was no difference between solvers and non-solvers at the first ($p = 0.517$) stage. The difference was marginally significant at the second stage ($p = 0.056$) and significant at the third stage ($p = 0.006$), indicating that solvers gradually became slower than non-solvers (**Table 10**). **Figure 6** shows mean movement times for successful and unsuccessful solvers in all three stages of the nine-dot problem solving.

Three-way GROUP \times SUCCESS \times STAGE interaction was not significant ($p = 0.68$), indicating the similar pattern of results between two groups (see **Tables 11, 12**). Two-way GROUP \times SUCCESS and GROUP \times STAGE interactions were also non-significant ($p = 0.15$ and $p = 0.56$, respectively).

Pause Duration

The same model was run for pause duration between lines drawing. A 2×3 repeated measures ANOVA with SUCCESS (solvers and non-solvers) as between-subjects factor and STAGE (first, second, and third) as within-subjects factor revealed no significant main effects. The two-way interaction was significant, $F(2,112) = 12.14$, $p < 0.001$, $\eta_p^2 = 0.18$, indicating different dynamics in pause durations in solvers and non-solvers across three stages. By using a t -test for pairwise comparisons, we observed no significant difference between solvers and non-solvers at the first ($p = 0.345$) and second stages ($p = 0.165$). But we found that at the third stage solvers made significantly longer pauses than non-solvers, ($p = 0.008$) (**Table 13**). **Figure 6** shows means and corresponding confidence intervals of the pauses time for solvers and non-solvers in three stages of the nine-dot problem solution.

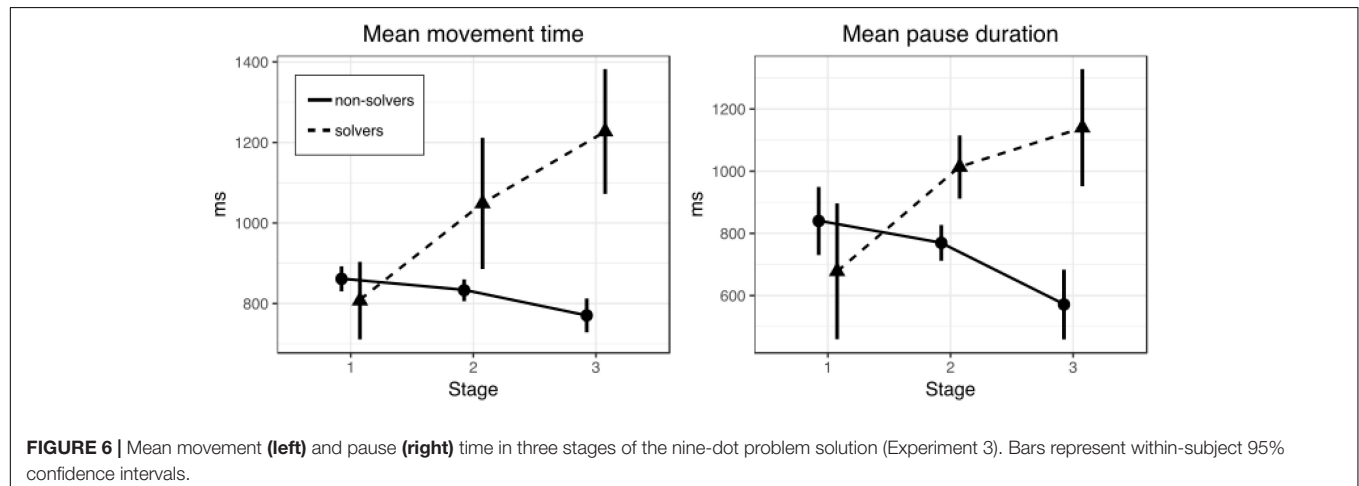
Then, we added GROUP factor (Relevant and Irrelevant training groups) to the model. The three-way interaction between

TABLE 9 | Solution rate in two experimental groups with relevant and irrelevant training.

	Non-solvers	Solvers	Total
Relevant training	22	7	29
Irrelevant training	15	14	29
Total	37	21	58

TABLE 10 | Mean and standard deviation of movement time in the three stages of the nine-dot problem solving.

	Solvers			Non-solvers			<i>M</i> _{stages}
	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	
First stage	807.32	371.7	1027.68	861.38	257.74	821.67	841.81
Second stage	1048.59	537.22		833.08	304.57		911.11
Third stage	1227.12	654.07		770.55	298.06		935.86

**TABLE 11** | Mean and standard deviation of movement time in the three stages of the nine-dot problem solving after relevant training.

	Solvers			Non-solvers			<i>M</i> _{stages}
	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	
First stage	648.93	216.11	843.37	877.16	282.1	823.62	822.01
Second stage	794.85	167.37		828.52	361.35		820.39
Third stage	1086.34	525.03		765.19	318.9		842.71

TABLE 12 | Mean and standard deviation of movement time in the three stages of the nine-dot problem solving after irrelevant training.

	Solvers			Non-solvers			<i>M</i> _{stages}
	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	
First stage	886.52	413.21	946.79	838.23	224.67	991.17	561.54
Second stage	1175.47	615.69		837.76	206.36		1001.83
Third stage	778.39	275.15		1297.51	717.59		1029

TABLE 13 | Mean and standard deviation of pause duration in the three stages of the nine-dot problem solving.

	Solvers			Non-solvers			<i>M</i> _{stages}
	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	<i>M</i>	<i>SD</i>	<i>M</i> _{total}	
First stage	678.01	398.82	943.72	839.94	717.77	726.9	781.31
Second stage	1013.28	682.07		769.43	606.88		857.72
Third stage	1139.79	851.31		571.33	368.81		777.15

“GROUP × SUCCESS × STAGE” was not significant ($p = 0.956$), thus indicating a similar pattern of results between two groups (see **Tables 14, 15**). Two-way “GROUP × SUCCESS” and “GROUP × STAGE” interactions were also not significant ($p = 0.935$ and $p = 0.129$, respectively).

Discussion

In Experiment 3, we aimed to test, whether non-specific movement training would result in a change of the nine-dot problem solution. During training, participants performed regular sequential movements more quickly than irregular, which

means that movements series was learned by them. In the Relevant training group, the regular sequence was identical to one of the solutions of the nine-dot problem, and as such, we expected that participants in this group would be more successful in the nine-dot problem. However, this was not the case as the Irrelevant training group participants solved the task successfully more often than Relevant training participants. Further statistical analysis showed, however, that this difference was not significant. In comparison to the Experiment 1, which had identical nine-dot problem session, the Irrelevant training group showed no significant difference in solution rates, whereas the Relevant training group had the significantly lower proportion of successful solutions than in Experiment 1. We don't think this result can be explained by the non-specific effect of training. A more probable interpretation is related to the overall lower solution rate in both groups in Experiment 3 than in Experiment 1. We then analyzed movement time and pause time depending on the solution success and group. In both cases, we observed the interaction between solution success and solution stage. Solvers tended to increase both movement times and pause times whereas non-solvers tended to decrease both movement and pause times. Training type (relevant to the nine-dot problem solution or not) did not affect movement and pauses times.

The latter result (i.e., the finding that preliminary motor training involving an irrelevant task does not influence motor activity during nine-dot problem solving) suggests that no transfer of the correct sequence of line segments acquired during the implicit learning session occurred during the solving of the nine-dot problem. The fact that participants did learn the correct sequence of movements while performing some irrelevant task is in accordance with the view that this sequence of movements played a purely instrumental role while approaching the target problem. However, the merely instrumental role played by motor activities was insufficient to target problem solving since no transfer of the learned sequence to the nine-dot problem was found. Therefore, we can argue that for the successful resolution of the nine-dot problem, the motor activity should also play a functional role.

GENERAL DISCUSSION

An Overview of Major Findings

Based on a preliminary theoretical analysis, we assumed that investigating on-line vs. off-line motor planning separately might be helpful in explaining the difference between successful and unsuccessful solvers of the nine-dot problem. We computed two quantities which are sensitive to the difference between on-line and off-line planning, the movement time and the pause duration, and then used them in order to compare successful and unsuccessful solvers of the nine-dot problem.

We reported three experiments in this study: Experiment 1 through Experiment 3, all of which showed similar results. All three showed that at the third stage of the solution process (the final one third of the block of trials) the successful solvers showed longer movement time than their unsuccessful counterparts. In Experiment 2, test takers also undertook a preliminary motor training prior to the test session. In this case, successful problem solvers slowed down their movements not only during the final third stage but also during the intermediate second stage. Also, our results indicate that successful problem solvers showed longer between-movement pauses at the final third stage in both Experiments 2 and 3 and at the intermediate second stage in Experiment 2. This result is in accordance with the critical role of the mental lookahead in finding the nine-dot problem solution, a theoretical position formulated by MacGregor et al. (2001). In agreement with the aforementioned study, our results show the increasing involvement of off-line planning (which is similar to the mental lookahead) at the late stages of the nine-dot problem solving.

Results of Experiment 2 do not support our assumption on the greater positive effect of practicing a non-dot turn with the relevant to the problem solution angle of 45 degrees over non-dot turn with an irrelevant angle of 26.6 degrees. Practicing non-dot turns of arbitrary angle actually caused some increase in the rate of successful solutions of the nine-dot problem. This result is in line with the empirical evidence showing an important role that non-dot turns play in successful solution of the nine-dot problem

TABLE 14 | Mean and standard deviation of pause duration in the three stages of the nine-dot problem solving after relevant training.

	Solvers			Non-solvers			
	<i>M</i>	<i>SD</i>	<i>M_{total}</i>	<i>M</i>	<i>SD</i>	<i>M_{total}</i>	<i>M_{stages}</i>
First stage	645.58	402.28	822.34	852.38	863.21	663.21	802.46
Second stage	836.66	646.56		662.2	606.62		704.31
Third stage	984.77	758.8		475.06	343.22		598.09

TABLE 15 | Mean and standard deviation of pause duration in the three stages of the nine-dot problem solving after irrelevant training.

	Solvers			Non-solvers			
	<i>M</i>	<i>SD</i>	<i>M_{total}</i>	<i>M</i>	<i>SD</i>	<i>M_{total}</i>	<i>M_{stages}</i>
First stage	694.23	411.3	1004.37	821.68	454.39	820.09	760.15
Second stage	1101.58	705.39		926.07	592.04		1011.13
Third stage	1217.3	910.95		712.51	370.34		

(Kershaw and Ohlsson, 2004; Öllinger et al., 2014). Results of Experiment 3 did not confirm our assumption. We expected that preliminary learning a motor pattern corresponding to a fraction of the nine-dot problem solution would help in solving this problem. However, the results of Experiment 3 suggest that learning a correct sequence of movements in the context of an irrelevant task does not affect a process of the nine-dot problem solving.

The Impact of Preliminary Motor Training on the Solution of the Nine-Dot Problem

It has been shown that preliminary motor training involving practicing different fractions of the correct solution of the nine-dot problem results in growing effectiveness of solving that problem (Weisberg and Alba, 1981; Lung and Dominowski, 1985). Kershaw and Ohlsson (2004) have come to a similar conclusion specifically regarding non-dot turns. We exploited two kinds of the preliminary motor training, a traditional one, which involved both instrumental and functional role of motor activity (problem solvers were practicing non-dot turns of 45 and 26.6 degrees), and another “implicit” training (participants implicitly learned a sequence of movements corresponding to a correct solution of the nine-dot problem) that took place during multiple attempts to perform an irrelevant task with hidden relevance to the target nine-dot problem. In the latter case, it turned out that the motor activity played an exclusively instrumental role in solving of the target problem.

The results obtained in the present study suggest that a preliminary training causes an increase in effectiveness of the nine-dot problem solving only if the movements involved in this training play a functional role in the solving of the nine-dot problem. It turned out that practicing non-dot turns regardless of their angle boosted the effectiveness of the solving process while the preliminary training, in which motor activity played an instrumental role only, did not affect the percentage of the correct solution of the nine-dot problem.

The Role of On-Line and Off-Line Planning in the Process of the Nine-Dot Problem Solving

A difference between successful and unsuccessful problem solvers allows for understanding what helps the successful solvers to solve the nine-dot problem. The obtained results from the abovementioned experiments provide valuable information for the analysis of the specific role of the on-line and off-line movement planning modes in the process of solving of that problem as well as their relative contribution to the successful problem solution.

There are two decisions that are to be made during the nine-dot problem solving: a problem solver has to select initial and final finger positions. However, this may be done in two modes. A problem solver might arrange a plan for upcoming motor activity (hand drawing the line segments connecting the dots) by arranging a certain sequence of line segments. These arrangements, i.e., off-line planning, occur in the mental space.

The off-line planning has a “long horizon,” meaning that several steps are being planned (MacGregor et al., 2001; Chronicle et al., 2004). However, this process goes in the well-established perceptual framework and does not transcend it. This way of movement planning does not help to go beyond the nine dots area because problem solvers usually select one of the visible dots as the movement final position. The second mode of motor planning is that the planning and execution processes go in parallel, which slows down the overt line drawing. In this case, a problem solver first chooses an initial position and then selects a direction of upcoming motion while the selection of a final position is temporarily postponed. During this process of slow line drawing a problem solver considers a wide range of possible final positions including those outside the visible nine dots area. This mode of motor planning has a wide but short horizon.

The two modes of motor planning, off-line and on-line modes, are not mutually exclusive. At the later stages of the solution process, an intensity of involvement of both planning modes is greater in successful than in unsuccessful problem solvers. Thus, one may infer that both modes of motor planning are required in order to successfully solve the nine-dot problem, each mode playing its specific role. One may hypothesize that the involvement of on-line planning mode gradually modifies the way by which the off-line planning mode operates. At the early stages of the solving process, the off-line planning is constrained by the initial perceptual description of the problem, i.e., its early representation. For example, relying exclusively on the spatial positions of nine dots and their specific arrangement in the form of square leads to all the planned movements start and end positions coincide with the visible dots and reside within the square area. Relying on on-line planning helps to gradually overcome these perceptual constraints, which in turn opens a way for adequate off-line planning and as a result of a successful solution of the nine-dot problem. All the above considerations lead to a conclusion that motor activity in its functional role is crucial for solving the nine-dot problem.

In order to account for the experimental results reported in the present work, we considered the role of two modes of the motor planning, the off-line and on-line modes. We believe that this approach can be generalized to those insight problems whose solutions substantially rely on some form of motor activity (the examples of problems of that sort were mentioned above). Substantial similarities can be found in all problems of that kind. At the early stages of the problem-solving process, an inadequate initial representation of the problem leads to activation of irrelevant motor programs which effectively hinder from finding the problem solution. As an example, an inadequate initial representation of the six matches problem leads to that solvers attempt to solve the problem (i.e., to arrange four equal triangles using six matches) by keeping all possible rearrangements of the matches confined to a single plane (Scheerer, 1963). A correct solution requires arranging matches into a tetrahedron in the three-dimensional space. Initial attempts to solve yet another insight problem, the 8-coin problem, are limited by moving coins

along the plane whereas the correct solution requires leaving the plane for the three-dimensional space (Ormerod et al., 2002). Relying on the on-line mode of the motor planning while solving the above-mentioned problems, like in the case of the nine-dot problem, could help to overcome the inadequate initial representations of these problems and allow the solvers to operate in the three-dimensional space. Of course, this possibility requires an experimental verification (see section Future Directions).

The results obtained in this work cannot be easily accounted for by dominant theories of insight problem solving. The representational change theory is based on the chunk decomposition, reencoding, elaboration and constraint relaxation as the major mechanisms of the insight problem solving (Schooler et al., 1993; Knoblich et al., 1999). In the framework of the theory, these mechanisms operate on the mental representation alone while any motor activity is considered in its pure instrumental role as a means for expressing the solution in the physical world. The major mechanisms considered in the framework of the criterion for satisfactory progress theory, are also purely mental upon their nature. They are closely related to the solvers' horizon of planning (lookahead) (MacGregor et al., 2001). Later, the lookahead concept has been linked to the spatial memory span (Chein et al., 2010). Note that neither of the theories predicts the change in the motor activity along the course of the insight problem solving.

One of the sources of the difficulty of the nine-dot problem traditionally considered in the literature is that during initial attempts to solve the problem the motor output is affected by irrelevant perceptual constraints imposed primarily by the square arrangement of the dots (Maier, 1930; Scheerer, 1963). We showed that successful solvers employ on-line planning for shaping their motor output and therefore that relying exclusively on the off-line planning mode is insufficient for reaching success. The relaxation of the negative impact of the perceptual grouping constraints takes place because of the influence the motor processes exert onto perceptual ones. This kind of motor-to-perception influence provides a new example of the functional role of motor activity during insight problem solving. We suggest that relying on the on-line motor planning constitute yet another possible mechanism of solving insight problems.

Methodological Innovations of the Present Study

An attempt to study the role of motor activity in the process of solving the nine-dot problem and other insight problems faces a difficulty: a lack of dependent variables quantifying the motor activity. In order to overcome the difficulty, in the present study, we modified a traditional way of presenting the problem and scoring the solving process. In our study, participants were asked to draw line segments with the tip of the index finger on the surface of a tablet computer. The graphical movements were recorded using the specially designed custom-made software. Then, the set of recordings corresponding to multiple attempts to solve the problem were analyzed with a

semi-automatic algorithm which is capable of breaking some entire recordings onto partially overlapping submovements. This allowed for separating periods of motion from the pauses between them and computing numerical estimates for movement times and pause durations. The obvious benefit of using such detailed description of solving-related motor activities is that it can be used to study the time course of the solution process.

The method that we applied allows for systematic comparison between successful vs. unsuccessful problem solvers based on the quantitative parameters of their motor activity. Using this method, we found new specific patterns of motor activity that differentiate successful and unsuccessful solvers. We hope that our approach would be helpful in further investigations of the functional role of motor activity in insight problem solving.

Limitations

The limitations of this study include a relatively small sample size and its predominantly female composition. Besides, the study is limited to analyzing the only one problem – nine-dot problem. Another limitation of the present study was that we did not verify whether solutions demonstrated by the participants were indeed insight solutions.

Future Directions

The proposed method makes it possible to implement several research directions. First, it seems reasonable to compare the process of solving various types of insight problems involving the motor component (for example, 6-coin, 8-coin, 6 matches etc.) from the perspective of the successful and unsuccessful solvers of the modes of motor planning. Second, a valuable contribution to understanding the mechanisms of insight problem solving would be identifying and analyzing the individual strategies in the course of solving these problems. Third, in order to uncover the details of the mechanisms of insight problem solving it worth to compare the impact of various experimental interventions (motor, oculomotor, verbal, etc.) in the form of prompting, priming or preliminary training on the process of solving insight problems involving the motor component. Finally, the mechanisms underlying the insight problem solving could be studied by comparing the parameters of motor activity shown by expert versus novice solvers. It is also interesting to compare the results obtained with the new method with the results of more traditional methods of fixating the process of solving insight problems (eye movements fixated with an eye-tracker, verbal protocols, video recording).

AUTHOR CONTRIBUTIONS

VS planned experiments, theoretical analysis of the results, and wrote the text of the article. NL conducted experiments, processed the data, theoretical analysis of the results, and wrote the text of the article. II conducted experiments, processed the data, and wrote the text of the article. AK processed the data, theoretical analysis of the results, and wrote the text of the article.

REFERENCES

- Anzai, Y., and Simon, H. A. (1979). The theory of learning by doing. *Psychol. Rev.* 86, 124–140. doi: 10.1037/0033-295X.86.2.124
- Chein, J. M., Weisberg, R. W., Streeter, N. L., and Kwok, S. (2010). Working memory and insight in the nine-dot problem. *Mem. Cognit.* 38, 883–892. doi: 10.3758/MC.38.7.883
- Chronicle, E., MacGregor, J., and Ormerod, T. (2004). What makes an insight problem? The roles of heuristics, goal conception, and solution recoding in knowledge-lean problems. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 14–27. doi: 10.1037/0278-7393.30.1.14
- Chronicle, E. P., Ormerod, T. C., and MacGregor, J. N. (2001). When insight just won't come: the failure of visual cues in the nine-dot problem. *Q. J. Exp. Psychol.* A 54, 903–919. doi: 10.1080/713755996
- Cleeremans, A., Destrebecqz, A., and Boyer, M. (1998). Implicit learning: News from the front. *Trends Cogn. Sci.* 2, 406–416. doi: 10.1016/S1364-6613(98)01232-7
- Grant, E. R., and Spivey, M. J. (2003). Eye movements and problem solving: guiding attention guides thought. *Psychol. Sci.* 14, 462–466. doi: 10.1111/1467-9280.02454
- 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
- Kershaw, T. C., and Ohlsson, S. (2004). Multiple causes of difficulty in insight: the case of the nine-dot problem. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 3–13. doi: 10.1037/0278-7393.30.1.3
- 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, 1534–1555. 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
- Köhler, W. (1921). *Intelligenzprüfungen an Menschenaffen [The Mentality of Apes]*. Berlin: Springer-Verlag. doi: 10.1007/978-3-642-47574-0
- Korneev, A. A., and Kurgansky, A. V. (2013). Internal representation of movement sequences built upon either static line drawings or trajectories of moving object. *Zhurnal vysshei nervnoi deiatel'nosti imeni I P Pavlova* 63, 437–450. doi: 10.7868/S0044467713040060
- Luchins, A. S. (1942). Mechanization in problem solving: The effect of *Einstellung*. *Psychol. Monogr.* 54, i-95. doi: 10.1037/h0093502
- Lung, C. T., and Dominowski, R. L. (1985). Effects of strategy instructions and practice on nine-dot problem solving. *J. Exp. Psychol. Learn. Mem. Cogn.* 11, 804–814. doi: 10.1037/0278-7393.11.1-4.804
- MacGregor, J. N., Ormerod, T. C., and Chronicle, E. P. (2001). Information processing and insight: a process model of performance on the nine-dot and related problems. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 176–201. doi: 10.1037/0278-7393.27.1.176
- Maier, N. R. F. (1930). Reasoning in humans. I. On direction. *J. Comp. Psychol.* 10, 115–143. doi: 10.1037/h0073232
- Nissen, M. J., and Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cogn. Psychol.* 19, 1–32. doi: 10.1016/0010-0285(87)90002-8
- Ohlsson, S. (1984). Restructuring revisited: II. An information processing theory of restructuring and insight. *Scand. J. Psychol.* 25, 117–129. doi: 10.1111/j.1467-9450.1984.tb01005.x
- Öllinger, M., Fedor, A., Brodt, S., and Szathmáry, E. (2017). Insight into the ten-penny problem: guiding search by constraints and maximization. *Psychol. Res.* 81, 925–938. doi: 10.1007/s00426-016-0800-3
- 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. (2014). The dynamics of search, impasse, and representational change provide a coherent explanation of difficulty in the nine-dot problem. *Psychol. Res.* 78, 266–275. doi: 10.1007/s00426-013-0494-8
- Orban de Xivry, J. J., Legrain, V., and Lefèvre, P. (2017). Overlap of movement planning and movement execution reduces reaction time. *J. Neurophysiol.* 117, 117–122. doi: 10.1152/jn.00728.2016
- 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
- Rhodes, B. J., Bullock, D., Verwey, W. B., Averbach, B. B., and Page, M. P. (2004). Learning and production of movement sequences: behavioral, neurophysiological, and modeling perspectives. *Human Move. Sci.* 23, 699–746. doi: 10.1016/j.humov.2004.10.008
- Scheerer, M. (1963). Problem-solving. *Sci. Am.* 208, 118–132. doi: 10.1038/scientificamerican0463-118
- 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
- Scott, S. H. (2012). The computational and neural basis of voluntary motor control and planning. *Trends Cogn. Sci.* 16, 541–549. doi: 10.1016/j.tics.2012.09.008
- Scott, S. H., Cluff, T., Lowrey, C. R., and Takei, T. (2015). Feedback control during voluntary motor actions. *Curr. Opin. Neurobiol.* 33, 85–94. doi: 10.1016/j.conb.2015.03.006
- Stanford, T. R. (2013). “Decision making, motor planning,” in *Encyclopedia of Computational Neuroscience*, eds D. Jaeger and R. Jung (New York, NY: Springer).
- Tessari, A., Bosanac, D., and Rumiati, R. I. (2006). Effect of learning on imitation of new actions: implications for a memory model. *Exp. Brain Res.* 173, 507–513. doi: 10.1007/s00221-006-0395-9
- Tessari, A., and Rumiati, R. I. (2004). The strategic control of multiple routes in imitation of actions. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 1107–1116. doi: 10.1037/0096-1523.30.6.1107
- Thomas, L. E., and Lleras, A. (2009). Swinging into thought: directed movement guides insight in problem solving. *Psychon. Bull. Rev.* 16, 719–723. doi: 10.3758/PBR.16.4.719
- Weisberg, R. W. (1995). “Prolegomena to theories of insight in problem solving: a taxonomy of problems,” in *The Nature of Insight*, eds R. J. Sternberg and J. E. Davidson (Cambridge, MA: MIT Press), 157–196.
- Weisberg, R. W., and Alba, J. W. (1981). An examination of the alleged role of “fixation” in the solution of several “insight” problems. *J. Exp. Psychol. Gen.* 110, 169–192. doi: 10.1037/0096-3445.110.2.169
- Werner, K., and Raab, M. (2013). Moving to solution: effects of movement priming on problem solving. *Exp. Psychol.* 60, 403–409. doi: 10.1027/1618-3169/a000213
- Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi: 10.3758/BF03196322
- Wong, A. L., and Haith, A. M. (2017). Motor planning flexibly optimizes performance under uncertainty about task goals. *Nat. Commun.* 8:14624. doi: 10.1038/ncomms14624
- Wong, A. L., Haith, A. M., and Krakauer, J. W. (2015). Motor planning. *Neuroscientist* 21, 385–398. doi: 10.1177/1073858414541484

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 Spiridonov, Loginov, Ivanchei and Kurgansky. 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.



Normative Data for 84 UK English Rebus Puzzles

Emma Threadgold^{1*†}, John E. Marsh^{1,2†} and Linden J. Ball^{1†}

¹ School of Psychology, University of Central Lancashire, Preston, United Kingdom, ² Department of Building, Energy and Environmental Engineering, University of Gävle, Gävle, Sweden

OPEN ACCESS

Edited by:

Kathryn Friedlander,
University of Buckingham,
United Kingdom

Reviewed by:

Ana-Maria Olteteanu,
Freie Universität Berlin, Germany
Steven M. Smith,
Texas A&M University, United States
Carola Salvi,
Northwestern University, United States

*Correspondence:

Emma Threadgold
ethreadgold1@uclan.ac.uk

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 09 July 2018

Accepted: 26 November 2018

Published: 13 December 2018

Citation:

Threadgold E, Marsh JE and Ball LJ
(2018) Normative Data for 84 UK
English Rebus Puzzles.
Front. Psychol. 9:2513.
doi: 10.3389/fpsyg.2018.02513

Recent investigations have established the value of using rebus puzzles in studying the insight and analytic processes that underpin problem solving. The current study sought to validate a pool of 84 rebus puzzles in terms of their solution rates, solution times, error rates, solution confidence, self-reported solution strategies, and solution phrase familiarity. All of the puzzles relate to commonplace English sayings and phrases in the United Kingdom. Eighty-four rebus puzzles were selected from a larger stimulus set of 168 such puzzles and were categorized into six types in relation to the similarity of their structures. The 84 selected problems were thence divided into two sets of 42 items (Set A and Set B), with rebus structure evenly balanced between each set. Participants ($N = 170$; 85 for Set A and 85 for Set B) were given 30 s to solve each item, subsequently indicating their confidence in their solution and self-reporting the process used to solve the problem (analysis or insight), followed by the provision of ratings of the familiarity of the solution phrases. The resulting normative data yield solution rates, error rates, solution times, confidence ratings, self-reported strategies and familiarity ratings for 84 rebus puzzles, providing valuable information for the selection and matching of problems in future research.

Keywords: problem solving, insight, rebus, norming, test validation

INTRODUCTION

Problem solving involves thinking activity that is directed toward the achievement of goals that are not immediately attainable (e.g., Newell and Simon, 1972). It is a central aspect of human cognition that arises across a range of contexts, from everyday activities to the attainment of major scientific advancements and the achievement of important technological, cultural, and artistic developments. Although problem solving can be fairly mundane (e.g., deciding what to make for your evening meal) it can also lead to solutions that are highly creative (e.g., a delicious new dish prepared by a master chef). This latter kind of “creative problem solving” is distinguished from other types of problem solving in that it involves the generation of solutions that are both original *and* effective, with the sole presence of either attribute being insufficient for a solution to be deemed creative (see Runco, 2018). Not surprisingly, creative problem solving is held in especially high regard in all areas of real-world practice.

Research on creative problem solving has burgeoned over the past 20 years, with a traditional assumption being that people solve such problems in one of two different ways, that is, either (i) through *analytic* processes, which involve conscious, explicit thinking that takes the solver closer to a solution in a slow, step-by-step manner (e.g., Fleck and Weisberg, 2004; Ball and Stevens, 2009); or (ii) through *insight* processes, which involve non-conscious, implicit thinking that gives rise

to a sudden and clear realization of how to make progress toward a solution (e.g., Sternberg and Davidson, 1995; Bowden and Jung-Beeman, 1998, 2003a; Jung-Beeman et al., 2004). According to the latter view, such flashes of insight are typically characterized as involving a major change in the representation of a problem, arising from largely tacit processes of problem elaboration, recoding or constraint relaxation (e.g., Ohlsson, 1992, 2011; Knoblich et al., 1999; see also Bowden et al., 2005).

Notwithstanding the possibility that creative problem solving can, in principle, occur in two distinct ways (i.e., either via explicit, analytic processes or via implicit, insight processes) the emerging consensus is that a good deal of the time people probably deploy a mix of both conscious analysis and non-conscious insight when tackling creative problems (e.g., Barr et al., 2014; Sowden et al., 2014; Gilhooly et al., 2015; Weisberg, 2015, 2018; Barr, 2018). This move away from polarized views of creative problem solving as involving *either* analytic processes or insight processes marks an important change in recent theorizing, which over the past couple of decades has tended to become sidetracked by rather narrow and somewhat esoteric debates focused on a very limited set of tasks and paradigms.

The welcome emergence of more nuanced and encompassing theories of creative problem solving has arguably been fueled not only through improved theory-driven experimentation (including neuroscientific studies; for a recent review see Shen et al., 2017), but also through the availability of a greater variety of problem-solving tasks that can be used by researchers in laboratory-based studies of problem-solving phenomena. This means that nowadays researchers are not just reliant on so-called “classic” insight tasks that often have their origins in Gestalt studies of problem solving (e.g., Duncker, 1945, candle problem or Maier, 1930, nine-dot problem), but that they can also make use of many other problems that may be solved to varying degrees by analysis or insight, such as remote associate tasks (RATs) (e.g., Mednick, 1968), matchstick algebra problems (e.g., Knoblich et al., 1999), magic tricks (e.g., Danek et al., 2014a,b) and rebus puzzles (e.g., MacGregor and Cunningham, 2008; Salvi et al., 2015), which are the focus of the present paper.

Classic insight problems suffer from a number methodological issues that have arguably limited their value in advancing an understanding of creative problem solving (for relevant arguments see Bowden and Jung-Beeman, 2003a; MacGregor and Cunningham, 2008). Most notably, there is a restricted pool of such classic insight problems from which researchers can draw, which means that studies using these problems often involve only a small number of items. In addition, classic insight problems can be very difficult to solve, with very few participants achieving a correct solution without some sort of hint being provided. Moreover, problem-solving times can be lengthy, often taking up to 10 min per problem. Classic insight problems are also heterogeneous and prone to being influenced by confounding variables (e.g., the amount of time that is available for solution generation itself is an important confounding factor that is often overlooked in theorizing; but see Ball et al., 2015). These problems may also yield ambiguous solutions that are difficult to quantify.

As an alternative to classic insight problems, researchers have turned in recent years toward the extensive use of compound remote associates (CRA) problems, which are conceptual descendants of the RAT first developed by Mednick (1968). CRA problems involve presenting participants with three words (e.g., pine, crab, sauce) for which they are required to produce a solution word which, when combined with the three words, generates three compound words or phrases (i.e., pineapple, crab apple, apple sauce). CRA problems have significant advantages over classic insight problems. Although variation of problem difficulty exists within CRA sets (Bowden and Jung-Beeman, 2003a; Salvi et al., 2015) they are comparatively easy to solve, fast to administer, more resistant to potentially confounding variables and typically yield unambiguous solutions (Bowden and Jung-Beeman, 2003a; MacGregor and Cunningham, 2008; but see Howe and Garner, 2018). Importantly, too, it is possible to construct a large number of CRA problems, as has recently been demonstrated by Olteteanu et al. (2017), who used computational methods to generate a repository of around 17 million American English CRA items based on nouns alone and meeting tight controls. Furthermore, CRA problems can be presented in compressed visual areas, rendering the problems suitable for electroencephalography (EEG; e.g., Bowden and Jung-Beeman, 2003a,b; Sandkühler and Bhattacharya, 2008) and functional magnetic resonance (fMRI; e.g., Bowden and Jung-Beeman, 2003a,b) procedures. In addition, CRA problems allow for control over stimulus presentation and response timing (e.g., Bowden and Jung-Beeman, 2003a) and lend themselves well to priming paradigms in which primes (e.g., Howe et al., 2016) solution hints (e.g., Smith et al., 2012) or solution recognitions can be presented across or within hemispheres (e.g., Bowden and Jung-Beeman, 2003b).

In the present paper we focus on rebus puzzles (e.g., MacGregor and Cunningham, 2008; Salvi et al., 2015), which are starting to feature more commonly in problem-solving research and have many of the benefits of CRAs, as well as some additional advantages. Rebus puzzles involve a combination of visual, spatial, verbal, or numerical cues from which one must identify a common phrase or saying. As an example, the rebus problem “BUSINES,” when correctly interpreted, yields the common phrase “Unfinished Business.” Such rebus problems have been used in research on creative problem solving-processes such as studies of fixation and incubation phenomena (Smith and Blankenship, 1989), with rebus problem-solving success also having been shown to be positively correlated with performance on remote associate problems, whilst being independent of general verbal ability (MacGregor and Cunningham, 2008).

Rebus puzzles are relatively easy to present to participants and have only single “correct” answers, which means that responses are straightforward to score. Importantly, however, the problems are moderately challenging to solve, although they are often solvable with persistent effort. The difficulty of rebus puzzles may arise, in part, from there being many ways in which they can be tackled (cf. Salvi et al., 2015), but may also be a consequence of the problem information initially *misdirecting* solution efforts because the solver draws upon implicit assumptions derived from the experience of normal reading (Friedlander and Fine,

2018, similarly suggest that normal reading may engender misdirection when solving cryptic crossword clues). Such self-imposed constraints may lead solvers to reach a point of *impasse*, where solution progress is not forthcoming, with such *impasse* needing to be circumvented by problem restructuring (see MacGregor and Cunningham, 2008; Cunningham et al., 2009). The challenges for solving rebus puzzles that arise from tacit, self-imposed assumptions can readily be seen in the rebus example “CITY,” whose solution is “capital city.” The font of the presented text is a superficial feature that would usually be ignored in normal reading, despite potentially carrying figurative meaning in the context of a rebus puzzle. Indeed, the difficulty of a rebus problem is believed to be a function of the number of implicit assumptions that need to be broken (MacGregor and Cunningham, 2008, 2009).

Another factor that makes rebus problems useful in problem-solving research is the observation that solvers often cannot report the details of the preceding processing that led to a solution, which is especially likely when such solutions are accompanied by an “Aha!” experience that is suggestive of an insight-based problem-solving process (MacGregor and Cunningham, 2008). Notwithstanding the fact that rebus puzzles can be solved via implicit, insight processes, there is also evidence that they are open to solution via analysis as well or a varying combination of both analysis and insight (MacGregor and Cunningham, 2008, 2009).

In sum, rebus puzzles offer a means by which a large pool of homogenous problems of different difficulty can be administered within a single session in order to investigate the processes of analysis and insight that underpin creative problem solving. Such rebus puzzles are rapid to administer and relatively under-represented in the problem-solving literature in comparison to tasks such as CRA problems. Despite the increasing use of rebus puzzles in problem-solving research, there exists very limited normative data relating to such problems in relation to their solution rates, solution times and phenomenological characteristics, with current norming studies being restricted (as far as we are aware) to the validation of a set of Italian rebus puzzles (Salvi et al., 2015). The lack of normative data is problematic given that rebus puzzles are linguistically context dependent, relating, as they do, to common words, sayings or phrases that exist in a particular language, including idiomatic expressions that have become culturally conventionalized. Language-specific normative data are, therefore, vital for advancing the use of rebus puzzles in problem-solving studies so that researchers can have confidence that the problems that they select for their experiments have desired characteristics to enable specific research questions to be studied.

To address the absence of normative data for English rebus puzzles, this paper presents normative data for 84 rebus items that are underpinned by common United Kingdom (UK) English phrases or sayings. The normative data that we obtained provide details of typical solution rates, error rates, and correct solution times (seconds) as well as standard deviations for all solution times. In addition, we obtained ratings of participants' confidence in their solutions, their familiarity with the solution

phrases as well as a self-report measure of the extent to which participants felt that they had solved the problem via a process of analysis vs. insight. The latter data were elicited to align with the emerging theoretical consensus that it is useful to view creative problem solving as involving a mix of processes that fall along a continuum ranging from analysis to insight.

We further note that an inspection of rebus puzzles revealed to us that there are several specific sub-types that involve very similar solution principles. This was also highlighted in the set of Italian rebus puzzles reported by Salvi et al. (2015), in which they identified 20 categories for the subset of rebus problems. On inspection of the UK English rebus puzzles, we categorized the puzzles into substantially fewer categories based on an observation of the specific solution principles that underpin these rebus items. We categorized the 84 rebus puzzles that we wished to norm into six specific categories relating to their structure and the types of cues necessary to solve each problem.

METHOD

Participants

The study involved 170 participants in total (125 female) with an age range of 19 to 70 years ($M = 36$ years-old, $SD = 12$ years). Participants received £3 in exchange for 30 min participation time and were recruited via the survey recruitment website “Prolific Academic.” Participants completed one rebus set each. All participants were UK nationals and native English speakers. This study was carried out in accordance with the recommendations of the British Psychological Society Code of Human Research Ethics. The protocol was approved by the Psychology and Social Work Ethics Committee (Ref: 397) at the University of Central Lancashire, UK. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Design

A total of 84 rebus puzzles were collated and divided into two equal sets, with 42 rebus puzzles per set (see **Appendix A** in Supplementary Material for the Set A items, **Appendix B** in Supplementary Material for the Set B items and **Appendix C** in Supplementary Material for three practice items that were used in the study). For each rebus puzzle, a normative solution rate and mean solution time (in seconds) was obtained. The maximum available solution time per item was 30 s. The dependent variables were the solution rate and solution time for each rebus, an error rate, a measure of confidence in the accuracy of the response to each rebus, and a measure of the extent to which each answer was solved via a process of analysis or insight. The confidence measure and the measure of analysis/insight phenomenology were each elicited using continuous sliding scales that participants used to register a response, resulting in scores ranging from 1 to 100. Thus, a higher score indicated a more confident response on the confidence scale and a more “insight-like” response on the analysis/insight scale. Furthermore, each rebus puzzle was allocated to one of six categories based on their underpinning solution principles, with these categories having been developed

for initial classification purposes (please refer to the Materials section below for a discussion of the development of these categories).

Materials

Rebus Puzzles

An initial set of 186 rebus puzzles were selected from copyright-free sources on the internet. It was ensured that the rebus puzzles all related to familiar UK English phrases, with the removal of any specifically American phrases. On inspection of the set of rebus items, it became clear that there were many common structural features across the puzzles. Therefore, six puzzle categories were developed to which each rebus item could be allocated so as to ensure that different types of rebus were presented in a balanced manner across item sets (see **Appendices A, B** in Supplementary Material). The six rebus categories that were developed are as follows: (1) a word, picture or number over another word, picture or number (for an example item see **Appendix A** in Supplementary Material, Item 1—“*feeling on top of the world*”); (2) a word, picture or number under another word, picture or number (see **Appendix A** in Supplementary Material, Item 5—“*try to understand*”); (3) a word presented within another word (see **Appendix A** in Supplementary Material, Item 6—“*foot in the door*”); (4) a play on words with numbers (see **Appendix A** in Supplementary Material, item 16—“*forty winks*”); (5) imagery (see **Appendix A** in Supplementary Material, Item 20—“*half hearted*”); and (6) spatial (see **Appendix A** in Supplementary Material, Item 36—“*parallel bars*”).

Drawing from the initial set of 186 rebus puzzles, each puzzle was allocated by two independent judges to one of the six constructed rebus categories. An inter-rater reliability analysis was then undertaken utilizing the Kappa statistic (Viera and Garrett, 2005) to determine the overall consistency in rebus categorization between the two judges. There was a statistically significant moderate agreement between the two judges, $\kappa = 0.59$ (95% CI = 0.50 to 0.67, $p < 0.001$). It was also evident from viewing the rebus puzzles that a number of them might be deemed to cross two or more categories. To account for this, and utilizing the Kappa scores, rebus items were selected for the norming study only when category agreement had been reached between the two judges. This resulted in a reduced pool of 126 rebus puzzles from the initial pool of 186. From this pool of 126 puzzles, 3 were selected to serve as practice items (see **Appendix C** in Supplementary Material) and 84 puzzles were randomly selected for norming, with 42 being allocated to Set A and 42 to Set B (see **Table 1** for details). The number of puzzles that were allocated to each puzzle category within Set A and Set B were balanced where possible. A number of puzzle categories were more commonly represented than others, with items falling into the imagery category being most prevalent, although it should be noted that this category also involves more varied items than the other categories. It can also be seen from **Table 1** that Categories 1 and 2 had the lowest representation in Sets A and B, although this relative under-representation may serve to allay concerns that at an abstract level the solution principle underpinning puzzles in these two categories is very similar.

TABLE 1 | The number of rebus items per puzzle category for Set A and Set B.

Puzzle category	Rebus set A	Rebus set B	Total
1. A word, picture or number, over another word, picture or number	4	3	7
2. A word, picture or number, under another word, picture or number	1	2	3
3. A word presented within another word	7	7	14
4. A play on words with numbers	7	7	14
5. Imagery	15	15	30
6. Spatial	8	8	16
Total	42	42	84

A further three rebus puzzles were selected as practice problems (see **Appendix C** in Supplementary Material). These problems served as practice items for both Set A and Set B items. These practice puzzles were chosen from the pool of problems for which an agreement had not been reached on a category, and had answers as follows: “all over again,” “once upon a time” and “long johns” (**Appendix C** in Supplementary Material). Problems for which an agreement had not been reached were selected as practice items so as not to provide a specific strategic advantage to the solving of any category of rebus puzzle in the norming study. Each participant received the same three practice problems in a fixed order, regardless of the rebus set that they had been allocated.

Phrase Familiarity Task

In order to solve rebus puzzles, a particular phrase or saying must be identified from the pictorial, number and word representation provided. A rebus phrase familiarity task was developed to test participants’ familiarity with the phrases (or answers) of each rebus puzzle. Following completion of the full set of rebus items, participants were presented with the phrases from the 126 Rebus where a category agreement had been reached, and a further 26 “pseudo phrases” developed by the experimenters (see **Appendix D** in Supplementary Material). The pseudo phrases were based on existing and well-known common UK English phrases. For example, “knock on metal” is a variant of the common phrase “touch wood.” These phrases therefore had an element of plausibility, whilst not being a common phrase or saying in UK English. The aim of the pseudo phrases was to ensure participants’ task engagement during the phrase familiarity rating task and thereby counteract any tendency toward purely confirmatory responding. Each phrase was presented to participants, and they were asked to respond with “yes” if the phrase was familiar to them, and “no” if the phrase was not familiar. Participants were informed that familiarity might stem from the experiment or from encountering these phrases in everyday life.

Procedure

Each participant completed the experiment individually and remotely via a desktop PC, laptop computer, or tablet.

Participants read an information sheet and indicated consent to participate in the experiment before proceeding. Each participant completed only one set of rebus puzzles (Set A or Set B). The experiment was constructed using Qualtrics experimental survey software and deployed through Prolific Academic, a survey recruitment platform. Each participant completed the set of rebus puzzles initially, followed by the phrase familiarity task.

The task instructions were presented on the screen for participants to read through prior to commencing the computerized rebus task. Participants were informed that they would be presented with a combination of words, pictures, or numbers on the computer screen and that their task was to identify the common word, phrase or saying represented by these words, pictures or numbers. Participants were also informed that they would complete 42 rebus puzzles in the study in addition to tackling three practice items to begin with. On completion of the three practice puzzles the answers were provided. This practice phase helped to ensure that participants were familiar with the general nature of rebus puzzles as well as with the response requirements of the study. The three practice items were identical for Set A and Set B. For each set of rebus puzzles, the presentation of the items was randomized by the Qualtrics programme.

All rebus puzzles were presented one at a time in black and white and were based within a square at the center of the computer screen covering an area of approximately 10 cm by 10 cm. The instructions required participants to read the rebus puzzle carefully, consider their answer, and when they had generated their final answer to input it in the text box provided. A maximum of 30 s was provided to view each rebus puzzle and generate and input an answer to it. The participant was able to see the timer display with the 30 s time limit. The clock was stopped when the participant moved onto the following page. This was to ensure that further thinking time was not taken when inputting an answer to the problem. If an answer was not provided within this 30 s time limit, the programme automatically advanced onto the next page.

Following each rebus puzzle a screen appeared asking participants to rate their confidence in the accuracy of their answer on a sliding scale ranging from 1 to 100, where 1 was labeled as “not at all confident” and 100 was labeled as “very confident.” Participants moved the cursor to the appropriate point on the scale to reflect their confidence in their answer for that problem. A “not-applicable” box was also provided for each rebus puzzle and participants were asked to select this box to register a response to the confidence question in all cases where they had not given an answer to the preceding puzzle.

Following the confidence judgment question, participants were next asked to provide a rating to indicate their perceived solution strategy, that is, whether they felt they had solved the preceding rebus puzzle more by analysis or more by insight (i.e., “Did you feel as if the problem was solved more by insight or more by analysis?”). It was emphasized that insight and analysis are two ends of a continuum, and therefore participants were asked to indicate if their answer was more “analytic-like,” or “insight-like” by responding on a sliding scale. An “insight” response was described as the following: “*Insight means that the*

answer suddenly (i.e., unexpectedly) came to your mind while you were trying to solve the problem, even though you are unable to articulate how you achieved the solution. This kind of solution is often associated with surprise exclamations such as ‘A-ha!’.” An analysis response was described as the following: “*Analysis means that you figured out the answer after you deliberately and consciously tested out different ideas before you found the right phrase or saying. In this case for instance, you are able to report the steps that you used to reach the solution.*” The ends of the response scale in relation to the analysis vs. insight question were alternated and counterbalanced across participants. A “not-applicable” box was also provided for participants to select in those cases where they had not given an answer to the preceding rebus puzzle. Participants were forced to respond by either moving the cursor from the mid-way point (50) on the sliding scale, or by selecting the “not applicable” box before proceeding to the next page.

On completion of the 42 rebus puzzles, participants completed a phrase familiarity task. This involved them rating a list of 152 phrases that were presented in a fixed, sequential order. In this task the participants were presented with the phrases from the 126 rebus puzzles for which a category judgment agreement had been reached by the raters, along with a further 26 “pseudo” rebus phrases (see **Appendix D** in Supplementary Material). Pseudo phrases were utilized to ensure that a number of phrases were likely to elicit a “no” response to the familiarity question. For each phrase, word, or saying, participants were asked to respond “yes” to indicate that the phrase was familiar to them, and a “no” to indicate that the phrase was not familiar. At the end of the experiment participants were debriefed and thanked for their participation time.

RESULTS

Fundamental Performance Characteristics of Each Rebus Puzzle

Performance data were collated for the 84 rebus puzzles across the two sets of items. Each participant completed only one set of 42 rebus puzzles, with 85 participants completing the 42 Set A items and another 85 participants completing the 42 Set B items. For each rebus puzzle we calculated the number of correct solutions and the number of incorrect solutions that had been provided by the 85 participants. This allowed us to calculate the percentage of correct solutions for a particular rebus item, which we subsequently refer to as the *solution rate*. Note that a response was counted as being an “incorrect solution” if an answer to the rebus puzzle had been provided by a participant that was not the correct phrase or saying. For example, in response to the rebus puzzle “try to understand,” incorrect solutions included “try to stand up” and “try to stand divided.” A “don’t know” response, or no attempt at an answer, was *not* counted as an “incorrect solution,” but was instead designated as being a null response.

In addition, for each correctly solved rebus puzzle we calculated the mean and standard deviation for its solution time (out of a maximum of 30 s). The solution time was the time spent on the rebus puzzle page, including the time to input

the answer. This was to ensure that any additional time spent contemplating the answer during the process of typing, was accounted for in the timing analysis. When 30 s had elapsed, the programme progressed to the next rebus puzzle. Furthermore, for each rebus puzzle we calculated a mean confidence rating for correct solution responses, where ratings could range from 1 (not at all confident) to 100 (very confident). For the insight vs. analysis rating, we again determined for each correctly solved rebus item the extent to which it was deemed to have been solved more by insight or more by analysis. The measurement scale ranged from 1 (analysis) to 100 (insight).

The various performance measures calculated for each rebus puzzle are presented in **Table 2**, with rebus items organized in the table in descending order of solution rate (i.e., from the easiest to the most difficult). As shown in **Table 2**, it is evident that the 84 rebus puzzles vary greatly in terms of their difficulty, with solution rates ranging from 95.29 to 0%, and with mean solution times for correct responses ranging from 8.68 to 22.64 s. We contend that the variability in both solution rates and solution times for this set of rebus puzzles is of great benefit for the selection of rebus stimuli for use in future experimental research. We note, in particular, that there are 50 rebus puzzles with a solution rate between 20 and 80%, which provides a good number of items for future use even when those puzzles are discounted that might be viewed as demonstrating either floor or ceiling effects. We also note that the performance data in **Table 2** provide good evidence that puzzles belonging to the same category can differ markedly in their difficulty, as indicated by wide variability in solution rates. For example, two rebus puzzles from Category 1 (i.e., Item 43—“long overdue”; Item 1—“feeling on top of the world”) have mean solution rates of 95.29–45.88%, respectively. This observation again supports the value of these presented norms for the effective selection and control of rebus stimuli in future studies.

Table 2 also shows that the mean confidence ratings for correctly solved rebus puzzles are all above the scale mid-point of 50, with the exception of just one item (i.e., Item 68—“partly cloudy”—with a confidence score of 17). These data indicate that when participants solve a puzzle they generally have above average confidence in the correctness of the solution, although such confidence stretches across the full range above the scale midpoint from 52.80 right up to 100. When it comes to item selection for future studies using rebus puzzles then the mean confidence data could be very useful for controlling for problem characteristics (e.g., enabling mean confidence scores for puzzles to be equated across different difficulty levels).

In relation to the performance measures for rebus puzzles that are concerned with self-perceived solution strategies (i.e., analysis vs. insight), **Table 2** indicates a good degree of variability in scores across the rebus puzzles, with scores ranging from 1 at the analysis end of the scale to 76 at the insight end. Interestingly, however, scores on this measure generally cluster between 35 and 65 (i.e., 15 points either side of the scale midpoint), with only a few puzzles having scores that extend beyond these lower and upper bounds. This finding suggests that *either* insight *or* analysis solution strategies may be deployed when solving a majority of these rebus items, with averaging of scores inevitably

leading to the bunching of scores around the scale midpoint. We view this observation positively, as it suggests that rebus puzzles provide an excellent way to explore underpinning problem-solving processes associated with insight-based solutions vs. analysis-based solutions.

Solution Strategies and Solution Correctness

Following on from the aforementioned point, we note that recent research has revealed that solutions to problems that are generated by a process of self-reported insight are more likely to be correct than solutions generated by a process of analysis. For example, Salvi et al. (2016) demonstrated this finding across CRA problems, anagrams, rebus puzzles and fragmented line drawings, with other researchers reporting the same effect with magic tricks (see Danek et al., 2014b; Hedne et al., 2016). In explaining this so-called “accuracy effect” in relation to insight solutions, Salvi et al. (2016; see also Danek and Salvi, 2018) propose that the effect is most likely to be attributable to the “all-or-nothing” manner in which insight solutions emerge into consciousness once non-conscious processing has been completed. In contrast, solutions that are based on analysis can be “guesses” that derive from conscious processing that is prematurely terminated, especially under time constraints. Such guesses would give rise to more errors of commission (i.e., incorrect responses) than errors of omission (i.e., timeouts) when compared to insight responses (for related evidence see Kounios et al., 2008).

In order to provide further corroboratory evidence for the existence of this consistent accuracy effect in relation to insight solutions, we applied a standard accuracy analysis to the present dataset to determine whether rebus puzzles that are solved via insight are more likely to be correct than rebus puzzles solved via analysis. Of all the solution responses designated as being based on insight (i.e., falling between 51 and 100 on the analysis/insight scale), an average of 65% ($SD = 27$) were correct. In contrast, of all the solution responses designated as being based on analysis (i.e., falling between 1 and 49 on the analysis/insight scale), an average of 54% ($SD = 27$) were correct. A paired-samples *t*-test revealed that insight solutions were indeed significantly more likely to be correct than analytic solutions, $t = 4.76, p < 0.001$.

Following Salvi et al. (2016), we also conducted a secondary analysis of the dataset with a narrower response window than the full 30 s that was available for solving each rebus puzzle. The analysis was similar to that just described, except that only those responses with latencies within a 2–10 s time-window were included. This approach helps to ensure a similar balance of insight and analytic responses in the dataset whilst also eliminating very fast responses made during the first 2 s, given that participants might inadvertently label these as insight-based (see Salvi et al., 2016). This revised analysis again revealed the predicted accuracy effect, with insight responses being significantly more likely to be correct ($M = 79\%$, $SD = 27$) than analytic responses ($M = 65\%$, $SD = 36$), $t = 4.69, p < 0.001$.

The previous approach to analyzing the link between solution strategies and solution correctness revolved around a

TABLE 2 | Normative data for 84 rebus puzzles (please refer to the text for further details).

Item	Answer	Puzzle set and puzzle category	Number of correct solutions	Number of incorrect solutions	Familiarity count for the solution phrase*	Solution rate (percentage)	Mean solution time (s)	SD solution time (s)	Mean confidence rating	Mean insight rating
43	Long overdue	B/1	81	4	84 [79]	95.29	8.02	3.26	90.45	64.00
44	Man overboard	B/1	80	5	84 [83]	94.11	8.94	4.65	96.29	58.51
62	Small talk	B/5	79	6	83 [78]	92.94	6.76	2.45	89.56	60.87
81	First aid	B/6	79	5	n/a	92.94	7.45	3.05	91.62	63.35
17	Four-wheel drive	A/4	77	5	78 [80]	90.59	9.55	3.79	80.84	57.87
58	Small print	B/4	77	8	81 [79]	90.59	7.50	2.97	88.60	65.73
69	Jack in the box	B/5	77	7	81 [81]	90.59	9.03	4.55	89.12	60.39
76	Wave goodbye	B/5	76	8	81 [76]	89.41	8.68	3.30	85.26	58.99
59	Go for it	B/4	75	9	81 [75]	88.21	8.33	3.42	85.24	55.61
25	Split personality	A/5	74	10	81 [80]	87.06	7.28	3.24	82.62	68.80
21	Big bad wolf	A/5	73	8	80 [81]	85.88	8.45	4.14	77.84	69.19
26	Top hat	A/5	73	11	79 [72]	85.88	7.35	3.37	74.42	65.82
31	Little sister	A/5	73	12	80 [76]	85.88	8.18	2.74	74.32	62.40
65	A big if	B/5	73	9	79 [62]	85.88	8.99	4.95	64.62	55.52
80	Upside down	B/6	73	8	81 [75]	85.88	9.81	4.67	81.22	57.32
30	Broken promise	A/5	72	12	81 [82]	84.71	9.38	4.86	78.16	63.31
4	Stay overnight	A/1	71	12	75 [78]	83.52	10.48	5.72	61.34	44.68
47	London underground	B/2	70	11	81 [81]	82.35	12.26	4.97	89.35	49.86
20	Half-hearted	A/5	69	16	82 [79]	81.18	8.70	3.99	82.71	58.45
15	Too little too late	A/4	67	7	84 [84]	78.82	12.66	5.67	78.20	61.96
16	Forty winks	A/4	67	14	76 [75]	78.82	10.29	5.24	78.70	76.00
57	All for one and one for all	B/4	67	11	81 [81]	78.82	16.15	5.70	82.13	50.33
52	You're in trouble	B/3	66	14	80 [79]	77.65	12.78	6.33	81.89	49.25
2	Bend over backwards	A/1	64	13	80 [83]	75.29	11.26	5.98	76.29	52.77
35	Cross breed	A/6	64	16	77 [67]	75.29	10.69	5.02	77.58	52.60
70	Big brother	B/5	64	14	81 [81]	75.29	8.68	4.02	75.02	55.81
46	Swim underwater	B/2	63	13	78 [40]	74.12	12.42	4.47	82.59	49.25
51	Tongue in cheek	B/3	63	11	83 [80]	74.12	11.62	4.85	89.49	50.14
53	Within reason	B/3	57	10	81 [79]	67.06	12.07	6.01	82.21	56.26
39	Side show	A/6	56	26	73 [70]	65.88	9.30	5.51	67.00	55.94
60	Tricycle	B/4	52	26	77 [68]	61.18	10.62	5.09	75.38	50.10

(Continued)

TABLE 2 | Continued

Item	Answer	Puzzle set and puzzle category	Number of correct solutions	Number of incorrect solutions	Familiarity count for the solution phrase*	Solution rate (percentage)	Mean solution time (s)	SD solution time (s)	Mean confidence rating	Mean insight rating
29	Shrinking violet	A/5	50	16	65 [60]	58.82	11.55	5.45	79.98	58.80
7	Burying your head in the sand	A/3	47	18	78 [81]	55.29	15.39	8.07	73.59	42.57
49	Deep in thought	B/3	46	20	82 [77]	54.12	10.66	4.75	89.39	53.79
54	One in a million	B/3	45	16	84 [83]	52.94	12.54	5.18	91.60	58.16
38	Pretty please	A/6	44	15	82 [80]	51.76	9.08	3.23	76.72	64.89
41	Falling in love	A/6	42	23	77 [81]	49.41	12.76	5.91	69.91	47.79
55	Tripod	B/3	42	32	73 [61]	49.41	10.71	5.58	74.25	56.59
11	Walk in the park	A/3	41	20	77 [83]	48.34	12.51	5.31	81.22	44.00
37	Uphill struggle	A/6	41	33	79 [79]	48.23	9.74	3.36	70.29	57.88
23	Shop "till you drop"	A/5	40	37	81 [83]	47.06	14.91	6.17	77.91	53.40
56	More often than not	B/4	40	33	81 [76]	47.06	12.11	5.61	86.93	58.48
1	Feeling on top of the world	A/1	39	22	84 [83]	45.88	14.06	7.43	77.61	56.03
3	Dinner on the table	A/1	39	29	75 [78]	45.88	15.06	6.73	61.46	41.26
19	Forgive and forget	A/4	39	35	81 [83]	45.88	14.47	5.74	78.76	41.35
72	Tongue in cheek	B/5	38	36	82 [80]	44.71	13.54	4.98	87.21	58.16
77	Warm up	B/6	36	30	83 [78]	42.35	11.82	4.83	83.08	51.49
40	Quite right	A/6	35	35	81 [77]	41.18	11.12	6.11	80.54	57.54
8	Pie in the sky	A/3	33	39	76 [75]	39.29	10.87	4.41	78.91	45.24
33	Downwind	A/5	32	44	71 [73]	37.64	10.80	6.37	68.79	47.81
50	Singing in the rain	B/3	31	13	82 [78]	36.47	16.82	6.53	92.13	50.35
12	Bad intention	A/3	30	30	78 [71]	35.29	17.41	7.54	73.30	39.90
45	Running on empty	B/1	30	23	80 [77]	35.29	13.74	4.98	74.80	60.21
71	Growing old	B/5	29	33	78 [77]	34.12	12.62	6.52	82.93	49.07
27	Extended family	A/5	28	31	78 [82]	32.94	12.21	5.45	74.14	47.21
75	Tall order	B/5	27	41	80 [77]	31.76	10.01	5.08	75.04	57.11
18	No one to blame	A/4	25	47	78 [77]	29.41	11.92	5.79	55.80	48.04
22	Balanced diet	A/5	25	35	83 [81]	29.41	11.26	5.80	84.00	36.85
79	Waterfall	B/6	24	54	77 [76]	28.24	10.62	4.20	62.58	57.54
32	Not bad	A/5	23	36	77 [81]	27.06	10.54	6.37	53.48	52.50
5	Try to understand	A/2	22	35	80 [77]	25.88	19.93	6.58	73.22	36.54
10	Reading between the lines	A/3	21	16	85 [84]	24.71	22.07	5.22	61.13	53.22
82	What goes up must come down	B/6	20	21	82 [83]	23.53	17.52	6.46	88.30	49.60
84	Round of drinks	B/6	20	46	83 [78]	23.53	14.75	5.03	83.25	42.80
6	Foot in the door	A/3	19	28	80 [79]	22.35	18.58	6.44	76.16	26.47

(Continued)

TABLE 2 | Continued

Item	Answer	Puzzle set and puzzle category	Number of correct solutions	Number of incorrect solutions	Familiarity count for the solution phrase*	Solution rate (percentage)	Mean solution time (s)	SD solution time (s)	Mean confidence rating	Mean insight rating
13	Six of one and half a dozen	A/4	19	32	71 [66]	22.35	19.25	4.82	71.90	44.90
73	Long time no see	B/5	19	29	83 [83]	22.35	17.19	6.21	91.63	34.42
83	Search high and low	B/6	17	30	83 [80]	20.00	14.34	6.78	86.00	43.76
14	Up for grabs	A/4	16	18	80 [80]	18.82	19.49	7.65	70.17	43.44
42	Your time is up	A/6	16	44	80 [83]	18.82	14.14	7.80	68.89	63.89
64	Through thick and thin	B/5	16	48	84 [81]	18.82	13.73	7.34	91.63	52.00
48	Put it in writing	B/3	12	37	80 [71]	14.12	17.91	6.70	80.33	32.08
61	Free for all	B/4	11	66	83 [82]	12.94	10.61	2.95	80.09	41.55
78	Up to no good	B/6	10	28	84 [82]	11.76	21.89	6.91	73.60	31.20
28	Vanishing point	A/5	8	65	68 [68]	9.41	15.04	6.02	82.38	60.38
74	Unfinished symphony	B/5	8	28	57 [57]	9.41	21.30	6.69	74.88	33.63
36	Parallel bars	A/6	5	49	54 [61]	5.88	14.44	9.35	52.80	54.00
67	Break time	B/5	3	70	n/a	3.53	8.99	4.51	60.00	20.33
68	Partly cloudy	B/5	3	56	72 [66]	3.53	22.64	6.82	17.00	36.33
34	Unfinished business	A/5	1	50	80 [82]	1.18	15.02		99.00	71.00
63	Split second timing	B/5	1	50	80 [65]	1.18	13.17		100	1.00
66	Emergency stop	B/5	1	50	81 [78]	1.18	12.83		100	1.00
9	Partridge in a pear tree	A/3	0	25	77 [77]					
24	A large overdraft	A/5	0	51	53 [71]					

Blank cells are provided where no correct answers were given for a particular rebus puzzle. For items 67 and 81 an "n/a" response is noted for the familiarity count for these solution phrases because of a computer error in registering responses. *Phrase familiarity count for the phrase in which the corresponding rebus puzzle had not been encountered.

dichotomous measure of solution strategies as being insight-based (above 51 on the analysis/insight scale) vs. analysis-based (below 49 on the analysis/insight scale). Conditionalizing solution correctness on solution strategy has become the standard approach in the literature for examining the existence of the accuracy effect. However, on the assumption that there is a very tight coupling between insight solutions and solution correctness it is also useful to test for the existence of a “correctness effect,” whereby correct solutions are more likely to be solved by insight than are incorrect solutions. This correctness effect should arise because of the “all-or-nothing” manner in which correct solutions typically arise via insight in comparison to the way in which analysis can promote incorrect guesses.

Determining the existence of a correctness effect involves conditionalizing self-reported solution strategies on the correctness of the proffered solution. To conduct the requisite analysis, we made use of participants’ exact ratings on the 1–100 analysis/insight scale, adding a greater degree of precision to the measure of analysis vs. insight than that which would arise from simply dichotomizing the scale at its midpoint. Our resulting analysis simply applied a paired samples *t*-test to compare participants’ mean solution strategy scores for all correct solutions vs. their mean solution strategy scores for all incorrect solutions. This test revealed that correct responses resulted in a significantly higher analysis/insight score ($M = 55.74$, $SD = 22.40$) than incorrect responses ($M = 47.81$, $SD = 18.62$), $t = 4.64$, $p < 0.001$. The observation that the mean analysis/insight score for correct response fell above the scale midpoint indicates a more insight-based solution strategy for correct solutions. In contrast, the observation that mean analysis/insight score for incorrect response fell below the scale midpoint indicates a more analysis-based solution strategy for incorrect solutions.

In sum, when considered together, the full set of analyses of the relation between solution strategies and solution correctness indicates a tight, bidirectional relationship in the form of both an accuracy effect (insight solutions are more likely to be correct than analytic solutions) and a correctness effect (correct solutions are more likely to be insight-based than incorrect solutions).

Solution Strategies, Solution Correctness, and Solution Confidence

In considering potential explanations of the accuracy effect for insight solutions, Danek and Salvi (2018) contemplate the viability of an account based on the notion that solvers might use their confidence in accurate responses as a metacognitive cue for reporting the solution as being based on insight. The essential idea here is that when accurate, solvers might feel highly confident about their solution and therefore retrospectively report having had an insight experience. As Danek and Salvi (2018) acknowledge, at first glance this account of the accuracy effect seems to gain support from the observation that confidence correlates highly with insight ratings (Webb et al., 2016; Danek and Wiley, 2017). However, Danek and Salvi (2018) counter that the studies that reveal a correlation between confidence and insight specifically mention “confidence” in their instructions

to participants, possibly inflating the observed correlation. Moreover, they note that solvers sometimes also feel confident about *incorrect* solutions (Danek and Wiley, 2017), suggesting that it is unlikely that the accuracy effect is solely based on high confidence serving as a metacognitive cue for insight ratings.

We agree with Danek and Salvi’s cautionary arguments and consider that a causal link between confidence judgments and insight ratings seems unlikely. Given that the present study elicited confidence ratings from participants for all generated solution responses, we analyzed the present dataset with a view to shedding further light on how solution confidence is related to solution strategy and solution correctness. A 2×2 Analysis of Variance (ANOVA) was conducted to determine the difference in confidence ratings according to solution correctness (correct vs. incorrect) and solution strategy (insight vs. analysis—again based on dichotomized scores).

The ANOVA revealed that there was no main effect of solution strategy, with confidence ratings for solutions generated via insight ($M = 59.56$, $SE = 1.62$) not differing significantly from confidence ratings for solutions generated via analysis ($M = 57.46$, $SE = 1.41$), $F_{(1,136)} = 1.47$, $MSE = 409.40$, $p = 0.23$. There was, however, a significant main effect of solution correctness, with confidence ratings being significantly higher for correct solutions ($M = 74.35$, $SE = 1.56$) in comparison to incorrect solutions ($M = 42.68$, $SE = 1.58$), $F_{(1,136)} = 273.28$, $MSE = 502.92$, $\eta_p^2 = 0.67$, $p < 0.001$. There was no solution strategy by solution correctness interaction, $F < 1$, $p = 0.42$. These results support the existence of heightened confidence for correct solutions over incorrect solutions whether or not the problem was solved via insight, suggesting that there is no unique and clear-cut link between perceived confidence and insight phenomenology, thereby supporting the arguments of Danek and Salvi (2018).

Solution Strategies, Solution Correctness, and Response Time

A 2×2 ANOVA was also conducted to determine the difference in mean solution times as a function of solution correctness (correct vs. incorrect responses) and solution strategy (insight vs. analysis). Given that solution-time data are often found to be positively skewed, thereby undermining the assumptions required for the pursuit of parametric data analysis, we first determined the skew in the dataset for each condition according to each set of rebus puzzles. We observed that two conditions demonstrated positive skew in their associated solution-time data, with skew values (i.e., 1.92 and 1.74) above typically accepted levels (e.g., Tabachnick and Fidell, 2013). As a result, a Log_{10} transformation was performed on the solution-time data for all conditions prior to running the ANOVA (see Table 3 for the natural and Log_{10} mean solution times for each condition).

For the transformed solution-time data the ANOVA revealed a significant main effect of solution strategy, with problems solved via insight being solved significantly faster ($M = 1.08$, $SE = 0.01$) than problems solved via analysis ($M = 1.17$, $SD = 0.01$), $F_{(1,136)} = 46.04$, $MSE = 0.02$, $\eta_p^2 = 0.25$, $p < 0.001$. This finding underscores how analysis is often a more laborious process

TABLE 3 | Mean natural solution times (s) and mean Log₁₀ solution times as a function of solution strategy (insight vs. analysis) and solution correctness (correct vs. incorrect).

	Puzzles solved by insight		Puzzles solved by analysis	
	Correct	Incorrect	Correct	Incorrect
Mean Natural Solution Times	10.93 (4.12)	15.76 (6.19)	13.32 (4.50)	18.20 (5.79)
Mean Log ₁₀ Solution Times	1.01 (0.15)	1.16 (0.18)	1.10 (0.15)	1.23 (0.17)

Standard deviations are shown in parenthesis.

than insight. There was a significant main effect of solution correctness, with mean solution times being significantly faster for correct responses ($M = 1.06$, $SE = 0.01$) in comparison to incorrect responses ($M = 1.20$, $SE = 0.01$), $F_{(1,136)} = 217.47$, $MSE = 0.01$, $\eta_p^2 = 0.61$, $p < 0.001$. This observation is unsurprising given that correct solutions are more likely to arise from a (fast) insight process than incorrect solutions. There was no solution strategy by solution correctness interaction, $F_{(1,136)} = 0.79$, $MSE = 0.01$, $p = 0.38$.

Phrase Familiarity

In **Table 2**, we also provide two familiarity counts for the solution phrase that was associated with each rebus puzzle, with each familiarity count having a maximum value 85, in line with the number of participants tackling each set of rebus puzzles. The importance of providing two familiarity counts for each particular solution phrase is to draw a distinction between a familiarity rating given to a solution phrase after the participant had encountered the corresponding rebus puzzle, compared to having not encountered the corresponding rebus puzzle. This distinction is made possible by the fact that each participant rated the familiarity for each of the 84 solution phrases, whilst only having attempted to solve 42 of the rebus puzzles relating to these phrases. The first familiarity count presented in **Table 2** is for the solution phrase from the set in which the corresponding rebus puzzle had been encountered. The second familiarity count (provided in square brackets in **Table 2**) is for the solution phrase from the set in which the corresponding rebus puzzle had not been encountered.

An independent samples *t*-test was conducted to determine if there was a significant difference between these two familiarity counts. This analysis revealed that phrase familiarity ($M = 78.61$, $SD = 5.98$) was significantly higher when the rebus puzzle corresponding to the solution phrase had been encountered in comparison to when the rebus puzzle corresponding to the solution phrase had not been encountered ($M = 76.41$, $SD = 7.43$), $t = 2.08$, $p = 0.039$. This suggests that there might be a small but reliable bias toward a judgment of familiarity being given for a solution phrase for which the previous rebus puzzle had been encountered, even though the “correct” solution phrase for each rebus has not been provided.

The familiarity data for solution phrases enabled us to explore a number of potentially interesting associations between phrase familiarity and the performance measures identified in

Table 2. These associations were explored using the item-based performance data (i.e., frequency counts and mean scores) for the 84 rebus puzzles that are depicted in **Table 2**. In order to explore patterns of association involving the familiarity data, we took the two familiarity count measures previously identified and transformed them into percentage familiarity scores. To reiterate, the first familiarity score was for the rating of a solution phrase from the set in which the corresponding rebus puzzle had been encountered. The second familiarity score was for the rating of the solution phrase from the set in which the corresponding rebus puzzle had not been encountered. Having computed the two percentage familiarity scores for each rebus puzzle we then correlated these independently with five performance measures for each rebus item, that is: its solution rate, its error rate (i.e., the percentage of incorrect solutions), the mean confidence in correct solutions, the mean analysis/insight score for correct solutions and the mean response time (seconds) for correct solutions.

Pearson correlation coefficients indicated that each familiarity score was not significantly associated with the solution rate ($r = 0.10$ and $r = 0.05$, respectively, both $ps > 0.05$). The absence of an association between phrase familiarity and solution success attests to the challenging nature of many of the rebus puzzles despite the fact that the underpinning solution phrase was well-known. For example, the two rebus puzzles with a 0% solution rate (Item 24—“large overdraft”; Item 9—“partridge in a pear tree”) still received scores of over 50% for the familiarity of their solution phrases. In other words, even when there is a good degree of familiarity with the underpinning solution phrase for a rebus puzzle, this does not necessarily translate into the ability to solve the rebus puzzle.

In terms of other observed associations, there was a weak but nevertheless significant negative correlation between the first familiarity rating (when the rebus puzzle corresponding to that particular solution phrase had been encountered) and error rate ($r = -0.24$, $p = 0.03$), indicating that as familiarity with the underpinning solution phrase increased, the percentage of incorrect solutions decreased. A similar pattern was found for the second familiarity measure ($r = -0.16$), but this failed to reach significance. Neither of the phrase familiarity scores was significantly associated with participants’ mean confidence in correct rebus solutions ($r = 0.20$ and $r = 0.14$, respectively), with their mean analysis/insight scores for correct solutions ($r = 0.12$ and $r = 0.16$) or with their mean response time for correct solutions ($r = 0.10$ and $r = 0.12$), all $ps > 0.05$.

Rebus Puzzle Categories

As discussed in the materials section, rebus puzzles were divided into six categories according to common solution principles (refer to **Table 1** for the distribution of categories across rebus puzzle Set A and Set B). The measurements presented in **Table 2** are reorganized in **Table 4** so as to show data collapsed across the six rebus categories. In other words, these reconfigured data provide an indication of how each of the dependent variables differs according to each particular rebus puzzle category.

The data in **Table 4** indicate that rebus puzzles in Categories 1, 2, and 4 gave rise to higher mean solution rates than

TABLE 4 | Normative data for each of the six rebus puzzle categories.

Puzzle category	Mean number of correct solutions	Mean number of incorrect solutions	Mean familiarity count for solution phrases	Mean solution rate (Percentage)	Mean solution time (s)	SD solution time (s)	Mean confidence rating	Mean insight rating
1. A word, picture or number, over another word, picture or number	57.71	15.42	80.29	67.89	11.65	5.54	76.89	53.92
2. A word, picture or number, under another word, picture or number	51.67	19.67	79.67	60.78	14.87	5.34	81.72	45.22
3. A word presented within another word	36.87	21.93	79.73	44.48	14.42	5.92	80.40	47.00
4. A play on words with numbers	48.62	23.92	79.38	57.19	12.53	4.96	77.90	53.49
5. Imagery	38.50	30.63	76.76	46.86	11.68	5.05	77.93	48.19
6. Spatial	36.38	28.94	78.40	42.79	12.47	5.51	75.84	53.22

those in Categories 3, 5, and 6, suggesting that the spatial and imagery related rebus puzzles are generally more challenging than those related to words, with the exception of the “word presented within another word” puzzles (Category 3), which are also more difficult than the other word-related rebus items. Nevertheless, the item-based data presented in **Table 2** reveal considerable variability in difficulty levels for items within each of the categories, ensuring that item selection in future studies can capitalize on such variability in situations where a puzzle-difficulty manipulation is a desirable feature of an experimental design.

With respect to mean analysis/insight ratings, the descriptive data in **Table 4** indicate very limited variability in ratings across the different rebus categories, with mean analysis/insight scores showing a narrow range from 45.22 to 53.22. A similar picture of homogeneity emerges for: (1) mean confidence ratings, which again show considerable similarity across categories, ranging from 75.84 to 81.72; and (2) mean solution times, which range from 11.65 to 14.87 s. Such high levels of similarity in people's performance measures across rebus categories support the usefulness of the present norming data to inform item selection for future studies.

We finally note that the unequal number of rebus puzzles in each of the rebus categories (including the particularly low number of puzzles in Categories 1 and 2) precludes the pursuit of formal, inferential analysis of the possible performance differences that might arise across rebus categories.

GENERAL DISCUSSION

Classic insight problem-solving tasks such as the candle problem (Duncker, 1945), two-string problem (Maier, 1930) and nine-dot problem (Maier, 1930) are complex and time-consuming to solve whilst also yielding potentially ambiguous solutions and being susceptible to the effects of confounding variables (cf. Ball et al., 2015). Furthermore, given the popularity of these tasks in the problem-solving literature and their exposure in research, the solutions to classic insight problems are often generally well-known. This has led to the advent of additional pools of insight-based problems, such as CRAs (e.g., Bowden and Jung-Beeman, 2003b; Wu and Chen, 2017), magic tricks (e.g., Danek et al.,

2014a,b) and rebus puzzles (e.g., MacGregor and Cunningham, 2008, 2009; Salvi et al., 2015).

The use of both CRAs and rebus puzzles is especially appealing, since in contrast to classic insight problems they are relatively simpler and yield unambiguous single-word answers (CRAs) or single phrases (rebus puzzles). They are easy to administer to participants and straightforward to record answers for and they are additionally relatively fast for solvers to generate solutions to. Moreover, multiple problems can be presented within a single session to maximize the number of observations per experimental condition, and therefore the reliability of the data obtained. The problems are also well-suited to study using fMRI (e.g., Kizilirmak et al., 2016) and EEG (e.g., Li et al., 2016) due to their simplicity and possibility for presentation within a compressed visual space. However, the utility of these insight problems in research is heavily dependent upon the knowledge of baseline problem difficulties and solution times (i.e., normative data).

In addition to the many positive features of CRAs and rebus puzzles that we have identified, we also note that they appear to share with classic insight problems the same kinds of underpinning component processes and phenomenological experiences (Bowden and Jung-Beeman, 2003a, 2007). For example, both CRAs and rebus puzzles have the potential to engender initial misdirection along ineffective solution avenues or the failure of effective retrieval processes that can culminate in impasse and a subsequent “Aha!” experience when a route toward a solution suddenly comes to mind (Salvi et al., 2015). Therefore, both CRAs and rebus puzzles can be used to address the degree to which participants differ in their tendency toward solving particular items via insight or analytic strategies.

The extant literature provides extensive normative data for CRAs, which have been normed for participant samples in the USA (Bowden and Jung-Beeman, 2003b), the UK (Sandkühler and Bhattacharya, 2008), China (Wu and Chen, 2017), and Italy (Salvi et al., 2016). To the best of our knowledge, however, there are very limited normative data for rebus puzzles, with the only data that are currently available being restricted to a set of 88 Italian rebus puzzles (Salvi et al., 2015). Due to the linguistically contextualized nature of rebus puzzles, however, it is important to extend the base of normative data for such problems to other

languages, including UK English. In setting out to address this gap in the literature we endeavored to undertake a norming study with a set of carefully-selected rebus puzzles for which we could obtain data relating to solution rates, error rates, solution times, solution confidence, self-reported solution strategies (insight vs. analysis), and familiarity with the solution phrases.

In **Table 2**, we provide normative data for each of the 84 rebus puzzles that we examined, which were assessed as two separate sets of 42 puzzles. Within **Table 2**, the data are depicted in descending order of their mean solution rate within the 30 s time limit available. Also reported in **Table 2** are the number of incorrect solutions, classified as attempts at a response that gave rise to incorrect words or phrases. Mean solution times (and standard deviations) are also displayed. Since rebus puzzles may differentially engender insight vs. analytic solution strategies, we additionally report data for participants' self-reported solution strategies. In **Table 4**, we provide normative data for rebus puzzles as a function of the rebus category within which they fell in terms of the underpinning solution principle. In **Appendices A, B** in Supplementary Material, all rebus puzzles are presented pictorially according to their presented set.

Since solutions to rebus puzzles are contingent on knowledge of the particular solution phrase underpinning the problem, we thought it critical to report data on the familiarity of each phrase that comprised a rebus solution. We observed that participants were largely familiar with the rebus solution phrases presented to them. Therefore, we can be confident that the rebus puzzles that were normed in the present study relate to well-known UK English phrases or sayings. We distinguished between two types of familiarity with the solution phrases, and found that the familiarity for a solution phrase in which the corresponding rebus puzzle had been attempted was significantly higher than the familiarity for a solution phrase in the absence of previously encountering the associated rebus puzzle. It is interesting to note that this bias existed even though the “correct” solution phrase for each rebus was not directly provided to the participants. The mere exposure to the associated rebus puzzle appeared to increase a subsequent familiarity rating for the solution phrase. Neither familiarity rating was associated with solution rate, mean confidence, mean insight or mean response time. Familiarity ratings were, however, associated with the percentage of incorrect solutions, in that greater familiarity led to fewer incorrect solutions, although this association was restricted to the familiarity rating for solution phrases for which the corresponding rebus puzzles had been attempted. The absence of significant associations between phrase familiarity and solution rate, mean confidence, mean insight and mean response time are unsurprising, given that we observed generally high familiarity levels for most of the rebus puzzle solution phrases.

More detailed analyses of the present dataset were also undertaken, which provide further support for a growing body of evidence demonstrating that solutions that arise from a self-reported insight process are more likely to be correct than solutions that arise via a process of analysis (e.g., Metcalfe, 1986; Salvi et al., 2015; Danek and Salvi, 2018). This particular advantage for insight responses appears to hold not just for rebus

puzzles, but also for CRA problems, magic tricks and anagrams (Danek and Salvi, 2018). Not only are insight solutions more likely to be correct than analytic solutions, they also arise more rapidly. However, these particular “insight” advantages were not seen to extend to people's self-rated confidence in solutions that were generated via insight (see also Hedne et al., 2016; Salvi et al., 2016).

We suggest that the rich seam of norming data reported here for rebus puzzles can be tapped to create different sets of stimuli that are closely matched on critical variables such as problem difficulty. This matching can be done either by hand, or preferably, via the use of stimulus matching software programs such as “Match” (Van Casteren and Davis, 2007) that automate the selection of groups of stimuli sets from larger pools through matching on multiple dimensions. In relation to the issue of controlling stimulus selection, it is also necessary to consider the structure of rebus puzzles and the resulting strategy that might be adopted to solve a particular problem. As noted in our method section (see also Salvi et al., 2015), given the structural similarity of some rebus puzzles, care must be taken to separate these problems to control for, or minimize, order and carry-over effects from one problem to subsequent ones. This is important when presenting a set of problems either within or between experimental blocks. That is, the solution for one problem with a particular structure (e.g., spatial), may influence the finding of a solution for a later encountered problem with a similar structure (e.g., via transfer or priming effects).

This latter issue is apparent if we consider Item 49 (“THODEEPUGHT”; solution: “deep in thought”) and Item 51 (“CHTONGUEEK”; solution: “tongue in cheek”). Here we see an example of two different problems from Category 3 (i.e., a word within a word), where the rebus is structured in such a way that the first word is quite literally presented “within” another word. Our categorization of the problems into different structural types that were validated through interrater reliability checks, can be used to help researchers to identify such overlap in rebus puzzles and thus avoid an issue of presenting problems underpinned by a similar structure or solution strategy. It remains unknown to what extent the transfer of problem structures assists solution rates or solution times for rebus puzzles from common categories. The present dataset does not permit an analysis of order effects according to each rebus puzzle within each category. However, descriptive statistics provided for each rebus puzzle do demonstrate a broad range of solution rates and solution times—even for problems within the same puzzle category—which is suggestive of minimal practice effects. Drawing on an example of two rebus puzzles from Category 3, solution rates for these two puzzles varied from 74.12 to 14.12%.

In conclusion, we hope that the materials and normative data presented here will arm researchers with important apparatus through which problem solving and creativity can be studied with UK English speaking participants. Like CRAs and their conceptual antecedents, RATs, rebus puzzles can be used across a broad range of domains to study problem solving and creative thinking, affect, psychopathologies and metacognitive processes.

AUTHOR CONTRIBUTIONS

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

FUNDING

The research reported in this article was supported by funding from the British Academy and Leverhulme Trust

REFERENCES

- 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. Reason.* 21, 76–96. doi: 10.1080/13546783.2014.934399
- Ball, L. J., and Stevens, A. (2009). “Evidence for a verbally-based analytic component to insight problem solving,” in *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds N. Taatgen and H. van Rijn (Austin, TX: Cognitive Science Society), 1060–1065.
- Barr, N. (2018). “Intuition, reason, and creativity: an integrative dual-process perspective,” in *The New Reflectionism in Cognitive Psychology*, ed G. Pennycook (Routledge), 99–124.
- Barr, N., Pennycook, G., Stolz, J. A., and Fugelsang, J. A. (2014). Reasoned connections: a dual-process perspective. *Think. Reason.* 21, 61–75. doi: 10.1080/13546783.2014.895915
- Bowden, E. M., 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. M., and Jung-Beeman, M. (2003a). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychon. Bull. Rev.* 10, 730–737. doi: 10.3758/BF03196539
- Bowden, E. M., and Jung-Beeman, M. (2003b). Normative data for 144 compound remote associate problems. *Behav. Res. Methods Instrum. Comput.* 35, 634–639. doi: 10.3758/BF03195543
- Bowden, E. M., 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
- Cunningham, J. B., MacGregor, J. N., Gibb, J., and Haar, J. (2009). Categories of insight and their correlates: an exploration of relationships among classic-type insight problems, rebus puzzles, remote associates and esoteric analogies. *J. Creat. Behav.* 43, 262–280. doi: 10.1002/j.2162-6057.2009.tb01318.x
- Danek, A. H., Fraps, T., von Müller, A., Grothe, B., and Öllinger, M. (2014a). 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., Fraps, T., von Müller, A., Grothe, B., and Öllinger, M. (2014b). Working wonders? Investigating insight with magic tricks. *Cognition* 130, 174–185. doi: 10.1016/j.cognition.2013.11.003
- Danek, A. H., and Salvi, C. (2018). Moment of truth: Why Aha! experiences are correct. *J. Creat. Behav.* doi: 10.1002/jocb.380. [Epub ahead of print].
- 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
- Duncker, K. (1945). On problem solving. *Psychol. Monogr.* 58, i–113. doi: 10.1037/h0093599
- Fleck, J. I., and Weisberg, R. W. (2004). The use of verbal protocols as data: an analysis of insight in the candle problem. *Mem. Cogn.* 32, 990–1006. doi: 10.3758/BF03196876
- Friedlander, K. J., and Fine, P. A. (2018). The penny drops: Investigating insight through the medium of cryptic crosswords. *Front. Psychol.* 9:904. doi: 10.3389/fpsyg.2018.00904
- Gilhooly, K. J., Ball, L. J., and Macchi, L. (2015). Insight and creative thinking processes: routine and special. *Think. Reason.* 21, 1–4. doi: 10.1080/13546783.2014.966758
- Hedne, M. R., Norman, E., and Metcalfe, J. (2016). Intuitive feelings of warmth and confidence in insight and non-insight problem solving of magic tricks. *Front. Psychol.* 7:1314. doi: 10.3389/fpsyg.2016.01314
- Howe, M. L., and Garner, S. R. (2018). Can false memories prime alternative solutions to ambiguous problems? *Memory* 26, 96–105. doi: 10.1080/09658211.2017.1332226
- Howe, M. L., Wilkinson, S., Garner, S. R., and Ball, L. J. (2016). On the adaptive function of children’s and adult’s false memories. *Memory* 24, 1062–1077. doi: 10.1080/09658211.2015.1068335
- Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity observed in people solving verbal problems with insight. *PLoS Biol.* 2:E97. doi: 10.1371/journal.pbio.0020097
- Kizilirmak, J. M., Hanne, T., Foltz-Schoofs, K., Schott, B. H., and Richardson-Klavehn, A. (2016). Neural correlates of learning from induced insight: A case for reward-based episodic encoding. *Front. Psychol.* 7:1693. doi: 10.3389/fpsyg.2016.01693
- 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, 1534–1555. doi: 10.1037/0278-7393.25.6.1534
- Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M., et al. (2008). The origins of insight in resting-state brain activity. *Neuropsychologia* 46, 281–291. doi: 10.1016/j.neuropsychologia.2007.07.013
- Li, Y., T. C., Tsai, A., Huang, A., and Lin, W. (2016). Different brain wave patterns and cortical abilities in relation to different creative potentials. *Creat. Res. J.* 28, 89–98. doi: 10.1080/10400419.2016.1125255
- MacGregor, J. N., and Cunningham, J. B. (2008). Rebus puzzles as insight problems. *Behav. Res. Methods* 40, 263–268. doi: 10.3758/BRM.40.1.263
- MacGregor, J. N., and Cunningham, J. B. (2009). The effects of number and level of restructuring in insight problem solving. *J. Problem Solv.* 2, 130–141. doi: 10.7771/1932-6246.1062
- Maier, N. R. (1930). Reasoning in humans. I. On direction. *J. Comp. Psychol.* 10, 115–143. doi: 10.1037/h0073232
- Mednick, S. A. (1968). Remote associates test. *J. Creat. Behav.* 2, 213–214. doi: 10.1002/j.2162-6057.1968.tb00104.x
- Metcalfe, J. (1986). Premonitions of insight predict impending error. *J. Exp. Psychol. Learn. Mem. Cogn.* 12, 623–634.
- Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. *Adv. Psychol. Think.* 1, 1–44.
- Ohlsson, S. (2011). *Deep Learning: How the Mind Overrides Experience*. Cambridge, MA: Cambridge University Press.
- Olteteanu, A. M., Schultheis, H., and Dyer, J. B. (2017). Computationally constructing a repository of compound remote associates test items in that was awarded to JM, LB, and ET (Grant No: SG162930). The data can be obtained by emailing the corresponding author.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02513/full#supplementary-material>

- American English with comRAT-G. *Behav. Res. Methods* 50, 1971–1980. doi: 10.3758/s13428-017-0965-8
- Runco, M. A. (2018). “Creative thinking,” in *The Routledge International Handbook of Thinking and Reasoning*, eds L. J. Ball and V. A. Thompson (Abingdon, OX: Routledge), 472–486.
- 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.
- Salvi, C., Constantini, G., Bricolo, E., Perugini, M., and Beeman, M. (2015). Validation of Italian Rebus puzzles and compound remote associate problems. *Behav. Res. Methods* 48, 664–685. doi: 10.3758/s13428-015-0597-9
- Sandkühler, S., and Bhattacharya, J. (2008). Deconstructing insight: EEG correlates of insightful problem solving. *PLoS ONE* 3:e1459. doi: 10.1371/journal.pone.0001459
- Shen, W., Yuan, Y., Liu, C., and Luo, J. (2017). The roles of the temporal lobe in creative insight: an integrated review. *Think. Reason.* 23, 321–375. doi: 10.1080/13546783.2017.1308885
- Smith, S. M., and Blankenship, S. E. (1989). Incubation effects. *Bull. Psychon. Soc.* 27, 311–314. doi: 10.3758/BF03334612
- Smith, S. M., Sifonis, C. M., and Angello, G. (2012). Clue insensitivity in remote associates test problem solving. *J. Problem Solv.* 4, 128–149. doi: 10.7771/1932-6246.1124
- Sowden, P. T., Pringle, A., and Gabora, L. (2014). The shifting sands of creative thinking: connections to dual-process theory. *Think. Reason.* 21, 40–60. doi: 10.1080/13546783.2014.885464
- Sternberg, R. J., and Davidson, J. E. (eds.). (1995). *The Nature of Insight*. Cambridge, MA: The MIT Press.
- Tabachnick, B. G., and Fidell, L. S. (2013). *Using Multivariate Statistics*. Harlow; Essex: Pearson Education Limited.
- Van Casteren, M., and Davis, M. (2007). Match: a program to assist in matching the conditions of factorial experiments. *Behav. Res. Methods* 39, 973–978. doi: 10.3758/BF03192992
- Viera, A. J., and Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Med.* 37, 360–363.
- 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
- Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Think. Reason.* 21, 5–39. doi: 10.1080/13546783.2014.886625
- Weisberg, R. W. (2018). “Problem solving,” in *The Routledge International Handbook of Thinking and Reasoning*, eds L. J. Ball and V. A. Thompson (Abingdon, OX: Routledge), 607–623.
- Wu, C. L., and Chen, H. C. (2017). Normative data for Chinese compound remote associate problems. *Behav. Res. Methods* 49, 2163–2172. doi: 10.3758/s13428-016-0849-3

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 Threadgold, Marsh and Ball. 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.



How Artificial Intelligence Can Help Us Understand Human Creativity

Fernand Gobet^{1*} and Giovanni Sala²

¹Department of Psychological Sciences, University of Liverpool, Liverpool, United Kingdom, ²Graduate School of Human Sciences, Osaka University, Suita, Japan

Recent years have been marked by important developments in artificial intelligence (AI). These developments have highlighted serious limitations in human rationality and shown that computers can be highly creative. There are also important positive outcomes for psychologists studying creativity. It is now possible to design entirely new classes of experiments that are more promising than the simple tasks typically used for studying creativity in psychology. In addition, given the current and future AI algorithms for developing new data structures and programs, novel theories of creativity are on the horizon. Thus, AI opens up entire new avenues for studying human creativity in psychology.

Keywords: artificial intelligence, bounded rationality, creativity, evolutionary computation, intelligence, simulation, scientific discovery, theory

OPEN ACCESS

Edited by:

Ian Hocking,
Canterbury Christ Church University,
United Kingdom

Reviewed by:

Colleen Seifert,
University of Michigan,
United States
Azlan Iqbal,
Universiti Tenaga Nasional, Malaysia
Masasi Hattori,
Ritsumeikan University, Japan

*Correspondence:

Fernand Gobet
fgobet@liv.ac.uk

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 15 May 2018

Accepted: 29 May 2019

Published: 19 June 2019

Citation:

Gobet F and Sala G (2019) How
Artificial Intelligence Can Help Us
Understand Human Creativity.
Front. Psychol. 10:1401.
doi: 10.3389/fpsyg.2019.01401

In psychology, research into creativity¹ has tended to follow well-trodden paths: simple tests of creativity (e.g., alternative uses test), correlations with measures of intelligence, and more recently neural correlates of creativity such as EEG and fMRI (e.g., Weisberg, 2006; Runco, 2014)². One line of research that has been little explored is to use progress in artificial intelligence (AI) to generate tools for studying human creativity.

Developments of AI have been impressive. DeepMind's AlphaGo has easily beaten the best human grandmasters in Go, a game that for many years had seemed beyond the reach of AI (Silver et al., 2016). IBM's Watson mastered natural language and knowledge to the point that it outclassed the best human players in Jeopardy! – a game show where contestants have to find the question to an answer (Ferrucci, 2012). Not less impressive, we are now on the brink of having self-driving cars and automated assistants able to book appointment by phone (Smith and Anderson, 2014). These developments raise profound issues about human identity; they also pose difficult but exciting questions about the very nature of human creativity and indeed rationality. But they also present novel opportunities for studying human creativity. Entirely new classes of experiments can be devised, going way beyond the simple tasks typically used

¹It is notably difficult to define “creativity,” and a large number of definitions exist with little agreement among researchers (see e.g., Hennessey and Amabile, 2010). In this article, we focus on what Boden (1990) calls “historical creativity” (concerning products that are considered novel by society at large) rather than “psychological creativity” (concerning products that are novel only for the agent producing them). Thus, if Joe Bloggs for the first time of his life realizes that a brick can be used as a pen holder, this is psychological but not historical creativity. If he is the first ever to claim that a brick can be used as an abstract rendition of Beethoven's 5th Symphony, this is both psychological and historical creativity according to Boden's definition.

²While the aim of this *Perspective Article* is not to provide a review of the extensive literature on creativity in psychology and neuroscience, a few additional pointers might be helpful to the reader: Cristofori et al. (2018); Kaufman and Sternberg (2019); and Simonton (2014).

so far for studying creativity, and new theories of creativity can be developed.

ARTIFICIAL INTELLIGENCE RESEARCH AND CREATIVITY

Using AI for understanding creativity has a long history and is currently an active domain of research with annual international conferences (for reviews, see Meheus and Nickles, 2009; Colton and Wiggins, 2012). As early as 1957, Newell, Simon, and Shaw had programmed Logic Theorist to prove theorems in symbolic logic. Not only did this research lead to an influential theory of problem-solving (Newell et al., 1958) but it also shed important light on human creativity, as Logic Theorist was able to prove some theorems in a more elegant way than Russell and Whitehead, two of the leading mathematicians of the twentieth century (Gobet and Lane, 2015). There are numerous examples of AI creativity in science today (Sozou et al., 2017). For example, at Aberystwyth University, a “robot scientist” specialized in functional genomics not only produced hypotheses independently but also designed experiments for testing these hypotheses, physically performed them and then interpreted the results (King et al., 2004).

In the arts, British abstract painter Harold Cohen all but abandoned a successful career as an artist to understand his own creative processes. To do so, he wrote a computer program, AARON, able to make drawings and later color paintings autonomously (McCorduck, 1990). More recently, several programs have displayed high levels of creativity in the arts. For example, a deep-learning algorithm produced a Rembrandt-like portrait (Flores and Korsten, 2016) and the program Aiva, also using deep learning, composes classical music (Aiva Technologies, 2018). An album of Aiva’s music has already been released, and its pieces are used in films and by advertising agencies. In chess, the program CHESTHETICA automatically composes chess problems and puzzles that are considered by humans as esthetically pleasing (Iqbal et al., 2016).

However, AI has had only little impact on creativity research in psychology (for an exception, see Oltețeanu and Falomir’s, 2015, 2016 work on modelling the Remote Associate Test and the Alternative Uses Test). There is only passing mention if at all in textbooks and handbooks of creativity (e.g., Kaufman and Sternberg, 2006; Runco, 2014), and mainstream research simply ignores it. In our view, this omission is a serious mistake.

THE SPECTER OF BOUNDED RATIONALITY

AI has uncovered clear limits in human creativity, as is well illustrated by Go and chess, two board games requiring creativity when played competitively. After losing 3–0 against computer program AlphaGo Master in 2017, Chinese Go grandmaster Ke Jie, the world No. 1, declared: “After humanity spent thousands of years improving our tactics, computers tell us that humans

are completely wrong... I would go as far as to say not a single human has touched the edge of the truth of Go” (Kahn, 2017). Astonishingly, this version of AlphaGo, which won not only all its games against Ke Jie but also against other leading Go grandmasters, was beaten 89–11 a few months later by AlphaGo Zero, a new version of the program that learns from scratch by playing against itself, thus creating all its knowledge except for the rules of the game (Silver et al., 2016, 2017).

Ke Jie’s remark is echoed by chess grandmasters’ comments (Gobet, 2018). In the second game of his 1997 match against Deep Blue, Kasparov and other grandmasters were astonished by the computer’s sophisticated and creative way of first building a positional advantage and then denying any counter-play for Kasparov. Kasparov’s surprise was such that he accused IBM and the programming team behind Deep Blue of cheating, a charge that he maintained for nearly 20 years. More recently, in the sixth game of the 2006 match between Deep Fritz and world champion Vladimir Kramnik, the computer played a curious rook maneuver that commentators ridiculed as typical of a duffer. As the game unfolded, it became clear that this maneuver was a very creative way of provoking weaknesses on Kramnik’s kingside, which allowed Deep Blue to unleash a fatal offensive on the other side of the board.

In general, these limits in rationality and creativity are in line with Simon’s theory of bounded rationality (Simon, 1956, 1997; Gobet and Lane, 2012; Gobet, 2016a), which proposed that limitations in knowledge and computational capacity drastically constrain a decision maker’s ability to make rational choices. These limits are also fully predictable from what we know from research in cognitive psychology. For example, Bilalić et al. (2008) showed that even experts can be blinded by their knowledge, with the consequence that they prefer standard answers to novel and creative answers, even when the latter are objectively better. Thus, when a common solution comes first to mind, it is very hard to find another one (a phenomenon known as the *Einstellung* effect). In Bilalić et al.’s chess experiment, the effect was powerful: compared to a control group, the strength of the *Einstellung* group decreased by about one standard deviation.

The power of long-term memory schemas and preconceptions is a common theme in the history of science and art and has often thwarted creativity. For example, in the early 1980s, the unquestioned wisdom was that stomach ulcers were caused by excess acid, spicy food, and stress. The genius of Marshall and Warren (1984) in their Nobel-winning discovery was to jettison all these assumptions before hypothesizing that a bacterium (*helicobacter pylori*) was the main culprit. Finding ways to overcome such mind-sets is an important task for fostering human creativity (Gobet et al., 2014), as they are common with normal cognition. In some instances, in order to be creative and explore new conceptual spaces, it is necessary to break these mind-sets, either by inhibiting some specific concepts or groups of concepts, or by eschewing concepts altogether. AI systems can use a large variety of different methods – some similar to those used by humans, some entirely dissimilar. Thus, they are less likely to be subject to such mind-sets and could provide humans with useful alternatives for developing creative products.

ARTIFICIAL INTELLIGENCE OFFERS NOVEL METHODS FOR STUDYING CREATIVITY

When considering the literature on creativity in psychology, it is hard to escape the feeling that something is amiss in this field of research. A considerable amount of research has studied simple tasks that are remote from real creativity in the arts and science – for example, alternative uses task, word generation task, and insight problems (see e.g., Runco, 2014) – but it is at the least debatable whether these tasks tell us much about real creativity. As support for this critique of the lack of ecological validity of many tasks used in the field, numerous experiments have found that these tasks correlate more with general intelligence (g) and verbal intelligence than with real-world creativity (Wallach, 1970; Silvia, 2015). In addition, in their review of the literature, Zeng et al. (2011) conclude that divergent-thinking tests suffer from six major weaknesses, including poor predictive, ecological, and discriminant validities. (For a more positive evaluation, see Plucker and Makel, 2010.) While some researchers have developed tasks that map more directly into the kind of tasks carried out in real-world creativity – see in particular the research on scientific discovery (Klahr and Dunbar, 1988; Dunbar, 1993) – this approach is relatively underrepresented in research into creativity.

A similar concern can be voiced with respect to experimentation and theory development. Although a fair amount of avenues have been explored – including generation and selection (e.g., Simonton, 1999), heuristic search (e.g., Newell et al., 1962), problem finding (e.g., Getzels and Csikszentmihalyi, 1976), systems theories (e.g., Gruber, 1981), explanations based on intelligence (e.g., Eysenck, 1995), and psychopathological explanations (e.g., Post, 1994) – entire experimental and theoretical spaces have been fully ignored or, in the best case, barely scratched. Clearly, this is due to the limits imposed by human bounded rationality, to which one should add the constraints imposed by the limited time resources available.

AI can help with both empirical and theoretical research. Empirically, it can simulate complex worlds that challenge human creativity; theoretically, it can help develop new theories by inhibiting some concepts (see above), making unexpected connections between known mechanisms or proposing wholly new explanations. Here we focus on scientific discovery, but similar conclusions can be reached for creativity in the arts.

A New Way of Designing Experiments

AI can be used as a new way to perform experiments on creativity. The central idea is to exploit current technology to design complex environments that can be studied with a creative application of the scientific method. Thus, these experiments go way beyond the simple tasks typically used in creativity research. Rather than studying creativity asking people to generate words that are related to three stimulus

words as in the Remote Associates Test (Mednick, 1962), one studies it by asking participants to find the laws of a simulated world. This is of course what Dunbar, Klahr, and others did in earlier experiments (Klahr and Dunbar, 1988; Dunbar, 1993). The key contribution here is to propose to use much more complex environments, including environments where the presence of intelligent agents approximates the complexity of studying phenomena affected by humans, as is the case in psychology and sociology. Thus, where standard programming techniques are sufficient for simulating physical worlds with no intelligent agents, AI techniques make it possible to simulate much more complex worlds, which incorporate not only physical and biological laws, but also psychosocial laws. In both cases, the participants' task is to reverse-engineer at least some of the laws of the domains – that is to make scientific discoveries about these domains. Thus, for example, participants must devise experiments for understanding the learning mechanisms of agents inhabiting a specific world. The mechanisms and laws underpinning these worlds can be similar to those currently postulated in science, or wholly different with new laws of physics, biology, or psychology. In that case, the situation is akin to scientists exploring life on a new planet.

These environments can be used with several goals in mind. First, they can test current theories of creativity and scientific discovery. The worlds can be designed in such a way that their understanding is facilitated by the mechanisms proposed by some theories as opposed to others (e.g., heuristic search might be successful, but randomly generating concepts might not, or *vice versa*). Additional questions include whether participants adapt their strategy as a function of the results they obtain and whether they develop new experimental designs where necessary. Second, these environments can be used to observe new empirical phenomena related to creativity, such as the generation of as yet unknown strategies. New phenomena are bound to occur, as the complexity of the proposed tasks is larger by several orders of magnitude than the tasks typically studied in psychology.

A third use is to identify creative people in a specific domain, for example in biology or psychology. As creativity is measured in a simulated environment that is close to the target domain, one is more likely to correctly identify individuals that might display creativity in the domain. If one wishes, one can correlate performance in the task and other behavioral measures with standard psychological measures such as IQ, motivation, and psychoticism.

A final use is to train people to be creative in a specific domain. Variables in the environment can be manipulated such that specific skills are taught, for example the efficient use of heuristics or standard research methods in science. The difficulty of finding laws can be manipulated as well: from a clear linear relation between two variables to non-linear relations between several variables with several sources of noise. The reader will have noticed that such environments are not dissimilar from some video games, and this game-like feature can be used to foster enjoyment and motivation, and thus learning.

Please note that we make no claim that training creativity in one domain will provide something like general creativity, as is sometimes proposed in the literature (e.g., De Bono, 1970). There is now very strong experimental evidence that skills acquired in a domain do not generalize to new domains sharing few commonalities with the original one (Gobet, 2016b; Sala and Gobet, 2017a), and this conclusion almost certainly also applies to creativity. One possible reason for this lack of far transfer is that expertise relies on the ability of recognizing patterns that are specific to a domain (Sala and Gobet, 2017b). It is possible to speculate that being creative relies, at least in part, on recognizing rare domain-specific patterns in a problem situation. For example, to go back to the example of discovering that stomach ulcers are caused by bacteria, Warren recognized the presence of bacteria in gastric specimens he studied with a microscope, although this was not expected as it was thought that the stomach was a sterile environment inhospitable for bacteria (Thagard, 1998). However, we do recognize that this is a hypothesis that should be tested, and it could turn out that, in fact, creativity is a general ability. This is an empirical question that can only be settled with new experiments, and the methods proposed in this paper may contribute to its answer.

Automatic Generation of Theories

As noted above, human bounded rationality has the consequence that humans only explore a very small number of subspaces within the space of all possible theories, and even these subspaces are explored only sparsely. Mind-sets and other biases mean that even bad hypotheses are maintained while more promising ones are ignored. AI can help break these shackles.

The subfield of AI known as computational scientific discovery has been active for decades, spearheaded by Herbert Simon's seminal work (Newell et al., 1962; Bradshaw et al., 1983). The aim is precisely to develop algorithms that can produce creative behavior in science, either replicating famous scientific discoveries or making original contributions (for a review, see Sozou et al., 2017). Due to space constraints, we limit ourselves to the description of only one approach – Automatic Generation of Theories (AGT) (Lane et al., 2014) – which is particularly relevant to our discussion as it excels in avoiding being stuck in local minima, contrary to human cognition which is notably prone to mind-sets, Einstellung effects, and other cognitive biases. In a nutshell, the central ideas of AGT are (1) to consider theories as computer programs; (2) to use a probabilistic algorithm (genetic programming) to build those programs; (3) to simulate the protocols of the original experiments; (4) to compare the predictions of the theories with empirical data in order to compute the quality (fitness) of the theories; and (5) to use fitness to evolve better theories, using mechanisms of selection, mutation, and crossover. Simulations have shown that the methodology is able to produce interesting theories with simple experiments. With relentless progress in technology, it is likely that this and other approaches in artificial scientific discovery will provide theoretical explanations for more complex human behaviors, including creativity itself.

Challenges

The two uses of AI proposed in this paper for studying creativity in psychology are not meant to replace current methods, but to add to the arsenal of theoretical concepts and experimental techniques available to researchers. Nor are they proposed as magic bullets that will answer all questions related to creativity. Our point is that these uses of AI present potential benefits that have been overlooked by psychologists studying creativity.

As any new approach, these uses raise conceptual and methodological challenges. Regarding the proposed method for collecting data, challenges include the way participants' results will be scored and compared, and how they will be used to test theories. A related challenge concerns the kind of theory suitable to account for these data; given the complexity and richness of the data, it is likely that computational models will be necessary – possibly models generated by the second use of AI we proposed.

Similarly, using AI for generating theories raises interesting practical and theoretical questions. Will the generated theories be understandable to humans, or will they only be black boxes providing correct outputs (predictions) given a description of the task at hand and other kind of information such as the age of the participants? Will their structure satisfy canons of parsimony in science? How will they link epistemologically to other theories in psychology, for example theories of memory and decision-making? Will they be useful for practical applications such as training experts to be creative in their specialty? In addition, there is of course the question as to what kind of AI is best suited for generating theories. We have provided the example of genetic programming, but many other techniques can be advanced as candidates, including adaptive production systems (Klahr et al., 1987) and deep learning (LeCun et al., 2015).

PROBLEMS AND PROSPECTS

Recent developments in AI signal a new relationship between human and machine. Interesting albeit perhaps threatening questions are posed about our human nature and, specifically, the meaning of creativity. These include philosophical and ethical questions. Can a product be creative if it is conceived by a computer? If so, who owns the research? Should computer programs be listed as co-authors of scientific papers? How will the synergy between human and computer creativity evolve? Should some types of creativity – e.g., generating fake news for political aims – be curtailed or even banned?

These developments also raise significant questions about human rationality, as discussed above. In doing so, they highlight the magnificent achievements of some human creators, such as Wolfgang Amadeus Mozart or Pablo Picasso. In addition, they have substantial implications for creativity in science and the arts. Entirely new conceptual spaces might be explored, with computer programs either working independently or co-designing creative products with humans. In science – the focus of this perspective article – this might lead to the development of novel research strategies,

methodologies, types of experiments, theories, and theoretical frameworks. Of particular interest is the possibility of mixing concepts and mechanisms between different subfields (e.g., between memory research and decision-making research), between different fields (e.g., psychology and chemistry), and even between science and the arts. As discussed above, there are also some new exciting opportunities for training. It is only with the aid of artificial creativity that we will break our mind-sets and reach a new understanding of human creativity.

REFERENCES

- Aiva Technologies. (2018). Available at: <http://www.aiva.ai> (Accessed September 08, 2018).
- Bilalić, M., McLeod, P., and Gobet, F. (2008). Inflexibility of experts: reality or myth? Quantifying the Einstellung effect in chess masters. *Cogn. Psychol.* 56, 73–102. doi: 10.1016/j.cogpsych.2007.02.001
- Boden, M. (1990). *The creative mind*. (New York: BasicBooks).
- Bradshaw, G., Langley, P. W., and Simon, H. A. (1983). Studying scientific discovery by computer simulation. *Science* 222, 971–975. doi: 10.1126/science.222.4627.971
- Colton, S., and Wiggins, G. A. (2012). “Computational creativity: the final frontier?” in *Proceedings of the 20th European conference on artificial intelligence*. eds. L. De Raedt, C. Bessière, D. Dubois, P. Doherty, P. Frasconi, F. Heintz, and P. Lucas (Montpellier, France: IOS Press), 21–26.
- Cristofori, I., Salvi, C., Beeman, M., and Grafman, J. (2018). The effects of expected reward on creative problem solving. *Cogn. Affect. Behav. Neurosci.* 18, 925–931. doi: 10.3758/s13415-018-0613-5
- De Bono, E. (1970). *Lateral thinking: Creativity step by step*. (New York: Harper & Row).
- Dunbar, K. (1993). Concept discovery in a scientific domain. *Cogn. Sci.* 17, 397–434. doi: 10.1207/s15516709cog1703_3
- Eysenck, H. J. (1995). *Genius: The natural history of creativity*. (New York: Cambridge University Press).
- Ferrucci, D. A. (2012). Introduction to “This is Watson”. *IBM J. Res. Dev.* 56, 1:1–1:15. doi: 10.1147/JRD.2012.2184356
- Flores, E., and Korsten, B. (2016). The Next Rembrandt. Available at: <http://www.nextrembrandt.com/> (Accessed September 08, 2018).
- Getzels, J. W., and Csikszentmihalyi, M. (1976). *The Creative vision: A longitudinal study of problem finding in art*. (New York: John Wiley & Sons).
- Gobet, F. (2016a). “From bounded rationality to expertise” in *Minds, models and milieux: Commemorating the centenary of Herbert Simon's birth*. eds. R. Frantz and L. Marsh (New York: Palgrave Macmillan), 151–166.
- Gobet, F. (2016b). *Understanding expertise: A multidisciplinary approach*. (London: Palgrave).
- Gobet, F. (2018). *The psychology of chess*. (London: Routledge).
- Gobet, F., and Lane, P. C. R. (2012). “Bounded rationality and learning” in *Encyclopedia of the sciences of learning*. ed. N. M. Seel (New York, NY: Springer).
- Gobet, F., and Lane, P. C. R. (2015). “Human problem solving – Beyond Newell et al.'s (1958): elements of a theory of human problem solving” in *Cognitive psychology: Revisiting the classic studies*. eds. M. W. Eysenck and D. Groome (Thousand Oaks, CA: Sage).
- Gobet, F., Snyder, A., Bossomaier, T., and Harre, M. (2014). Designing a “better” brain: insights from experts and savants. *Front. Psychol.* 5:470. doi: 10.3389/fpsyg.2014.00470
- Gruber, H. E. (1981). *Darwin on man: A psychological study of scientific creativity*. Rev. edn. (Chicago: University of Chicago Press).
- Hennessey, B. A., and Amabile, T. M. (2010). Creativity. *Annu. Rev. Psychol.* 61, 569–598. doi: 10.1146/annurev.psych.093008.100416
- Iqbal, A., Guid, M., Colton, S., Krivec, J., Azman, S., and Haghighi, B. (2016). *The digital synaptic neural substrate: A new approach to computational creativity*. (Switzerland: Springer International Publishing).
- Kahn, J. (2017). Robots are going to take our jobs and make us look like fools while doing it. Available at: [https://medium.com/bloomberg/robots-](https://medium.com/bloomberg/robots-are-going-to-take-our-jobs-and-make-us-look-like-fools-while-doing-it-ec25b05a5910)

AUTHOR CONTRIBUTIONS

Both authors conceptualized the paper. FG wrote the first draft of the paper and GS contributed to drafting its final version.

FUNDING

GS is a JSPS International Research Fellow (grant number: 17F17313).

- are-going-to-take-our-jobs-and-make-us-look-like-fools-while-doing-it-ec25b05a5910 (Accessed May 11, 2018).
- Kaufman, J. C., and Sternberg, R. J. (eds.) (2006). *The international handbook of creativity*. (Cambridge, UK: Cambridge University Press).
- Kaufman, J. C., and Sternberg, R. J. (eds.) (2019). *The Cambridge handbook of creativity*. (Cambridge: Cambridge University Press).
- King, R. D., Whelan, K. E., Jones, F. M., Reiser, P. G. K., Bryant, C. H., Muggleton, S. H., et al. (2004). Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* 427, 247–252. doi: 10.1038/nature02236
- Klahr, D., and Dunbar, K. (1988). Dual space search during scientific reasoning. *Cogn. Sci.* 12, 1–48. doi: 10.1207/s15516709cog1201_1
- Klahr, D., Langley, P., and Neches, R. (1987). *Production system models of learning and development*. (Cambridge, MA: MIT Press).
- Lane, P., Sozou, P., Addis, M., and Gobet, F. (2014). “Evolving process-based models from psychological data using genetic programming” in *Proceedings of the 50th anniversary convention of the AISB: Computational scientific discovery symposium*. eds. M. Addis, F. Gobet, P. Lane, and P. Sozou (London: AISB).
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Marshall, B. J., and Warren, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* 823, 1311–1315.
- McCorduck, P. (1990). *AARON'S code: Meta-art, artificial intelligence, and the work of Harold Cohen*. (New York: W. H. Freeman & Co).
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* 69, 220–232. doi: 10.1037/h0048850
- Meheus, J., and Nickles, T. (2009). *Models of discovery and creativity*. (New York: Springer).
- Newell, A., Shaw, J. C., and Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychol. Rev.* 65, 151–166. doi: 10.1037/h0048495
- Newell, A., Shaw, J. C., and Simon, H. A. (1962). “The process of creative thinking” in *Contemporary approaches to creative thinking*. Vol. 3, eds. H. E. Gruber, G. Terrell, and Wertheimer (New York: Atherton Press), 63–119.
- Olteanu, A.-M., and Falomir, Z. (2015). comRAT-C: a computational compound remote associates test solver based on language data and its comparison to human performance. *Pattern Recogn. Lett.* 67, 81–90. doi: 10.1016/j.patrec.2015.05.015
- Olteanu, A.-M., and Falomir, Z. (2016). Object replacement and object composition in a creative cognitive system. Towards a computational solver of the alternative uses test. *Cogn. Syst. Res.* 39, 15–32. doi: 10.1016/j.cogsys.2015.12.011
- Plucker, J. A., and Makel, M. C. (2010). “Assessment of creativity” in *The Cambridge handbook of creativity*. eds. J. C. Kaufman and R. J. Sternberg (Cambridge: Cambridge University Press), 48–73.
- Post, F. (1994). Creativity and psychopathology: a study of 291 world-famous men. *Br. J. Psychiatry* 165, 22–34. doi: 10.1192/bjp.165.1.22
- Runco, M. A. (2014). *Creativity theories and themes: Research, development, and practice*. (New York: Academic Press).
- Sala, G., and Gobet, F. (2017a). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Curr. Dir. Psychol. Sci.* 26, 515–520. doi: 10.1177/0963721417712760
- Sala, G., and Gobet, F. (2017b). Experts' memory superiority for domain-specific random material generalizes across fields of expertise: a meta-analysis. *Mem. Cogn.* 45, 183–193. doi: 10.3758/s13421-016-0663-2

- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489. doi: 10.1038/nature16961
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. (2017). Mastering the game of Go without human knowledge. *Nature* 550, 354–359. doi: 10.1038/nature24270
- Silvia, P. (2015). Intelligence and creativity are pretty similar after all. *Educ. Psychol. Rev.* 27, 599–606. doi: 10.1007/s10648-015-9299-1
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychol. Rev.* 63, 129–138. doi: 10.1037/h0042769
- Simon, H. A. (1997). *Models of bounded rationality*. Vol. 3, (Cambridge, MA: The MIT Press).
- Simonton, D. K. (1999). *Origins of genius*. (Oxford: Oxford University Press).
- Simonton, D. K. (ed.) (2014). *The Wiley-Blackwell handbook of genius*. (Oxford, UK: Wiley-Blackwell).
- Smith, A., and Anderson, J. (2014). *Predictions for the state of AI and robotics in 2025*. (Washington, DC: Pew Research Center).
- Sozou, P. D., Lane, P. C., Addis, M., and Gobet, F. (2017). “Computational scientific discovery” in *Springer handbook of model-based science*. eds. L. Magnani and T. Bertolotti (New York: Springer), 719–734.
- Thagard, P. (1998). Ulcers and bacteria: discovery and acceptance. *Stud. Hist. Philos. Sci. C* 29, 107–136.
- Wallach, M. A. (1970). “Creativity” in *Carmichael’s manual of child psychology*. ed. P. H. Mussen (New York: Wiley), 1273–1365.
- Weisberg, R. W. (2006). *Creativity*. (New York: Wiley).
- Zeng, L., Proctor, R. W., and Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creat. Res. J.* 23, 24–37. doi: 10.1080/10400419.2011.545713

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 Gobet and Sala. 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.



Use or Consequences: Probing the Cognitive Difference Between Two Measures of Divergent Thinking

Richard W. Hass^{1*} and Roger E. Beaty²

¹ College of Humanities and Sciences, Thomas Jefferson University-East Falls, Philadelphia, PA, United States, ² Department of Psychology, Pennsylvania State University, University Park, PA, United States

OPEN ACCESS

Edited by:

Phillip A. Fine,
University of Buckingham,
United Kingdom

Reviewed by:

Kenneth James Gilhooly,
Brunel University London,
United Kingdom
Boris Forthmann,
Universität Münster, Germany

*Correspondence:

Richard W. Hass
richard.hass@jefferson.edu

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 30 April 2018

Accepted: 06 November 2018

Published: 27 November 2018

Citation:

Hass RW and Beaty RE (2018) Use or
Consequences: Probing the Cognitive
Difference Between Two Measures of
Divergent Thinking.
Front. Psychol. 9:2327.
doi: 10.3389/fpsyg.2018.02327

Recent studies have highlighted both similarities and differences between the cognitive processing that underpins memory retrieval and that which underpins creative thinking. To date, studies have focused more heavily on the Alternative Uses task, but fewer studies have investigated the processing underpinning other idea generation tasks. This study examines both Alternative Uses and Consequences idea generation with a methods pulled from cognitive psychology, and a novel method for evaluating the creativity of such responses. Participants were recruited from Amazon Mechanical Turk using a custom interface allowing for requisite experimental control. Results showed that both Alternative Uses and Consequences generation are well approximated by an exponential cumulative response time model, consistent with studies of memory retrieval. Participants were also slower to generate their first consequence compared with first responses to Alternative Uses, but inter-response time was negatively related to pairwise similarity on both tasks. Finally, the serial order effect is exhibited for both tasks, with Consequences earning more creative evaluations than Uses. The results have implications for burgeoning neuroscience research on creative thinking, and suggestions are made for future areas of inquiry. In addition, the experimental apparatus described provides an equitable way for researchers to obtain good quality cognitive data for divergent thinking tasks.

Keywords: creativity, divergent thinking, memory search, default mode network, semantic memory

1. INTRODUCTION

Creative thinking studies have long depended on classic divergent thinking tasks as operationalizations of the construct. With recent emergence of studies using a variety of neuroimaging techniques to examine the cognitive roots of performance on the tasks, there became a need for more probing cognitive analyses of divergent thinking. To some extent, this has been done (Beaty et al., 2014; Forthmann et al., 2016; Acar and Runco, 2017; Hass, 2017a) but such analyses have focused almost exclusively on responses to the Alternative Uses task, in which participants are asked to generate as many creative uses for common objects as possible within a specified time period (usually 2–3 min).

This study was motivated by several perceived gaps in both methodology and theory of another oft-used divergent thinking task: the Consequences task (Wilson et al., 1954; Torrance, 1974). First, when participants generate responses to prompts from the consequences task (e.g., “imagine that humans no longer needed to sleep”), it is not altogether clear whether the idea generation process

unfolds in a similar fashion to idea generation to Alternative Uses prompts (e.g., “think of creative uses for a brick”). Second, it is somewhat more difficult for judges to agree on creativity ratings assigned to consequences responses (Silvia et al., 2008; Hass et al., 2018). Indeed, it seems that scoring consequences tasks involves an increase in cognitive load over the scoring of alternative uses tasks (Forthmann et al., 2017). Finally, though other researchers are beginning to examine response time distributions as evidence of cognitive processing during divergent thinking (Acar and Runco, 2017), the full scope of analyses that can be done with response times has not fully been explicated.

The novel components of this study follow from the points just raised. In this paper we present a web-based data collection methodology for divergent thinking tasks (and indeed any kind of creative thinking task that urges multiple responses), which was designed using the tools created by the psiTurk group (McDonnell et al., 2012). There are indeed other methods by which one can use electronic means of collecting DT data, but the importance of this web-based tool is that psiTurk was designed specifically to allow researchers to collect precise cognitive data from workers on the Amazon Mechanical Turk platform (MTurk). As such, it allows for researchers of all levels to easily collect creative thinking data from a more representative sample than is often available on university campuses. In addition, we illustrate how cognitive theory can be applied to response time data culled from both Alternative Uses and Consequences tasks. Finally, we use a newly validated scale for measuring creativity of responses, along with human rated similarity of responses to compare and contrast the response generation process across these two tasks.

1.1. Divergent Thinking and Memory Processes

The study of divergent thinking in general has spanned generations of creativity researchers. Though the tasks that measure divergent thinking are disparate (e.g., Forthmann et al., 2018), and may not be interchangeable (cf. Silvia, 2011; Runco et al., 2016), this study was focused on cognitive analyses of the acts of generating alternative uses for objects, and generating consequences of impossible situations. Specifically, the central question of this analyses was whether or not the memory processes involved in generating these two types of divergent thinking responses overlap, or are distinct. Do answer that question, methods culled from the cognitive science of memory recall were used in conjunction with methods from the creativity literature. This section summarizes the relevant aspects of the cognitive science of memory recall.

Several past results in the literature on memory retrieval provide a context for the current study. First, in one of the foundational studies on divergent thinking, Christensen et al. (1957) plotted the number cumulative responses to various cues as a function of time elapsed. Along with alternative uses cues, the authors plotted results from the classic Bousfield and Sedgewick (1944) study of cumulative responding and semantic memory retrieval, in which they derived the well known negative exponential function which describes the decreasing output rate

for generating category exemplars like fruits and animals. Wixted and Rohrer (1994) reviewed the results of subsequent studies concluding that the function is evidence of a repeated sampling of semantic space during memory retrieval, which is then depleted leading to more false retrievals, and an exponential slowing of retrieval rate (see also Raaijmakers and Shiffrin, 1981). Hass (2017a) found similar exponential slowing of response rates when participants generated uses for objects, but also found that generating uses yields lower response totals, and that response arrays were looser in terms of pairwise semantic relationships (cf. Troyer et al., 1997).

The Christensen et al. (1957) study on divergent thinking provided more direct evidence of differences, not only between creative idea generation and memory retrieval, but among different idea generation prompts. First, the output totals for divergent thinking cues were among the lowest reported (the lowest being output totals for words containing the letters M, T, or D, a very constraining memory retrieval task). Second, the cumulative response curves for two divergent thinking prompts: alternative uses for a brick, and impossibilities (“think of all of the impossible things”), were more linear than those for classic memory retrieval cues (e.g., U.S. cities). The “impossibilities” prompt is similar to the more common “consequences” prompt, with the latter simply specifying an impossibility of which participants generate consequences, while the former involves participants generating impossibilities with no specific context. So on that basis, there may be little difference in cumulative output when comparing alternative uses prompts to consequences prompts. However, in the Christensen and colleagues study, participants generated ideas for over 10 min, and the cumulative response functions were plotted across 2 min blocks. It may be that a more granular analysis of cumulative responding will yield subtle differences in the output functions when alternative uses and consequences prompts are compared. Indeed, there is reason to believe that differences should exist between the two tasks, and the argument forwarded presently is that alternative uses responding and consequences responding may rely on different contributions of episodic and semantic memory and also reasoning.

1.1.1. Episodic and Semantic Memory and Divergent Thinking

Much of the existing work on characterizing the contributions of memory retrieval to divergent thinking has focused on semantic memory (e.g., Gilhooly et al., 2007; Abraham and Bubic, 2015; Kenett et al., 2016; Hass, 2017a,b). One line of research has examined how individual differences in semantic retrieval ability (i.e., verbal fluency or “broad retrieval ability”) relates to divergent thinking fluency and originality. Silvia et al. (2013) administered a battery of verbal fluency tasks, corresponding to lower-order facets of retrieval ability (e.g., associational fluency; listing as many words in a given category as possible), and found that a higher-order “retrieval ability” factor comprised of the lower-order factors strongly predicted the quantity and quality of ideas generated on the Alternative Uses task, suggesting that the general ability to fluently retrieve a range of concepts from semantic memory is central to verbal

divergent thinking performance (see also Benedek et al., 2012; Avitia and Kaufman, 2014). Subsequent work has found that both controlled access to semantic memory (via verbal fluency) and the underlying structure of semantic concepts in memory contribute to divergent thinking (Beaty et al., 2014; Benedek et al., 2017), lending support to so-called “dual-process” models of creative cognition that emphasize the involvement of both top-down (executive) and bottom-up (associative) processes (Barr et al., 2015; Sowden et al., 2015).

Aside from the contributions of broad retrieval abilities, recent analyses of divergent thinking using network analysis (Kenett et al., 2014) and response time analysis (Hass, 2017a) have illustrated that the structure of semantic memory influences divergent thinking responding (see also Forthmann et al., 2016). Kenett and colleagues showed that a more “flexible” semantic network structure relates to high-divergent thinking ability and self-reported creative achievement (Kenett et al., 2016), likely reflecting an organization of semantic memory that is more conducive to establishing more remote conceptual links. Building off of Hass’s work, Xu (2017) further showed that when constraining participants to only think of “new” ideas during alternative uses responding, the response time functions were more linear, yielding higher predicted output totals, and higher originality, compared to phases in which participants were instructed to think of “old” ideas. Finally, there seems to be a robust serial order effect in alternative uses responses such that early responses earn lower creativity ratings than responses generated later in the responding interval (Christensen et al., 1957; Beaty and Silvia, 2012; Hass, 2017b; Wang et al., 2017).

A recent verbal protocol analyses of divergent thinking, Gilhooly et al. (2007) showed that the retrieval of known uses for objects from episodic memory dominates initial alternative uses responding. That result provides an explanation for the serial order effect such that known object uses should be rated as less creative than uses created on the spot by participants. The results presented by Gilhooly and colleagues also spurred a number of studies designed to test whether an “episodic specificity induction”, an exercise where participants are trained to retrieve details from “recent experiences”, affected the fluency and flexibility of divergent thinking responding (e.g., Madore et al., 2015, 2016, 2017). Madore et al. (2015) showed that the induction enhanced the number of categories of uses (also known as flexibility) during divergent thinking, but did not enhance the number of objects generated in an association task. Similarly, Madore et al. (2016) showed that the induction enhanced responding on both an alternative uses and a consequences task. However, in the latter study, the effects were constrained to counts of participant-rated “old” vs. “new” responses (following Benedek et al., 2014). Madore et al. (2016) pointed out that the participants reported generating many more “new” responses on the consequences task, leading them to conclude that the task relies less on recalling specific episodes compared with alternative uses responding.

This characterization is also in line with increasing evidence from functional brain imaging research. Several functional MRI studies have reported activation within a set of brain regions collectively known as the default network (DN) when

participants are engaged in creative thinking tasks in the scanner. The DN shows robust engagement during episodic memory retrieval and episodic future simulation tasks, which require the flexible recombination of episodic content (e.g., people, places, and actions) to reconstruct past experiences and imagine possible future experiences (Buckner et al., 2008). As noted above, Madore et al. (2015) have shown that an episodic specificity induction selectively enhances performance on the AUT, potentially reflecting the involvement of constructive episodic retrieval mechanisms (Schacter and Madore, 2016).

A recent fMRI study involved administering the episodic induction in the scanner and found that the induction was associated with increased divergent thinking performance, which corresponded to increased activity within the left anterior hippocampus (Madore et al., 2017), a region within the DN involved in episodic simulation. Several other studies have reported functional connectivity (i.e., correlation in neural responses) between regions of the DN and regions involved in cognitive control associated with creative task performance (Green et al., 2015; Beaty et al., 2017a,b, 2018; Gao et al., 2017; Zhu et al., 2017; Bendetowicz et al., 2018; Chen et al., 2018; Shi et al., 2018; Sun et al., 2018; Vartanian et al., 2018). Cooperation between DN and control regions is thought to reflect an interplay between idea generation and idea evaluation, retrieving possible solutions from memory and modifying them to fit task constraints (Beaty et al., 2016; Beaty and Schacter, 2018).

1.2. Differentiating Uses and Consequences Tasks

The points raised in the preceding discussion lean heavily on the use of the Alternative Uses task as the proxy measure of creative thinking (but see Addis et al., 2016). Given that discussion, it seems clear that Alternative Uses responding begins with a memory search process similar to the search that unfolds when people generate members of a well learned category. However, as responding continues, people rely less on known instances of an object’s use, and begin to exploit properties of objects to discover new uses via some sort of simulation process. Individual differences in the ability to generate creative uses has been tied both to fluid intelligence and to functional connectivity between cognitive control brain regions and memory-related regions within the DN. However, it is unclear if these conclusions extend to idea generation when the Consequences task is used as the proxy measure of creative thinking.

As mentioned, Madore et al. (2016) provided evidence that the reliance on episodic memory retrieval is weaker for the consequences task, but that result requires further investigation. There are other mechanisms that may be at work during consequences responding beyond memory retrieval and episodic simulation. To name one, the consequences task may require a form of counterfactual reasoning (Byrne, 2002; Abraham and Bubic, 2015) such that participants must consider what would happen if an enduring property of the world changed (e.g., gravity ceased to exist). However, when cognitive psychologists study counterfactual reasoning, the experimental methods often

require participants to learn about novel situations and then create counterfactuals using reasoning (e.g., about placing bets Dixon and Byrne, 2011). Analysis usually focuses on how participants' reasoning changes based on information contained in the description of the event (e.g., contrasting "normal" behavior of the agent with "extraordinary" behavior). In those studies, counterfactual reasoning is a given, and the goal is to discover how context influences the course of reasoning. In the consequences task specifics, a participant is supplied with a counterfactual antecedent (*if humans no longer need sleep*) and must then supply as many consequences (*then humans will not need to do X*) as possible (Forthmann et al., 2017). The goal of most studies using consequences tasks is simply to provide a proxy of creative thinking that can be correlated with other variables, or contrasted across groups. Thus, it may be difficult to ascertain whether or not counterfactual reasoning is at work during consequences responding. Still, the general hypothesis that can be tested currently is that consequences generation entails a lengthier processes compared with uses generation due primarily to additional reasoning that might be required.

The aim of this study was to use response time models, human rated semantic similarity, and a newly validated rating scale for DT responses to attempt to distinguish the course of alternative uses responding from consequences responding. There are three distinct predictions that follow from such analyses and the information reviewed in previous sections. First, the serial order effect for alternative uses responding seems to be a function of the early reliance on episodic retrieval and then the continued use of episodic and semantic simulation to derive more and more remote associations between known properties of objects and novel uses for those objects. Given that the consequences task often results in "new" responses, which may indicate that episodic memory is less of a factor, we hypothesize that the serial order effect should either be flatter for consequences responses. That is, it may be that when responding to consequences items, instead of searching quickly for a specific episode (which indeed seems impossible) participants instead arrive at a consequence through some type of reasoning (possibly counterfactual reasoning). For example, if given the prompt to think of [creative] consequences that would result if humans no longer needed sleep, a participant might search for knowledge related to sleep, and then use counterfactual reasoning to derive successive consequences (i.e., consider what might [not] happen if those facts about sleep became false). This, in turn, would yield a potentially more creative response earlier in the response sequence, thus affecting the rate of change in the relationship between the order of responding and creativity (Prediction 1).

The second predicted difference between consequences and alternative uses responding is in the dynamics of response times. There are two sub-predictions here. First, if it is the case that consequences responding is not a simple function of memory search (i.e., involves counterfactual reasoning, or some other process), the initial response time for a consequences prompt should be slower than the initial response time for alternative uses. Previous analyses suggest that on average people take between 2 and 4 s to generate their first use in an alternative uses task (Hass, 2017a). Theories of semantic memory search

suggest that this initial response latency is a function of the initial encoding of the cue, and the initialization of search processes (Wixted and Rohrer, 1994). If this initial encoding for consequences responding also involves counterfactual reasoning (or other processes), then the latency to the first response should be longer. Second, if it is the case that the consequences responding continually requires new creation of counterfactual consequences, the rate of responding should also be affected. As reviewed, Hass (2017a) showed that alternative uses responding is consistent with the negative exponential rate of search that is typical of semantic memory search. Explanations for the negative exponential rate usually center around the fact that semantic memory is a finite store and repeated search and recall of information will lead to a depletion of to-be-recalled information, exponentially slowing search. If it is the case that consequences responding does not simply involve search and retrieval from episodic and semantic stores, then a negative exponential function is not likely to fit response times. This also follows from the analysis by Xu (2017), which showed that when participants are constrained to only generating "new" alternative uses responses (i.e., avoiding the initial reliance on episodic stores), the cumulative response function appears more linear than exponential. More specifically, the rate of the cumulative response function is slower in the latter case. Since Madore et al. (2016) demonstrated that consequences response arrays are dominated by "new" responses, then consequences response curves should also be more linear than alternative uses response curves.

The final prediction tested in this analysis involves the semantic similarity of successive responses. In ordinary memory search using free-recall paradigms (e.g., naming all the animals one knows), participants often generate clusters of similar responses in short succession (e.g., farm animals such as cow, pig, goat, etc.). Several explanations for the phenomenon exist that are out of the scope of the current paper (cf. Gruenewald and Lockhead, 1980; Herrmann and Pearle, 1981; Troyer et al., 1997; Hills et al., 2012; Abbott et al., 2015), but generally pertain to the question of whether memory itself is a clustered representation, and/or whether search processes exploit certain features of the memory store. Hass (2017a) showed that clustering is not as readily apparent in Alternative Uses responding, though there was some relationship between inter-response time (IRT) and human rated similarity. However, since alternative uses responding relies to some extent on known associations, semantic similarity should be more strongly related to IRT in that task compared with the consequences task. This prediction is more tentative since it is plausible that both analyses show a weak relationship between IRT and semantic similarity, but the prediction is consistent with the prior research on the differential contributions of episodic memory to the two tasks.

The above logic is dependent upon the type of instructions used in the tasks. Two recent analyses showed that instructions to "be creative" while generating divergent thinking responses (as opposed to instructions to "think of as many responses as possible") leads to lower output totals (Nusbaum et al., 2014; Forthmann et al., 2016), but higher creativity ratings. The boost in creativity is moderated by fluid intelligence, with

more intelligent seemingly being better able to jump to more “creative” strategies (Nusbaum et al., 2014) throughout the task. In addition, the number of associations afforded by each DT prompt word (indexed by word-frequency) affected fluency, and to a lesser extent creativity and interacted with instruction type (Forthmann et al., 2016). So clearly, the type of instructions given to participants affects the kinds of memory processes in question here. In this paper, we opted to provide a middle ground between “be-creative” and “be-fluent” instructions because we used a 3-min time limit, but wanted to elicit an adequate number of responses per person for the purposes of evaluating the negative-exponential model of recall. This decision impacts the interpretation of our results and will be discussed later.

The novel components of this study include the various methods used to probe the predictions described above, which are not commonly applied to DT data. In addition, a web application was created to obtain the data. The app, which will be described in the section 2, relies on software created by the psiTurk project (McDonnell et al., 2012), a free and open-source set of python code that allows for experimental data to be collected in a controlled manner using participants recruited from MTurk. As will be described, the app and several helper functions are freely available to be adapted for use and can be downloaded from OSF and from the psiTurk experiment exchange (via github). The novelty of this component is that it allows researchers that may lack on-campus labs and participant pools to obtain reliable data regarding the cognitive processes involved in creative idea generation. The psiTurk code acts as an interface between user-generated HTML and JavaScript code and the MTurk platform, and several helpful features of that code enable controls on participants’ workflow. In addition, the psiTurk code allows for data management and storage without the usual databasing infrastructure overhead that is needed for other apps and web interfaces. In this way, the novelty of the web app pertains to its ability to provide tools to small labs and independent researchers that might not otherwise be available to them.

2. METHODS

2.1. Participants

Seventy-two participants (49 females) were recruited from MTurk. Participants were paid \$2 US for successful completion of the experiment (i.e., accepting the HIT on Mturk, and proceeding through the entire experiment). Ages ranged from 19 to 69 years ($M = 38.96$, $SD = 12.29$) and 79% of the participants were caucasian (8% African American, 4% Hispanic/Latino, 9% Other). All participants consented to participate electronically, and the experimental procedure was approved by the first author’s Institutional Review Board.

2.2. Materials

The experimental materials consisted of the experimental web-app, coded in JavaScript, the HTML pages that supported other parts of the experiment, and the supporting Python code that interfaced with MTurk. All are available via the psiTurk experiment exchange (<http://psiturk.org/ee/>

PaY8pUQUx2yd2wraXHEiLA). Information and tutorials about the process of creating an experiment using psiTurk are available at <http://psiturk.org>.

2.2.1. Physical Features of the psiTurk app

The web-app was written in JavaScript, and was laid out similar to a Matlab experiment used in previous studies (e.g., Hass, 2017a). Main instruction pages were presented to the participant along with specific instruction pages that preceded the two experimental blocks (one for alternative uses and the other for consequences), all with adequate font size. While responding, the cue was present on the screen in large font, and underneath the cue was a response field (an HTML text-entry field) labeled with the following text: “type responses here; press ENTER after EACH response.” Participants had full control of the response field with their keyboard and could use the backspace button to edit a response before pressing enter. When ENTER was pressed, the response field cleared so that the next response could be entered. When all tasks were completed, a survey page appeared with questions about age, sex, ethnicity, and a rating scale for engagement in the task (1 = not at all engaging; 10 = very engaging). A submit button appeared at the bottom of the survey page, which submitted the work to MTurk, and thanked the participant for participating.

In addition to collecting information about the type of browser the participant was using, when each browser event occurred (e.g., pressing a submit button), the main experimental data of interest were collected via the text-entry field. JavaScript functions were implemented to record the elapsed time between the presentation of the prompt and the first keypress for each response (response time), the latency between the first keypress of a response and the pressing of ENTER (entry time), and the actual text typed (response). Response time and the actual responses served as the primary data for analysis. Entry time was retained but not analyzed for this study.

2.2.2. Creativity and Similarity Ratings

In addition to the response times collected via the app, 3 independent sets of ratings were obtained for the responses participants entered. Two raters were recruited from MTurk following the procedure detailed by Hass et al. (2018). Raters were supplied with two 5-point semantic differential scales, one created for Alternative Uses responses, and the other created for Consequences responses. As described by Hass and colleagues, the wording of the semantic differentials was created to assess how creative the responses were vis a vis the process by which the responses were generated. Raters were supplied with spreadsheets, one per prompt, and assigned a rating to each unique response from each prompt. Inter-rater reliability was evaluated using the intra-class coefficient, with guidelines for interpretation supplied by Cicchetti (2001). The inter-rater reliability estimates for Alternative Uses prompts ranged from fair to good (Brick ICC(2,2) = 0.50, Hammer ICC(2,2) = 0.70, Car Tire ICC(2,2) = 0.49). Inter-rater reliability estimates from Consequences prompts were generally fair (No Gravity ICC(2,2) = 0.52, 12-Inches ICC(2,2) = 0.48, No Sleep ICC(2,2) = 0.49).

A separate set of two raters provided ratings of similarity on a 4-point semantic differential (Hass, 2017a). These raters were not recruited from MTurk, but were undergraduate research assistants at the first author's institution. The raters were supplied with spreadsheets that gave the order of response, the participant who generated the response, and a blank cell to indicate the similarity between each pair of successive responses per participant per prompt. These raters achieved good to excellent reliability (Brick ICC(2,2) = 0.77, Hammer ICC(2,2) = 0.77, Car Tire ICC(2,2) = 0.85; No Gravity ICC(2,2) = 0.68, 12-Inches ICC(2,2) = 0.81, No Sleep ICC(2,2) = 0.79).

2.3. Procedure

MTurk is a service where “human intelligence tasks” (HITs) are posted with descriptions and an offer of payment. The psiTurk command line interface allows for posting batches of HITs, which appear on MTurk as ads. When a participant clicks on a HIT, a brief description is presented. For this experiment, the description advertised that this was an experiment about creative thinking in which there were going to think of creative ideas for six different prompts. They were also told that the experiment should last about 30 min. Once a participant “accepted” the HIT (meaning that he or she intended to participate), he or she was allotted 60 min to actually complete the experiment. Generally, if MTurkers participants do not leave enough time to finish HITs, they are free to “release” them for another MTurker to accept. Sixty minutes was more than enough time for participants to accept and complete the hit, and only two people failed to submit their work within the 60 min time period, both of which waited too long between accepting the HIT and beginning the experiment.

Immediately upon beginning the HIT, a pop-up appeared in a participant's browser containing a consent form, which could also be printed. To give consent, participants simply clicked “I agree,” and the experiment was launched. Two general instruction pages were then loaded, the first screen explained that there would be 6 experimental trials lasting 3 min each, and a practice trial lasting 30 s. The second screen explained that the experiment required them to type on their keyboard, and that the experiment would be split into 3 blocks, the 30-s practice block, and two 9-min experimental blocks. They were told that they could take short breaks between blocks, but reminded that they must finish the HIT within the allotted time.

Each block, including the practice block, contained additional instructions specific to the task. For the practice block, instructions were provided about the experimental interface, that it would contain a cue and a response field where they were to continue to type responses until the cue changed. They were told that the practice block was simply designed to orient them to the use of the response field. The practice prompt was to type “all the colors [you] know.” Participants were reminded in the instructions, and on the text-entry page to type enter after each response, and to keep thinking of responses for the entire time. A START button was visible on the bottom of the instruction page, and clicking it began the practice trial.

At the end of the practice block, and each subsequent block, the prompt field was cleared from the screen and a

message appeared for 5 s, stating, “Good job! The next task is loading, please wait.” Another instruction page appeared for each experimental block, and participants were told that they could take a short break, but reminded that the HIT would expire in 60 min. The experiment did not proceed until the participant read the instructions for the block and clicked a START button on the bottom.

The order of the experimental blocks was counterbalanced: half of the participants began with the Alternative Uses prompts, and the other half began with the Consequences prompts. Within each block, the order of the prompts were randomized by JavaScript. The instructions for the Alternative Uses block read:

In the next set of tasks, the goal is to think of uses for objects. Please be as creative as you like. When you press Start, the name of a that object will appear on the screen. As soon as you think of something, type it into the field and press ENTER. Do this as many times as you can in 3 min. After 3 min on one category, the prompt will change to a new category, and after the next 3 min a third category will appear. This phase will last 9 min. Remember, it is important to try to keep thinking of responses and to type them in for the entire time for each prompt. Please type them in one at a time as they come to you, and press enter after entering each one.

The instructions for the Consequences block was similar, with the following change: participants were told that “a statement will appear on the screen. The statement might be something like *imagine that humans walked with their hands*. For 3 min, try to think of any and all consequences that might result from the statement. Please be as creative as you like.”

The prompts for the Alternative Uses task were brick, hammer, and car tire, and the prompts for the Consequences task were to imagine the consequences of “humans no longer needing sleep,” “humans becoming 12 inches tall,” and “gravity ceasing to exist.” On the text-entry page for Alternative Uses prompts, the text read “How can you use a(n) *OBJECT*?” to remind the participants that they must generate uses, not just associates for the object. On the text-entry page for Consequences prompts, the text read “What would happen if *SCENARIO*?”, again to remind them to generate consequences. In each case, the name of the object, or the scenario appeared in capital letters. Custom JavaScript functions recorded the response time (initial keypress), entry time (latency between initial keypress and pressing ENTER), and the actual text of each response. Data were saved to a dynamic MySQL instance hosted on Amazon Web Services, and parsed using a set of customized R functions which are downloadable here.

Prompts remained on the screen for 3 min, and were separated by a 5-s break, in which the prompt field cleared as well as the text-entry box, and a message stated “Good job! The next prompt is loading.” At the end of the final experimental block, the screen again displayed the “Good job” message, and the post-experiment questionnaire. Participants indicated their responses to questions about age, gender, ethnicity, and task engagement using drop-down menus. When they were finished, they pressed the submit button, and a thank-you message appeared on the

TABLE 1 | Descriptive statistics for fluency (number of responses) across prompts (AU, Alternative Uses; C, Consequences, see text for full description of prompts).

Prompt	Mean	SD	Median	Skew	Kurtosis
Brick (AU)	10.12	4.52	10	0.86	0.71
Hammer (AU)	9.52	4.36	9	0.73	−0.07
Car Tire (AU)	9.39	4.37	9	0.79	0.51
12-Inches (C)	8.42	3.60	8	0.50	−0.49
No Gravity (C)	7.91	3.83	7	1.00	1.27
No Sleep (C)	8.84	4.53	6	0.91	0.26

screen. They were then directed back to MTurk, and received payment when the batch of HITs was completed and approved. All participants that submitted results successfully back to MTurk were paid, regardless of whether they completed the task correctly. Inspection of the data revealed one instance of an error in the logging of responses and 4 instances of participants neglecting to press ENTER to log responses. Thus, the final sample size was 67 participants.

3. RESULTS

All data and analysis scripts and functions are available via the first author's Open Science Framework (osf.io/eux2k). Data parsing and analysis was performed using the R Statistical Programming Language (R Core Team, 2016), including the following packages: psych (Revelle, 2017), RMySQL (Ooms et al., 2017), jsonlite (Ooms, 2014), dplyr (Wickham et al., 2017), lme4 (Bates et al., 2015), lattice (Sarkar, 2008), and ggplot2 (Wickham, 2009). In all sections below, response times (RTs) represented the time between the presentation of the prompt and the time of the initial keypress leading to each response. This is consistent with recall studies that use voice-key technology to record response times, which are then defined according to the time of the initial voice onset of each response (e.g., Rohrer et al., 1995).

Four statistical analyses were planned: a descriptive analysis of the relationship between cumulative response counts and elapsed time (cumulative RT), a test of the difference in time to the first response (initial RT) across the two prompt types, a test of whether the relationship between pairwise similarity and inter-response time (IRT) differed by prompt-type, and a test of whether the serial order effect varied by prompt type. To aid interpretation of these analyses, descriptive statistics for fluency across the 6 prompts are listed in **Table 1**. Notably, fluency was, on average, significantly larger for Alternative Uses prompts ($M = 9.68$, $SD = 4.01$) than for Consequences prompts ($M = 8.39$, $SD = 3.49$), $t_{(66)} = 3.44$, $p = 0.00$, $d = 0.42$.

3.1. Cumulative Response Curves by Prompt

The purpose of this is to examine whether there are differences in the cumulative response function across tasks. In **Figure 1** plots of the average number of responses given by participants across successive 10-s blocks are shown. The plot is imprecise, such that toward the end of the interval, some of the slower

participants had generated few responses, which resulted in the fluctuations seen on the right hand side of the plot. However, the plot suggests that a negatively accelerating cumulative RT function should provide an adequate fit to individual data across the prompts.

The function that is often used to approximate the trends seen in **Figure 1** is an exponential function, in which the cumulative number of responses at time t is a curvilinear function that flattens out (reaches an asymptote) as time grows. The function was first derived by Bousfield and Sedgewick (1944) and is given by:

$$R(t) = a * (1 - e^{-\lambda t}) \quad (1)$$

where $R(t)$ is the cumulative number of responses at time t and e is the exponential function. The constant a represents the “asymptotic level of responding” or the total number of items available for retrieval (Bousfield and Sedgewick, 1944). The constant λ is the rate of the exponential decay (deceleration), and was parameterized in terms of the inverse relation $\lambda = \frac{1}{\tau}$. In this parameterization, τ is the theoretical mean response time, which is a more interpretable parameter than λ in this context. Though Wixted and Rohrer (1994) suggested that τ can provide an index of search set size, here, mathematically, a larger τ represents a more linear cumulative response function (Xu, 2017), which was of interest in this analysis. In both cases, a larger τ represents a smaller λ , and following Bousfield and Sedgewick (1944) the equation represents the proportion of to be retrieved items left to be sampled at time t (see also Gruenewald and Lockhead, 1980).

Following from earlier work (Hass, 2017a; Xu, 2017), nonlinear least squares estimates of the asymptote (a) and mean response time (τ) were obtained for each participant using the “nls” function in R. **Table 2** gives the quartiles of these estimates, for each prompt. Participants with fewer than 3 responses per prompt were excluded, but only for that prompt. In addition, as **Table 2** illustrates, a few additional participants' estimates were not returned due to failure of the nls algorithm to converge. The results in **Table 2** are consistent with **Figure 1**, in that the largest of the estimates of the τ parameter came from RTs for the Consequences of being 12 inches tall prompt. The results also illustrate that the exponential model predicts higher theoretical totals for fluency for the Consequences prompts than were actually observed (**Table 1**). These results are all consistent with Consequences prompts producing more linear cumulative response curves than Alternative Uses prompts.

A statistical test for the last assertion is difficult to perform because due to the nature of nonlinear least squares estimates. However, a statistical test of the difference among the various response curves is possible using the discretized data that were the basis for **Figure 1** in a mixed-effects regression model. The dependent variable in this model is cumulative responding with a discrete, integer predictor indexing which of the 18 10-s bins a response was output. A quadratic term for time-bin was added to the model to approximate the curvature of the exponential function. To test for the variation in curve shapes, the model included a fixed-effect of prompt (coded as

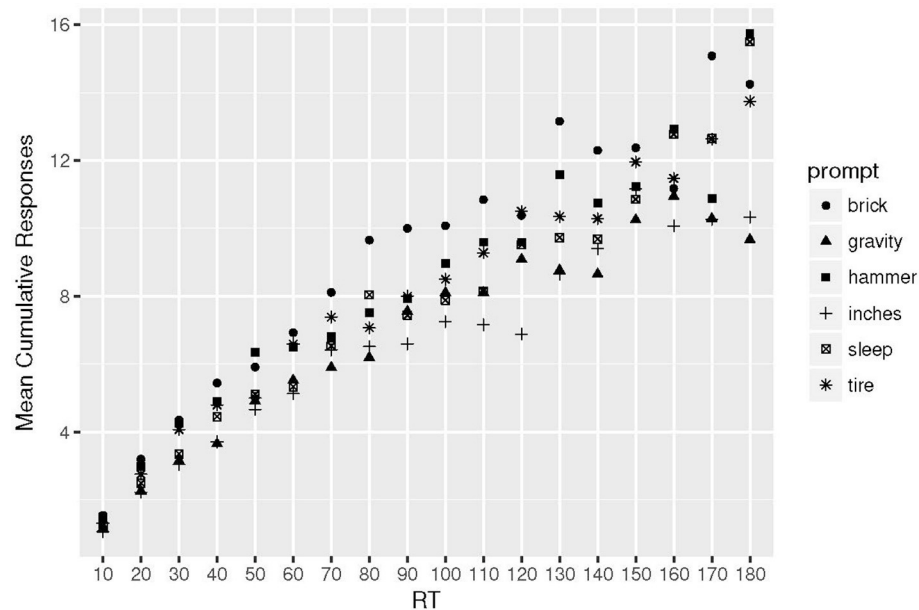


FIGURE 1 | Plot of the Mean number of cumulative responses in successive 10 s blocks. Alternative Uses prompts: brick, hammer, tire; Consequences prompts: gravity, inches, sleep (see section 2 for full description).

TABLE 2 | Median, Q1 and Q3 for the nonlinear least-squares estimates of asymptotic responding level (a), mean response time (τ) across prompts (AU, Alternative Uses; C, Consequences, see text for full description of prompts).

Prompt	n	a estimates			τ estimates		
		Q1	Median	Q3	Q1	Median	Q3
Brick (AU)	63	7.85	10.80	16.72	31.72	51.24	86.01
Hammer (AU)	59	7.97	10.80	16.20	33.63	64.05	107.98
Car Tire (AU)	59	8.17	11.24	18.50	38.26	65.53	121.26
12-Inches (C)	57	8.58	12.04	18.73	66.97	102.90	172.05
No Gravity (C)	59	7.40	10.03	14.81	41.97	77.68	143.89
No Sleep (C)	56	6.97	10.13	18.61	51.08	87.78	164.49

The scale of a is number of responses, whereas τ is reported in seconds.

a treatment contrast with the Brick task as the baseline) along with a cross-level interaction between the quadratic time-bin and prompt. Random intercepts and slopes per participant per prompt were also modeled. The numeric results are given in **Table 3**. Not surprisingly, the coefficients for linear and quadratic time-bin were significant, along with contrasts for output total. Importantly, the interactions between prompt and the quadratic time-bin term were significant for the No Gravity and 12-Inches prompt, with negative coefficients illustrating that these curves had less pronounced quadratic components (i.e., they were more linear) than the Brick curve. The curve for the No Sleep prompt did not significantly differ in its quadratic component. Thus, there is evidence that cumulative response times are more linear for 2 of the Consequences prompts compared to the Brick prompt.

3.2. First Response Latency by Condition

As an additional test of the processing differences between Alternative Uses and Consequences items, the RTs for first responses on the three Alternative Uses prompts were averaged, as were the RTs for the first responses to the 3 Consequences prompts. This seemed feasible given the results above, that all 6 prompts are well approximated by the exponential function, with varying parameters. The RT averages were skewed, due mainly to a few participants who took a long time to begin responding (which was later found to be a flaw in the design of the app). To test for a difference in initial RT across the tasks, without an assumption of normality, a Wilcoxon signed rank test (with continuity correction) was performed in R. The test was significant such that initial RTs were shorter for the Alternative Uses prompts than for the consequences prompts, $z = -2.16, p = 0.03, r = -0.26$. The effect size is small to medium using Cohen's (1988) guidelines for effect size r (Fritz et al., 2012), suggesting that there may be a small increase in initial processing involved when generating responses to Consequences prompts.

3.3. Pairwise Similarity by Prompt

The third planned analysis examined the relationship between pairwise similarity and IRT. Theoretically, if the Alternative Uses task involves searching through a memory store that is more highly clustered, there should be a stronger relationship between pairwise similarity and IRT for those prompts compared with Consequences prompts. That is, theoretically, short IRTs would indicate less remote association between successive responses. The consequences task, which may depend only on semantic memory, and also on other reasoning processes, should

TABLE 3 | Results of the Mixed-effects regression model of the RT curves, with cumulative response total as the dependent variable and 10-s block number as the discrete RT variable.

Fixed effects	Coefficient	<i>t</i>	<i>p</i>
Intercept	0.72		
RT (discrete)	1.04	49.55	< 0.001
RT-quadratic	−0.02	−12.48	< 0.001
No Gravity	−1.16	−4.58	< 0.001
Hammer	−0.62	−2.35	0.022
12-Inches	−1.46	−5.49	< 0.001
No Sleep	−1.37	−4.99	< 0.001
Car Tire	−0.91	−3.19	0.002
No Gravity*RT-quadratic	−0.008	−6.71	< 0.001
Hammer*RT-quadratic	−0.003	−2.69	0.007
12-Inches*RT-quadratic	−0.007	−6.41	< 0.001
No Sleep*RT-quadratic	−0.002	−1.88	0.060
Car Tire*RT-quadratic	−0.002	−1.71	0.087

Random effects	Variance
Participant	2.88
No Gravity	3.40
Hammer	3.81
12-Inches	3.86
No Sleep	4.16
Car Tire	4.47
Residual	2.20

The baseline level for the contrasts was the Brick prompt.

theoretically have a looser relationship between IRT and pairwise similarity. That hypothesis was tested by fitting a linear mixed effects model, with pairwise similarity rating as the dependent variable, IRT as a level-1 independent variable, and prompt-type (Uses v. Consequences) as a level-2 variable (fixed effect). Random intercepts for prompt (all 6 levels) and participant were included in the model to account for the repeated measures nature of the design. Modeling a cross-over interaction between prompt-type and IRT did not improve the fit of this model [$\chi^2_{(1)} = 0.20$, $p = 0.65$], meaning that there was no significant difference in the slope of the IRT - similarity relationship across the two prompt types.

Table 4 contains the full results of the model with no interaction term. The fixed effect of condition was not significant indicating a non-significant difference in average pairwise similarity across the two prompt types. However, the IRT - similarity association was significant, such that as IRTs increased, pairwise similarity tended to decrease. **Figure 2** illustrates these trends, and also shows that indeed, there seems to be little difference in the IRT-similarity slopes. However, the Figure also illustrates a clear nonlinear pattern in the results: short IRTs show a variety of different pairwise similarity values, but as IRTs increased, similarity decreases. Indeed, a quantile-quantile plot of residuals suggested that the model over-predicts pairwise similarity for short IRTs, and under-predicts pairwise similarity for long IRTs. So a more conservative conclusion is

TABLE 4 | Results of the Mixed-effects regression model with pairwise similarity as the dependent variable.

Fixed effects	Coefficient	<i>t</i>	<i>p</i>
Intercept	2.12	18.260	
IRT	−0.01	−8.50	< 0.001
Prompt-type	−0.20	−1.30	0.25

Random effects	Variance
Participant	0.038
Prompt	0.036
Residual	0.784

Consequences was the baseline Prompt-Type. The random effect of Prompt is the variance component across all 6 prompts.

that the relationship between IRT and pairwise similarity does not systematically vary by prompt-type, and that the linearity of the relationship may be overstated by the model. Contrary to the hypothesis, the two tasks seem to show the same degree of relationship between IRT and pairwise similarity.

3.4. Serial Order by Task

The final question asked in this analysis was whether the serial order effect varied by prompt type. Again, a mixed-effects model was fit, this time with creativity ratings as the dependent variable, the order of the response as the level-1 predictor, and a level-2 predictor for prompt-type (Alternative uses vs. Consequences). Response order was rescaled with the first response denoted by 0. To remove the potential of outliers (highly fluent individuals) to affect these results, serial order analysis was limited to the first 14 responses. This value was chosen because 95% of participants gave 14 or fewer responses on the consequences prompts. The 95th percentile of fluency for the Alternative uses prompts was around 17. To make this analyses equitable, the smaller of the two values was chosen.

Again, random intercepts for prompt-type and participant were included to model the repeated measures nature of the design. Following Beaty and Silvia (2012), both linear and quadratic order effects were modeled. An interaction between prompt type and the linear serial order term did not improve the model fit [$\chi^2_{(1)} = 1.04$, $p = 0.31$], nor did a quadratic serial order term improve the fit [$\chi^2_{(1)} = 0.86$, $p = 0.35$]. So the best model was that including a linear serial order term and a fixed-effect of prompt type, along with the random intercepts described above. The quantile-quantile plot of the residuals from this model suggested that the residuals did conform to normality, unlike the IRT model. **Table 5** contains the full output from the final model. There are significant linear and quadratic trends, which replicates the results of earlier serial order effects analyses (Beaty and Silvia, 2012). In addition, the ratings from Alternative Uses tasks were significantly lower at the onset of responding, but with no interaction, **Figure 3** illustrates that serial order effects are the same, albeit offset for the two prompt types. As such, it seems that participants begin with more creative responses to the Consequences prompts compared to the Alternative Uses

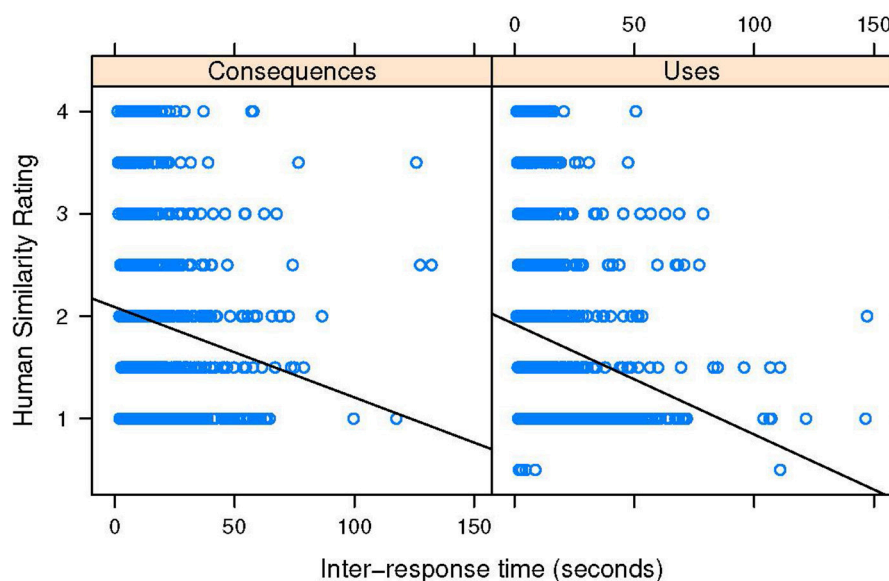


FIGURE 2 | Scatter plots of inter-response time and pairwise similarity for the two prompt conditions. Solid line represents ordinary least squares regression. See **Table 3** for the actual regression results from mixed-effects modeling.

TABLE 5 | Results of the Mixed-effects regression model of the serial order effect (Creativity as the dependent variable).

Fixed effects	Coefficient	<i>t</i>	<i>p</i>
Intercept	2.598		
Order (linear)	0.139	12.93	< 0.001
Order (quadratic)	−0.008	−8.94	< 0.001
Prompt-type	−0.554	−4.18	0.006
Random effects	Variance		
Participant	0.026		
Prompt	0.026		
Residual	0.456		

Consequences was the baseline Prompt-Type. The random effect of Prompt is the variance component across all 6 prompts.

prompts, but that the serial order effect remains in tact for both prompt types.

4. DISCUSSION

The present study was motivated by theoretical and practical issues. The theoretical motivations will be discussed first in light of the data. To address a hole in the burgeoning research on memory processes in creative thinking, the Alternative Uses task was compared to the Consequences task using a variety of metrics derived from existing analyses of memory retrieval. A recent analysis (Madore et al., 2016) suggested that the Consequences task may be less dependent on well-learned episodic information, and the results of this analysis are

consistent with that interpretation. First, in **Figure 1**, the rate of exponential growth of cumulative responses on the consequences prompts was, on average, slower. This can be seen for example, by examining the points where $t = 60$. There are clearly two clusters of points, the bottom of which consist of the mean cumulative number of responses for the 3 consequences prompts, which appear nearly 2 units lower than the three points representing the 3 alternative uses prompts. The separation between these points begins around 20 s, and is clear through about 70 s, where the mean cumulative responses become more variable. The individual fits of the exponential response time function in **Table 2** confirm along with the regression analysis that for at least 2 of the consequences prompts, output was more linear. This is consistent with Xu's results that when participants are instructed only to generate "new" Alternative Uses for objects in a creative task, the rate of exponential growth ($1/\tau$) of the response time function is slower, as it was here for Consequences responding. Xu also showed that when constraining participants to think of only "new" uses, their output totals are smaller, which is again consistent with the current analysis.

The slower rate of responding may be a function of additional processes operating during the Consequences task, such the initial time to respond to consequences prompts was significantly longer than the initial response time for alternative uses prompts. This suggests that either the encoding of the cue and initial search of memory takes longer for consequences prompts, or that in addition to encoding the cue and searching memory, consequences responding requires additional cognitive processing to continue. Unfortunately, the current analysis could not disentangle encoding from additional processes, but it is likely that future behavioral or neuroscientific studies will be able to do so. As mentioned, one candidate process involves

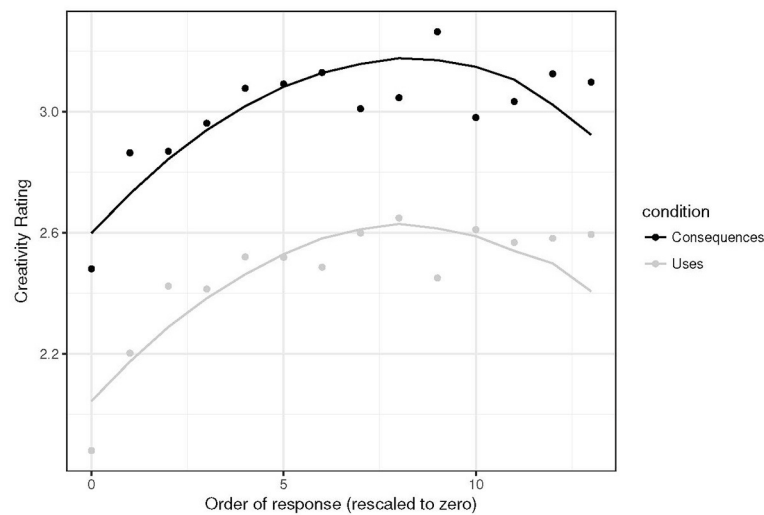


FIGURE 3 | Data (dots) and model predictions (line) for the serial order effects across the two prompt types. Note that the order of responses was re-scaled with 0 as the first response.

counterfactual reasoning about the impossible events represented by consequences prompts. However, it is also likely that the Consequences tasks are executively more demanding, and that it is an executive slow-down that is occurring, rather than a superposition of memory and reasoning processes.

Finally, the results of the serial order analysis provide further evidence that during consequences generation, participants are better able to generate more creative responses from the beginning of the response interval. However, there was still a serial order effect for Consequences prompts, meaning that remote association may form the core of Consequences generation, as it does for Alternative Uses generation. That interpretation is supported by the lack of a difference in the relationship between inter response time (IRT) and pairwise similarity across the two types of prompts. Though the IRT-similarity relationship does not seem to be linear, the amount of pairwise similarity did not vary significantly across the two types of prompts. This suggests that either the type of knowledge accessed during generation of both types of ideas is not likely to be strongly associated with other knowledge to the task, or that some executive process intervenes to override local cues during the generation of creative responses (cf. Troyer et al., 1997; Hills et al., 2012; Hass, 2017a). An answer to that question rests upon further analysis of the existence of semantic clusters of responses in these arrays, which is beyond the scope of the current study. Indeed, while norms exist to identify clusters in semantic categories such as animals (Troyer et al., 1997) there are currently no published norms for Alternative Uses responses, and norms for Consequences responses are proprietary. Though many researchers use their own systems for categorizing DT responses (for the purposes of flexibility scoring), a normative system for such categorization would be helpful to further probe the regularities of the search process involved by enabling more thorough computational modeling of idea generation.

4.1. Implications for Further Cognitive and Neuroscientific Studies

In the introduction, several pieces of new research pointing to a specific set of cortical structures within the default network supporting creative thinking were reviewed. Because this network shows reliable activation during tasks involving episodic retrieval and simulation (Gerlach et al., 2011), it has been hypothesized that activation of these regions in studies of creative thinking reflect the involvement of episodic retrieval mechanisms (Addis et al., 2016; Madore et al., 2017). Perhaps the clearest evidence for a role of episodic retrieval comes from Madore et al. (2017), who found that an episodic specificity induction boosted performance on the alternate uses task, which corresponded to increased activity within the left anterior hippocampus of the default network. Another recent study by Benedek et al. (2018) found that default network regions (hippocampus and medial prefrontal cortex) are involved in both the recall of original object uses and the imagination of novel object uses (i.e., the generation of “old” and “new” ideas, respectively; Benedek et al., 2014) compared to a control task that does not require creative thinking. Contrasting old and new idea generation directly, however, revealed selective engagement of the left supramarginal gyrus (SMG) during the generation of new ideas (Benedek et al., 2014, 2018). In light of the SMG’s role in cognitive control processes and constrained memory retrieval, Benedek and colleagues hypothesized that the generation of new ideas involves more executively-demanding mental simulations that are less relevant for the retrieval of old ideas from episodic memory. Critically, however, neuroimaging work has largely focused on the Alternate Uses task, so the extent to which similar brain regions are involved in Consequences generation remains an open question. Taken together with recent behavioral and neuroimaging work on old and new ideas, the current results suggest that, because Consequences responses tend to be more

“new” than “old,” one might expect executive brain regions to come online to support such complex search and retrieval processes. This is among the speculations relayed previously about the Consequences task, and the current results suggest that it may be advantageous to begin to compare the uses and consequences prompts in the scanner.

4.2. Practical Considerations for Using MTurk and Psiturk

Despite the success of this project, and the building of a useable interface to conduct these kinds of experiments using MTurk workers, there are a few practical issues to consider in follow-up studies. First, MTurkers are very sensitive to the directions. In pilot testing, participants tended to not press ENTER unless explicitly instructed to do so on the text-entry page, meaning data were lost. The current app includes instructions which are very specific and repeatedly remind the participant to press enter and to continue thinking of responses. Even so, at least 1 participant per prompt exhibited atypical initial response times (e.g., initial RT > 30 s), which may be due to distraction. Though the age range of participants on MTurk is larger than that for normal laboratory based psychology experiments, using MTurk successfully, and getting work approved usually requires that people are computer savvy. That is, the small number of long latencies is not expected to be a function of the age of the participants. Even though one participant reported being 69 year old, a majority of participants were between the ages of 24 and 50 year old. The relationship between age and initial latency was not tested, however, and may be a relevant research question for future studies.

The app, as it is now constructed, does not allow the participant to take an extended break within an experimental block, only between blocks. Within a block, the prompt would change after 3 min plus a 5 s delay. If the participant became distracted, there was no way for him or her to notice the fact that the next task started, and latencies were biased by the distraction. Again, this was rare, but the fact that it happened more than once means that initial steps must be taken to control the flow of the program, or to set exclusion criteria. Since exclusion criteria set prior to the experiment were simply designed to filter out participants who did not follow directions or who did not respond to all tasks, it was decided that these atypical latencies should be retained for transparency. Due to the nature of the analysis, these latencies did not greatly affect the results, but the app has been updated to include a button press (space-bar) between each prompt presentation, so that MTurkers can move at their own pace.

4.3. Theoretical Limitations and Alternative Explanations

Aside from practical considerations, a few limitations and alternative explanations exist. First, the residuals of the linear mixed-effects regression of similarity on IRT and prompt type violated the assumption of normal residuals, and the model may be overshooting the relationship between IRT and pairwise similarity. As mentioned, the lack of norms for responses on

both kinds of prompts used in this study makes other IRT analyses difficult, but such analyses are necessary before firmer conclusions are made about the IRT-similarity relationship during creative thinking tasks. As an alternative, ordinal multilevel regression models could be used, as the similarity scale can be treated as ordinal. For example, Forthmann et al. (2016) used linear response trees to examine the interaction of word frequency and be-creative instructions with very interpretable results.

In addition, there may be an alternative explanation for the difference in creativity ratings between Alternative Uses and Consequences responses. The rating scales used to rate the responses do differ with respect to the scoring criteria in a nature relevant to this difference in response length. For consequences responses, the maximum creativity rating for consequences responses (5 out of 5) is “very imaginative/detailed consequence”, while the maximum creativity rating for uses is “very imaginative / re-contextualized use.” The rationale for the difference between the two is that the scores are then specific to the goals of the tasks. In constructing those scales, it was reasoned that a very creative consequence should be one in which a detailed thought process was carried out. However, Consequences responses may simply earn higher ratings because they contain a greater number of words on average (Forthmann et al., 2017). This issue of scoring differences across creative thinking tasks is at the heart of a larger debate in creativity about domain-specificity (Baer, 2011). That said, the participants were not aware of the criteria used for rating their responses, and are not explicitly told to be detailed in the responses to either task. So from that perspective, the fact that Consequences responses tend to be longer seems to be related to the nature of the prompt rather than an artifact of the rating procedure. Still more research into the reasoning processes that underpin the Consequences task is necessary to shed more light on this issue.

Finally, the instructions given to participants deviated from the “be creative” instructions used in more recent studies. The choice was made to use the current instructions in order to facilitate higher fluency totals. This, however, is a limitation of the method, as it can be argued that instructing participants to “be creative as [they] like[d]” leaves open the question as to whether all participants interpreted the tasks in the same way. This is an important caveat, though the results are in line with predictions based on studies that used “be creative” instructions. It would be advantageous, however, to investigate whether the exponential parameters fit to RT data in the current study would change when “be creative” instructions are used, compared to “be fluent” instructions.

4.4. Concluding Remarks

The goals of this study were both practical and theoretical. On the practical side, a workable web interface for collecting creative thinking data from MTurk workers is now available to the scientific research community. Moreover, the data generated by participants in the web environment are consistent with data generated in the lab (cf. Hass, 2015, 2017a), and can be used

to test cognitive hypotheses. This is a novel development as it can allow for researchers with limited lab space and lack of participant pools to collect valuable data about cognitive processing in divergent thinking. In addition, since MTurkers represent a more wide-ranging demographic than undergraduate participant pools, the results may be more externally valid.

On the theoretical side, the results suggest that both Alternative Uses and Consequences tasks tap the same general processes, and conform to the serial order effect. This is a novel result as the serial order effect has never been explored with Consequences prompts. However, there were subtle differences such that the initial processing time for Consequences responding is slightly longer than Alternative Uses, and Consequences responses seem to earn higher creativity ratings from the start of the idea generation process. The latter effect may be due to additional reasoning required to either search for or evaluate potential Consequences before they are output. Evidence from cumulative RT analysis provides some support for that assertion, but it is hoped that future research with computational models and brain imaging techniques will provide more insight. At the same time, both Alternative Uses and Consequences tasks can continue to be used as measures of divergent thinking.

REFERENCES

- Abbott, J. T., Austerweil, J. L., and Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychol. Rev.* 122, 558–569. doi: 10.1037/a0038693
- Abraham, A., and Bubic, A. (2015). Semantic memory as the root of imagination. *Front. Psychol.* 6:325. doi: 10.3389/fpsyg.2015.00325
- Acar, S., and Runco, M. A. (2017). Latency predicts category switch in divergent thinking. *Psychol. Aesthet. Creat. Arts* 11, 43–51. doi: 10.1037/aca0000091
- Addis, D. R., Pan, L., Musicaro, R., and Schacter, D. L. (2016). Divergent thinking and constructing episodic simulations. *Memory* 24, 89–97. doi: 10.1080/09658211.2014.985591
- Avitia, M. J., and Kaufman, J. C. (2014). Beyond g and c: the relationship of rated creativity to long-term storage and retrieval (Glr). *Psychol. Aesthet. Creat. Arts* 8, 293–302. doi: 10.1037/a0036772
- Baer, J. (2011). Why grand theories of creativity distort distract and disappoint. *Int. J. Creat. Prob. Solving* 21, 73–100. Available online at: <http://psycnet.apa.org/record/2011-10112-006>
- Barr, N., Pennycook, G., Stolz, J. A., and Fugelsang, J. A. (2015). Reasoned connections: a dual-process perspective on creative thought. *Think. Reason.* 21, 61–75. doi: 10.1080/13546783.2014.895915
- 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
- Beaty, R. E., Benedek, M., Silvia, P. J., and Schacter, D. L. (2016). Creative cognition and brain network dynamics. *Trends Cogn. Sci.* 20, 87–95. doi: 10.1016/j.tics.2015.10.004
- Beaty, R. E., Christensen, A. P., Benedek, M., Silvia, P. J., and Schacter, D. L. (2017a). Creative constraints: brain activity and network dynamics underlying semantic interference during idea production. *NeuroImage* 148, 189–196. doi: 10.1016/j.neuroimage.2017.01.012
- Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., et al. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1087–1092. doi: 10.1073/pnas.1713532115
- Beaty, R. E., and Schacter, D. L. (2018). “Episodic memory and cognitive control: Contributions to creative idea production,” in *The Cambridge Handbook of the*

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of APA ethics guidelines for ethical treatment of participants. The protocol was approved by the Institutional Review Board, Thomas Jefferson University. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed for this study can be found on the Open Science Framework repository for **this study**.

AUTHOR CONTRIBUTIONS

RH conceived of the study and ran the analyses. RH and RB wrote the paper.

ACKNOWLEDGMENTS

The authors thank Kelsey Korst and Marisa Rivera for help with response coding.

Neuroscience of Creativity, eds R. E. Jung and O. Vartanian (New York, NY: Cambridge University Press), 249–260.

- Beaty, R. E., and Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychol. Aesthet. Creat. Arts* 6, 309–319. doi: 10.1037/a0029171
- Beaty, R. E., Silvia, P. J., and Benedek, M. (2017b). Brain networks underlying novel metaphor production. *Brain Cogn.* 111, 163–170. doi: 10.1016/j.bandc.2016.12.004
- Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., and Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Mem. Cogn.* 42, 1186–1197. doi: 10.3758/s13421-014-0428-8
- Bendetowicz, D., Urbanski, M., Garcin, B., Foulon, C., Levy, R., Bréchemier, M.-L., et al. (2018). Two critical brain networks for generation and combination of remote associations. *Brain* 141, 217–233. doi: 10.1093/brain/awx294
- Benedek, M., Jauk, E., Fink, A., Koschutnig, K., Reishofer, G., Ebner, F., et al. (2014). To create or to recall? Neural mechanisms underlying the generation of creative new ideas. *NeuroImage* 88, 125–133. doi: 10.1016/j.neuroimage.2013.11.021
- Benedek, M., Kenett, Y. N., Umdasch, K., Anaki, D., Faust, M., and Neubauer, A. C. (2017). How semantic memory structure and intelligence contribute to creative thought: a network science approach. *Think. Reason.* 23, 158–183. doi: 10.1080/13546783.2016.1278034
- Benedek, M., Könen, T., and Neubauer, A. C. (2012). Associative abilities underlying creativity. *Psychol. Aesthet. Creat. Arts* 6, 273–281. doi: 10.1037/a0027059
- Benedek, M., Schües, T., Beaty, R. E., Jauk, E., Koschutnig, K., Fink, A., et al. (2018). To create or to recall original ideas: brain processes associated with the imagination of novel object uses. *Cortex* 99, 93–102. doi: 10.1016/j.cortex.2017.10.024
- Bousfield, W. A., and Sedgewick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *J. Gen. Psychol.* 30, 149–165. doi: 10.1080/00221309.1944.10544467
- Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011
- Byrne, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends Cogn. Sci.* 6, 426–431. doi: 10.1016/S1364-6613(02)01974-5

- Chen, Q., Beaty, R. E., Wei, D., Yang, J., Sun, J., Liu, W., et al. (2018). Longitudinal alterations of frontoparietal and frontotemporal networks predict future creative cognitive ability. *Cereb. Cortex* 28, 103–115. doi: 10.1093/cercor/bhw353
- Christensen, P. R., Guilford, J. P., and Wilson, R. C. (1957). Relations of creative responses to working time and instructions. *J. Exp. Psychol.* 53, 82–88. doi: 10.1037/h0045461
- Cicchetti, D. (2001). Methodological commentary the precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. *J. Clin. Exp. Neuropsychol.* 23, 695–700. doi: 10.1076/jcen.23.5.695.1249
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Dixon, J. E., and Byrne, R. M. J. (2011). “If only” counterfactual thoughts about exceptional actions. *Mem. Cogn.* 39, 1317–1331. doi: 10.3758/s13421-011-0101-4
- Forthmann, B., Gerwig, A., Holling, H., Çelik, P., Storme, M., and Lubart, T. (2016). The be-creative effect in divergent thinking: the interplay of instruction and object frequency. *Intelligence* 57, 25–32. doi: 10.1016/j.intell.2016.03.005
- Forthmann, B., Holling, H., Zandi, N., Gerwig, A., Çelik, P., Storme, M., et al. (2017). Missing creativity: the effect of cognitive workload on rater (dis-) agreement in subjective divergent-thinking scores. *Think. Skills Creat.* 23, 129–139. doi: 10.1016/j.tsc.2016.12.005
- Forthmann, B., Lips, C., Szardenings, C., Scharfen, J., and Holling, H. (2018). Are speedy brains needed when divergent thinking is speeded-or unspeeded? *J. Creat. Behav.* 3, 53–11. doi: 10.1002/jocb.350
- Fritz, C. O., Morris, P. E., and Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *J. Exp. Psychol.* 141, 2–18. doi: 10.1037/a0024338
- Gao, Z., Zhang, D., Liang, A., Liang, B., Wang, Z., Cai, Y., et al. (2017). Exploring the associations between intrinsic brain connectivity and creative ability using functional connectivity strength and connectome analysis. *Brain Connect.* 7, 590–601. doi: 10.1089/brain.2017.0510
- Gerlach, K. D., Spreng, R. N., Gilmore, A. W., and Schacter, D. L. (2011). Solving future problems: default network and executive activity associated with goal-directed mental simulations. *NeuroImage* 55, 1816–1824. doi: 10.1016/j.neuroimage.2011.01.030
- Gilhooly, K. J., Fioratou, E., Anthony, S. H., and Wynn, V. (2007). Divergent thinking: strategies and executive involvement in generating novel uses for familiar objects. *Brit. J. Psychol.* 98, 611–625. doi: 10.1111/j.2044-8295.2007.tb00467.x
- Green, A. E., Cohen, M. S., Raab, H. A., Yedibalian, C. G., and Gray, J. R. (2015). Frontopolar activity and connectivity support dynamic conscious augmentation of creative state. *Hum. Brain Mapp.* 36, 923–934. doi: 10.1002/hbm.22676
- Gruenewald, P. J., and Lockhead, G. R. (1980). The free recall of category examples. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 225–240. doi: 10.1037/0278-7393.6.3.225
- Hass, R. W. (2015). Feasibility of online divergent thinking assessment. *Comput. Hum. Behav.* 46, 85–93. doi: 10.1016/j.chb.2014.12.056
- Hass, R. W. (2017a). Semantic search during divergent thinking. *Cognition* 166, 344–357. doi: 10.1016/j.cognition.2017.05.039
- Hass, R. W. (2017b). Tracking the dynamics of divergent thinking via semantic distance: analytic methods and theoretical implications. *Mem. Cogn.* 45, 233–244. doi: 10.3758/s13421-016-0659-y
- Hass, R. W., Rivera, M., and Silvia, P. (2018). On the dependability and feasibility of layperson ratings of divergent thinking. *Front. Psychol.* 9:1343. doi: 10.3389/fpsyg.2018.01343
- Herrmann, D. J., and Pearle, P. M. (1981). The proper role of clusters in mathematical models of continuous recall. *J. Math. Psychol.* 24, 139–162. doi: 10.1016/0022-2496(81)90040-7
- Hills, T. T., Jones, M. N., and Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychol. Rev.* 119, 431–440. doi: 10.1037/a0027373
- Kenett, Y. N., Anaki, D., and Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Front. Hum. Neurosci.* 8:407. doi: 10.3389/fnhum.2014.00407
- Kenett, Y. N., Beaty, R. E., Silvia, P. J., Anaki, D., and Faust, M. (2016). Structure and flexibility: investigating the relation between the structure of the mental lexicon, fluid intelligence, and creative achievement. *Psychol. Aesthet. Creat. Arts* 10, 377–388. doi: 10.1037/aca0000056
- Madore, K. P., Addis, D. R., and Schacter, D. L. (2015). Creativity and memory: effects of an episodic-specificity induction on divergent thinking. *Psychol. Sci.* 26, 1461–1468. doi: 10.1177/0956797615591863
- Madore, K. P., Jing, H. G., and Schacter, D. L. (2016). Divergent creative thinking in young and older adults: extending the effects of an episodic specificity induction. *Mem. Cogn.* 44, 974–988. doi: 10.3758/s13421-016-0605-z
- Madore, K. P., Thakral, P. P., Beaty, R. E., Addis, D. R., and Schacter, D. L. (2017). Neural mechanisms of episodic retrieval support divergent creative thinking. *Cereb. Cortex*. doi: 10.1093/cercor/bhx312. [Epub ahead of print].
- McDonnell, J. V., Martin, J. B., Markant, D. B., Coenen, A., Rich, A. S., and Gureckis, T. M. (2012). psiturk (version 1.02). Available online at: <https://github.com/NYUCCL/psiTurk>
- Nusbaum, E. C., Silvia, P. J., and Beaty, R. E. (2014). Ready, set, create: what instructing people to “be creative” reveals about the meaning and mechanisms of divergent thinking. *Psychol. Aesthet. Creat. Arts* 8, 423–432. doi: 10.1037/a0036549
- Ooms, J. (2014). The jsonlite package: a practical and consistent mapping between json data and r objects. arXiv:1403.2805 [stat.CO].
- Ooms, J., James, D., DebRoy, S., Wickham, H., and Horner, J. (2017). *RMySQL: Database Interface and 'MySQL' Driver for R*. R package version 0.10.13.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Raaijmakers, J. G., and Shiffrin, R. M. (1981). Search of associative memory. *Psychol. Rev.* 88, 93–134. doi: 10.1037/0033-295X.88.2.93
- Revelle, W. (2017). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, IL: Northwestern University. R package version 1.7.8.
- Rohrer, D., Wixted, J. T., Salmon, D. P., and Butters, N. (1995). Retrieval from semantic memory and its implications for Alzheimer's disease. *J. Exp. Psychol.* 21, 1127–1139. doi: 10.1037/0278-7393.21.5.1127
- Runco, M. A., Abdulla, A. M., Paek, S. H., Al-Jasim, F. A., and Alsuwaidi, H. N. (2016). Which test of divergent thinking is best. *Creativity* 3, 4–18. doi: 10.1515/ctra-2016-0001
- Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. New York, NY: Springer.
- Schacter, D. L., and Madore, K. P. (2016). Remembering the past and imagining the future: identifying and enhancing the contribution of episodic memory. *Mem. Stud.* 9, 245–255. doi: 10.1177/1750698016645230
- Shi, L., Sun, J., Ren, Z., Chen, Q., Wei, D., Yang, W., et al. (2018). Large-scale brain network connectivity underlying creativity in resting-state and task fMRI: cooperation between default network and frontal-parietal network. *Biol. Psychol.* 135, 102–111. doi: 10.1016/j.biopsycho.2018.03.005
- Silvia, P. J. (2011). Subjective scoring of divergent thinking: examining the reliability of unusual uses, instances, and consequences tasks. *Think. Skills Creat.* 6, 24–30. doi: 10.1016/j.tsc.2010.06.001
- Silvia, P. J., Beaty, R. E., and Nusbaum, E. C. (2013). Verbal fluency and creativity: general and specific contributions of broad retrieval ability (Gr) factors to divergent thinking. *Intelligence* 41, 328–340. doi: 10.1016/j.intell.2013.05.004
- Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., et al. (2008). Assessing creativity with divergent thinking tasks: exploring the reliability and validity of new subjective scoring methods. *Psychol. Aesthet. Creat. Arts* 2, 68–85. doi: 10.1037/1931-3896.2.2.68
- Sowden, P. T., Pringle, A., and Gabora, L. (2015). The shifting sands of creative thinking: connections to dual-process theory. *Think. Reason.* 21, 40–60. doi: 10.1080/13546783.2014.885464
- Sun, J., Liu, Z., Rolls, E. T., Chen, Q., Yao, Y., Yang, W., et al. (2018). Verbal creativity correlates with the temporal variability of brain networks during the resting state. *Cereb. Cortex*. doi: 10.1093/cercor/bhy010. [Epub ahead of print].
- Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking: Norms-Technical Manual Research Edition*. Princeton, NJ: Personell Press.
- Troyer, A. K., Moscovitch, M., and Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology* 11, 138–146. doi: 10.1037/0894-4105.11.1.138

- Vartanian, O., Beatty, E. L., Smith, I., Blackler, K., Lam, Q., and Forbes, S. (2018). One-way traffic: the inferior frontal gyrus controls brain activation in the middle temporal gyrus and inferior parietal lobule during divergent thinking. *Neuropsychologia* 118, 68–78 doi: 10.1016/j.neuropsychologia.2018.02.024
- Wang, M., Hao, N., Ku, Y., Grabner, R. H., and Fink, A. (2017). Neural correlates of serial order effect in verbal divergent thinking. *Neuropsychologia* 99, 92–100. doi: 10.1016/j.neuropsychologia.2017.03.001
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.
- Wickham, H., Francois, R., Henry, L., and Muller, K. (2017). *dplyr: A Grammar of Data Manipulation*. R package version 0.7.4.
- Wilson, R. C., Guilford, J. P., Christensen, P. R., and Lewis, D. J. (1954). A factor-analytic study of creative-thinking abilities. *Psychometrika* 19, 297–311. doi: 10.1007/BF02289230
- Wixted, J. T., and Rohrer, D. (1994). Analyzing the dynamics of free recall: an integrative review of the empirical literature. *Psychon. Bull. Rev.* 1, 89–106. doi: 10.3758/BF03200763
- Xu, D. (2017). *Oppositional Processes in Divergent Thinking*. PhD thesis, Arizona State University.
- Zhu, W., Chen, Q., Xia, L., Beaty, R. E., Yang, W., Tian, F., et al. (2017). Common and distinct brain networks underlying verbal and visual creativity. *Hum. Brain Mapp.* 38, 2094–2111. doi: 10.1002/hbm.23507

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 Hass and Beaty. 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.



Exploring the Creative Process: Integrating Psychometric and Eye-Tracking Approaches

Dorota M. Jankowska¹, Marta Czerwonka¹, Izabela Lebuda² and Maciej Karwowski^{2*}

¹ Department of Educational Sciences, The Maria Grzegorzewska University, Warsaw, Poland, ² Institute of Psychology, University of Wrocław, Wrocław, Poland

OPEN ACCESS

Edited by:

Ian Hocking,
Canterbury Christ Church University,
United Kingdom

Reviewed by:

Emma Threadgold,
University of Central Lancashire,
United Kingdom
Jonathan Plucker,
Johns Hopkins University,
United States

*Correspondence:

Maciej Karwowski
maciej.karwowski@uwr.edu.pl;
maciek.karwowski@gmail.com

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 21 May 2018

Accepted: 19 September 2018

Published: 09 October 2018

Citation:

Jankowska DM, Czerwonka M,
Lebuda I and Karwowski M (2018)
Exploring the Creative Process:
Integrating Psychometric
and Eye-Tracking Approaches.
Front. Psychol. 9:1931.
doi: 10.3389/fpsyg.2018.01931

This exploratory study aims at integrating the psychometric approach to studying creativity with an eye-tracking methodology and thinking-aloud protocols to potentially untangle the nuances of the creative process. Wearing eye-tracking glasses, one hundred adults solved a drawing creativity test – The Test of Creative Thinking-Drawing Production (TCT-DP) – and provided spontaneous comments during this process. Indices of visual activity collected during the eye-tracking phase explained a substantial amount of variance in psychometric scores obtained in the test. More importantly, however, clear signs of methodological synergy were observed when all three sources (psychometrics, eye-tracking, and coded thinking-aloud statements) were integrated. The findings illustrate benefits of using a blended methodology for a more insightful analysis of creative processes, including creative learning and creative problem-solving.

Keywords: psychometrics, creative process, Test of Creative Thinking-Drawing Production, eye-tracking, thinking aloud

INTRODUCTION

While scholars generally agree that creativity leads to ideas and products that are novel (original) and meaningful (useful, relevant) (see Runco and Jaeger, 2012), much less agreement is observed when it comes to the creative process. For good reasons, though: dynamism of the process and variety of mechanisms involved in generation and explorations of ideas make it challenging to capture.

Diverse conceptualizations of how, when, and why people create have resulted in a set of quite isolated measurement approaches that, taken together, make effective synthesis of previous findings difficult. Are there any common findings that may be considered as regularities of the creative process, despite the methods applied? Or perhaps different measurement approaches, by definition, can capture only some aspects of the creative process?

In this article, we analyze different perspectives and include a variety of methods by integrating the more traditional psychometric approach (usually based on scores obtained in creative thinking tests) with an analysis of metacognitive and self-regulation mechanisms engaged within the process (measured by think-aloud protocols), and the parameters that reveal the way attention functions during this process (measured by the eye-tracking methodology). We posit that such triangulation

holds the promise to result in a more complex and comprehensive look at the process itself. The approaches we use in the study described below not only give us an opportunity to measure real-time (Schwarz, 2012) dynamics of the process, but also potentially catch the interaction between the person (or actor; see Glăveanu, 2013) and the outcome (or product; see Botella et al., 2013), and to include metacognitive aspects of the process into our analyses.

There are different theoretical and methodological views on the nature of the creative process and its measurement. Below, we briefly focus on three of them, which we consider most relevant from the perspective of our investigation. The first, most classic perspective divides the creative process into a number of different, sequential or recursive stages, or phases (e.g., Wallas, 1926). More contemporary extensions describe the process in terms of the most important mental operations and behaviors within each of the stages (Isaksen et al., 2000; Sawyer, 2012). The role of cognitive processes (e.g., Mumford et al., 1991, 1997) or different facets (Amabile, 1996) during the process of creating or problem solving is analyzed as well. The classic stage models often utilize a wide range of methods, starting from qualitative interviews (Perry, 1999) or observations (Csikszentmihalyi, 1996; Dunbar, 1997), all the way to historical case studies of eminent creators (Weisberg, 1986) or computational models based on the archival study of individual creative episodes taken from the notebooks of scientists (e.g., Langley et al., 1987; Kulkarni and Simon, 1988).

The second view on the creative process, i.e., the creative cognition approach (Finke et al., 1992), emphasizes and micro-analytically investigates the cognitive mechanisms as the core of the creative thought. For example, the geneplore model (Finke et al., 1992) specifies the creative process as a set of basic cognitive processes that increase the likelihood of a creative output. The nature of both generative and exploratory processes – two main phases in the geneplore model – can be described and potentially modeled thanks to the understanding of detailed operations and processes engaged in both these phases. The creative cognition approach largely benefited from a convergence strategy (Ward, 2001): anecdotal facts about great creative discoveries served as an inspiration that allowed to hypothesize specific phases and mental operations involved in creative thinking that were subsequently rigorously examined in controlled laboratory experiments. Studies inspired by the creative cognition approach usually focus on intensive, experimental laboratory task – the so-called creative generation tasks. Participants are presented with open-ended problems (e.g., drawing animals that may exist on other planets) and their solutions are scored in terms of creativity or originality. Importantly, the creative cognition approach is focused on both the outcome and the process itself.

The third approach refers to psychometrics that has for decades been the predominant approach to understanding individual differences in creativity and, to a lesser extent, the creative process. Not surprisingly, psychometricians tended to rely on divergent thinking tests or other standardized methods, such as scales or questionnaires (see Hocevar, 1981). Although the expansion of purely psychometric work has been criticized as providing, at best, a fragmented and incomplete (see e.g.,

Baer, 1994; Glăveanu, 2014) and, at worst, invalid picture of creativity, usefulness of this approach goes without saying (see Plucker and Runco, 1998 for a discussion). Psychometric works identified several important effects in the psychology of creativity, such as the serial-order effect (Beatty and Silvia, 2012), the threshold hypothesis (Jauk et al., 2013; Karwowski et al., 2016a), or the fourth-grade slump in creativity (Krampen, 2012). What's more, current works that utilize psychometric tasks tend not only to explore the overall creativity of the outcomes created (e.g., responses in the unusual uses tasks), but also the dynamics of the process itself (Gilhooly et al., 2007; Beatty and Silvia, 2012; Silvia et al., 2017).

Exploring the Dynamics of the Creative Process

The creative process has been examined with the use of both qualitative and quantitative methods that permit differentiating the stages of the creative process, e.g., among professional artists (Botella et al., 2013) or screenplay writers (Bourgeois-Bougrine et al., 2014). Moreover, contemporary studies have demonstrated an analysis of temporal dynamics in creative ideation (Gilhooly et al., 2007; Beatty and Silvia, 2012; Silvia et al., 2017) – not necessarily restricted to individuals, but also in dyads (Glăveanu et al., 2018). They also applied visual-verbal protocols to explore the meshed modes of creative thinking in time (Pringle and Sowden, 2017). Likewise, novel techniques and measures that allow for better understanding of the creative process are being developed. Conceptual Clock-face in testing the role of distance in conceptual processing (Hocking and Vernon, 2017) and Mode Shifting Index to assess shifts between associative and analytic modes of creative thought (Pringle and Sowden, 2016) are but two examples. In the same vein, data analysis methods are becoming more sophisticated – researchers not only combine qualitative and quantitative strategies (Bourgeois-Bougrine et al., 2014), but they also routinely use multivariate analyses (with semantic component analysis; Botella et al., 2013) and multilevel modeling (Glăveanu et al., 2018). The natural advantage of more synergistic approaches lies in the potential to combine a wider set of data, allowing for a more complex understanding of the creative process.

Thus, we observe a growing number of studies that focus on developing or adapting methods, and it seems that creativity scholars strive to combine different perspectives and use more blended approach. Does it mean that psychometric measures of creative abilities and creative process will eventually be discarded? We doubt it. Instead, we suggest putting more effort into *dynamizing creativity tests* that would allow for a more fruitful analysis of the process rather than the output alone. Such a blended, multi-method approach was effectively used to analyze the dynamic nature of the creative learning process (e.g., Gajda et al., 2017). Gajda et al. (2017) combined the qualitative (observations, audio-recorded interactions) and quantitative (measures of creativity and academic achievement) method to explore interpersonal characteristics of creativity in a classroom. Here, we take a more intrapersonal orientation by exploring

the role of strategies and self-regulation during a creativity test.

How Do Methodological Innovations Inform Our Understanding of the Process?

Previous findings showed that individuals implement various strategies – from the more structured (Sawyer, 2012) to more isolated (Root-Bernstein and Root-Bernstein, 2001) – during the creative process. Other classifications defined strategies as experiential – derived from episodic personal memory – or semantic, thus based on abstract, conceptual knowledge (Walker and Kintsch, 1985; Vallee-Tourangeau et al., 1998). Gilhooly et al. (2007) classified the strategies obtained using think-aloud protocols into Memory, Property, Broad Use and Disassembly strategies – and demonstrated that different strategies operate at the initial and later creative process stages.

Analysis of strategies involved in the performance phase (e.g., critical thinking, ideation, imagery) seems to be specifically relevant to their relationship with self-regulated learning (Rubenstein et al., 2017). According to this approach, creative process strategies may support learning strategies, because of their relevance for self-regulation. Effective self-regulation processes are crucial for successful transformation of creative ideas into creative products (Ivcevic and Nusbaum, 2017). Two broad groups of self-regulation processes in creativity have been identified in previous studies: [1] revising and re-strategizing, and [2] sustaining and maintaining effort. The first set involves continual exploration and revision (Csikszentmihalyi and Getzels, 1971; Csikszentmihalyi, 1988). The second set involves both planning and implementation operations. While appreciating this broad categorization, we posit that even a more detailed analysis of strategies may be necessary to describe different idiosyncrasies of the creative process. On the broad level, it seems that strategies focused on generation and exploration may be important during the initial phases of the creative process, while those related to monitoring and control activities are engaged more steadily across the entire process – meaning not only generation of initial ideas, but also their combinations and polishing. In the literature, theoretical premises and empirical evidence demonstrate the impact of metacognitive strategies on creativity (Pesut, 1990) and creative problem solving (Hargrove and Nietfeld, 2015); thus, we assume a non-trivial role of metacognitive strategies and mechanisms that refer to affective-evaluative activity. It is also widely accepted that different creative self-beliefs are engaged in the creative process and their role is vital in initiating the activity, but especially in expending effort in the creative process (see e.g., Karwowski and Beghetto, 2018). Therefore, we expect that self-efficacy or affect-based evaluative behaviors may be prominent during the final stages of the creative process as well.

Another rapidly developing line of a creative process analysis applies neuroscience and behavioral methods (e.g., Beaty et al., 2016, 2018). Although a detailed overview of these approaches is outside of the scope of this article (see Benedek, 2018 for a review), these studies provide compelling evidence of

the integrative character of the creative process. It has been demonstrated that the creative process integrates brain default and executive networks, thus providing the evidence that mind-wandering as well as controlled thinking are simultaneously engaged in the creative process, and – importantly – free as well as controlled processes play roles during all phases of the process. In a similar vein, eye-tracking methods are applied to follow the attention mechanisms involved in the process (e.g., Vartanian, 2009; Beaty et al., 2014; Agnoli et al., 2015; Benedek, 2018). Researchers linked types of processing within the creative process with focused or defocused (Howard-Jones, 2002; Gabora and Ranjan, 2013), or internal versus externally directed attention (Benedek, 2018). Indeed, data obtained thanks to the eye-movement tracking methodology are quite informative for understanding the shifts and subprocesses during idea generation and evaluation (e.g., Ueda et al., 2015). Recent studies demonstrate that the idea generation phase is accompanied by reduced micro-saccade activity and by longer and more frequent blinks (Benedek et al., 2017; Walcher et al., 2017). Findings also suggest that solving insight problems goes with more extended blinks and more gaze aversion (Salvi et al., 2015). An oculometric measure (specifically eye-blink rate; EBR) is an attentional marker of mind-wandering during creative thinking (e.g., Baird et al., 2012; Hao et al., 2015).

The Present Study

The main goal of this exploratory study was to attempt to integrate a relatively static, psychometric approach with a more dynamic, process-based analysis of attention and metacognition functioning during the creative process. To this end, we explored how participants solved the figural creativity test [Test of Creative Thinking-Drawing Production (TCT-DP)], but instead of focusing solely on its outcome (or the final score), we used eye-tracking methodology and thinking-aloud protocols with the hope to provide a more nuanced and dynamic analysis of the process.

MATERIALS AND METHODS

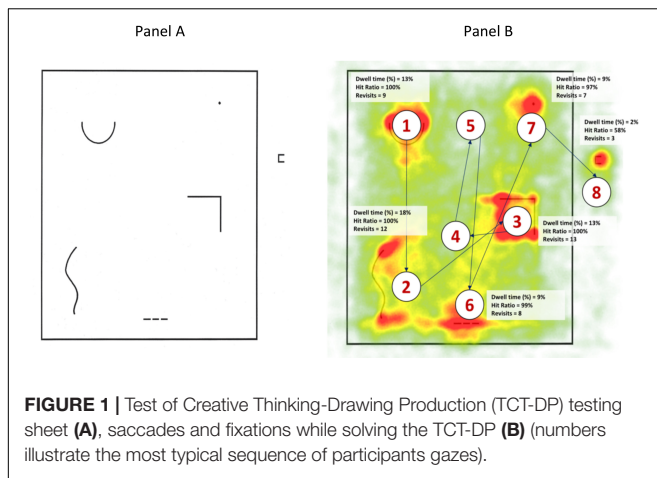
Participants

One hundred participants (50 female and 50 male) aged between 18 and 40 years ($M = 28.82$, $SD = 7.33$) participated in this study. Participants were recruited on the main streets in the center of Warsaw, the capital of Poland and invited to the lab. Participants were remunerated for their time with a one-time payment of 50 PLN (equivalent of approximately 12 euro).

Measures

Test for Creative Thinking-Drawing Production (TCT-DP)

We used Urban and Jellen (1996) TCT-DP. Participants were asked to complete a drawing with six elements placed asymmetrically on a test sheet (see **Figure 1**, panel A). Assessment of the TCT-DP includes fourteen detailed criteria: (1) continuations, (2) completions, (3) new elements, (4) connections made with a line, (5) connections that contribute to



a theme, (6) boundary breaking that is fragment-dependent, (7) boundary breaking that is fragment-independent, (8) perspective, (9) humor and affectivity, (10) unconventionality: manipulation of the test material; (11) unconventionality: surrealistic or abstract elements; (12) unconventionality: use of symbols or signs; (13) unconventionality: unconventional usage of the given fragments and (14) speed. As this study was untimed, we relied on 13 instead of 14 criteria. Speed, as an optional, additional criterion of the test (Urban and Jellen, 1996), was omitted, because the methodology used (especially think-aloud protocols) made the process longer than usual. The final TCT-DP result is the sum of points obtained in all tested criteria. The total score in TCT-DP (without considering speed) may range between 0 and 66 points. Previous studies (e.g., Karwowski et al., 2016b) confirmed validity and reliability of the TCT-DP. In this study, the internal consistency of TCT-DP was comparable to previous studies ($\alpha = 0.74$). The TCT-DP was scored independently by two coders (second and third author with an excellent reliability: $r = 0.987$).

Eye-Tracking

Participants solved the TCT-DP wearing eye-tracking glasses (SensoMotoric Instruments, SMI) with a temporal resolution of 120 Hz. We used the manufacturer's software to calibrate and compute the eye movement parameters: fixations and saccades. Before the study, there was a 4-point calibration procedure. Six main indices were analyzed for each of the area of interests – main elements of the TCT-DP (AOI, see **Figure 1**), specifically: (1) entry time – time in milliseconds to the first fixation within the AOI, (2) dwell time – total time in milliseconds spent on all AOIs in total, (3) hit ratio – the number of participants who fixated within the AOI, (4) revisits – the number of revisits to the specific AOI, (5) average fixation – length of the average fixation within the AOI, and (6) length of the first fixation within the AOI.

Thinking-Aloud Protocols

Metacognitive and self-regulation mechanisms engaged in the creative process were measured by participants' statements and activities during the process of completing the TCT-DP.

The think-aloud statements were audio-recorded after securing proper consent from all participants.

Procedure

The study was conducted individually and lasted between 20 and 45 min. After a short introduction of the goals of the study, obtaining informed consent, and calibrating the eye-tracking glasses, participants filled the TCT-DP. They were instructed to complete a drawing and think aloud during the process – this request was repeated by the researcher if participants tended to draw in silence. Additionally, the following points were emphasized: (1) participants should perform the test in a way that they would if they were not thinking aloud; (2) they should verbalize all thoughts that occur while solving this test, and (3) they should be natural about their reactions. Moreover, participants solved one additional test and filled one questionnaire outside of the scope of this study.

Ethics Statement

This study was carried out after obtaining written informed consent from all subjects. All subjects were informed about the goals of the study and provided informed consent. The protocol was approved by the first author's Institutional Review Board (decision number 128-2016/2017).

RESULTS

The results are presented in four steps. We start with a basic description that illustrates how the process of filling the TCT-DP looked. Then, we switch to the question of whether it is possible to predict psychometric results obtained in the TCT-DP based on eye-tracking results. The third step involves a more detailed analysis of metacognitive strategies and activities during the creative process among individuals who obtained the highest and lowest scores in the TCT-DP. The last step of analyses examined whether metacognitive strategies are related to visual activity during the creative process, as measured by eye-tracking indices.

The Process of Completing the TCT-DP

The average total score obtained in the TCT-DP was in line with previous studies on similar samples in Poland (Gralewski et al., 2016; Karwowski et al., 2016b): $M = 19.15$, $SD = 9.45$. Thus, the results did not suggest that the use of eye-tracking glasses combined with retrospective think-aloud influenced the results due to verbal overshadowing or having the glasses *per se*.

As illustrated on **Figure 1** (panel B), almost all participants focused on five main elements of the test placed within the border, yet almost half of them (42%) completely ignored and omitted the small unfinished square outside. The most typical path of saccades included exploration of the element placed in the upper-left side of TCT-DP (the semicircle) as first, switching to the curve placed in the bottom-left corner next, and then exploring the center-right unfinished square (right angle), bottom dashed line, and the upper-right dot. On average, participants spent most time looking at the bottom curve shape (18% of the total dwell

TABLE 1 | Pearson's correlations and 95% confidence intervals between the total score in the Test of Creative Thinking-Drawing Production (TCT-DP) and main indices obtained in ET study.

TCT-DP with:	Pearson's <i>r</i>	Lower 95% CI	Upper 95% CI
Entry time (ms)	0.49***	0.33	0.63
Dwell time (ms)	0.41***	0.23	0.56
First fixation duration (ms)	−0.03	−0.23	0.17
Revisits	0.46***	0.30	0.61
Fixation count	0.44***	0.27	0.59
Dwell time within AOI (%)	−0.35***	−0.51	−0.16
Average fixation duration (ms)	−0.03	−0.23	0.17

N = 100, ****p* < 0.001 – given several independent tests, the Holm–Bonferroni correction was used to control for Type-I error.

time), then the semicircle and the unfinished square (13% each), while less time was devoted to looking at the line and dot (both 9%), and the unfinished small square outside the frame (2%). Overall, almost 2/3 (63%) of all registered glances were assigned to the six elements of the tests – the remaining ones were linked to the places between elements. The number of revisits to main elements of interests ranged from 7 in the case of the dot to 13 in the case of the unfinished square.

Visual Activity and TCT-DP Results

To examine the extent to which the basic indices registered during eye-tracking are able to predict the psychometric results obtained in the TCT-DP, we proceeded with a two-step procedure. First, we estimated Pearson's correlations between the main indices obtained in the eye-tracking study and the total score of the TCT-DP. Second, we used hierarchical clustering to identify groups with different profiles of gaze distribution and compared TCT-DP scores across the groups. Given the exploratory character of our study and a large number of independent tests, in all cases, we used the Holm–Bonferroni sequential correction for multiple comparisons (Holm, 1979).

As illustrated in **Table 1**, the links between ET indices and TCT-DP scores were significant and robust in terms of the effect size. The more time the participants spent on looking at the main areas of interest, the higher their scores were. Similarly, the more fixations within the AOIs were recorded and revisits to AOIs were found, the higher the scores in the test were. A negative correlation was obtained between the percentage of the dwell time within a single AOI and the total score in TCT-DP. In other words, the more intensively and dynamically the participants explored the test sheet, the higher their scores were, while the more exclusive focus on the certain part of the test resulted in lower scores on average.

Hierarchical cluster analysis (Revelle, 1979; Yim and Ramdeen, 2015) on standardized scores of all ET indices performed with the use of Ward agglomeration technique suggested a four-cluster solution (see **Figure 2**, panel A). We decided to proceed with this solution and indeed four clusters differed in the profile of their visual activity while solving the test (**Figure 2**, panel B).

As illustrated in **Figure 2** (panel B), the first two clusters were characterized by generally low visual focus during the process; the only difference between cluster 1 and 2 was more focus within a certain AOI observed in cluster 1. Thus, participants assigned to the first cluster generally entered into the test quickly (low entry time), briefly scanned all elements, and then tended to focus on selected elements. In the case of people from cluster 2, even more quick and scanning-like functioning was observed. Cluster 3 was composed of participants who focused quite intensively on a certain AOI from the very beginning and proceeded around this specific element. Cluster 4 consisted of people with a more analytical approach – it took them a while to focus on a certain element (relatively long entry time and quick first fixation duration) – then they switched between elements (revisits and many fixations with low dwell time within specific AOIs). In other words, cluster 3 consisted of individuals who seemed to more deliberately compare and combine elements, but their process was more dynamic, while cluster 4 suggested a more analytical approach to solving the test (see **Table 2**).

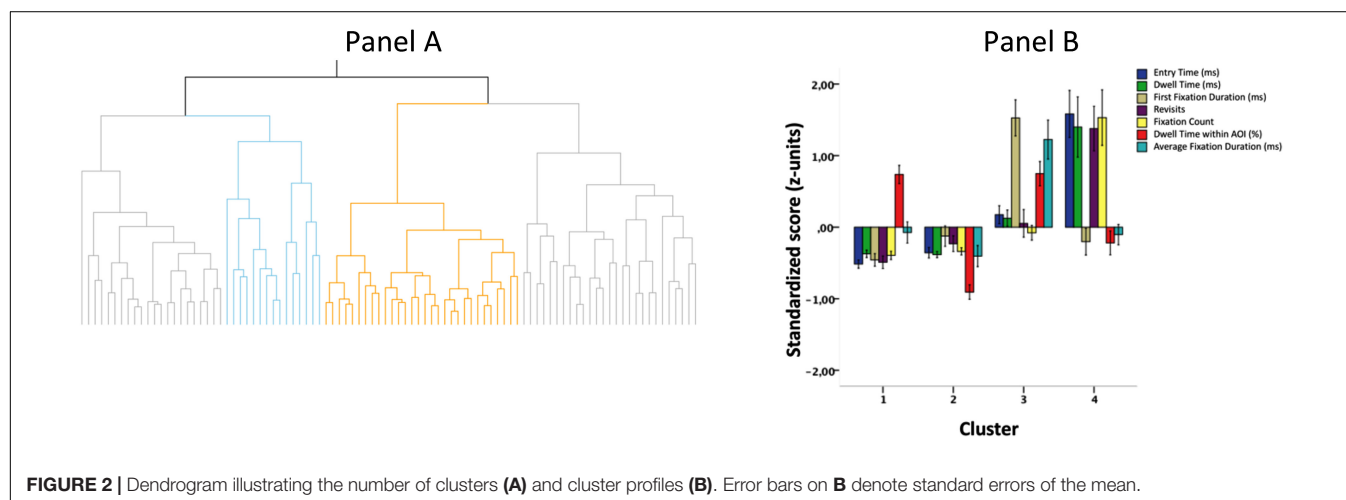
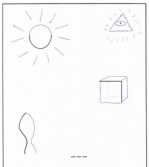





FIGURE 2 | Dendrogram illustrating the number of clusters (A) and cluster profiles (B). Error bars on B denote standard errors of the mean.

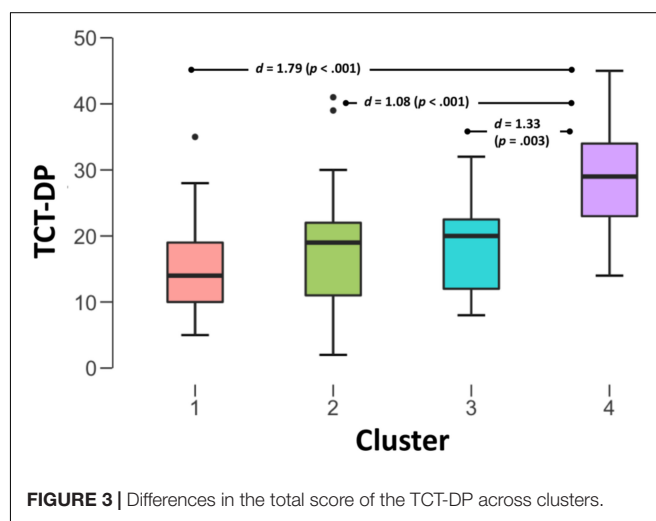
TABLE 2 | Profiles of visual activity while solving the test – analysis of clusters.

TCT-DP: drawing productions	Segments of thinking aloud
Cluster 1 	“The only thing that comes to mind quickly is some abstract cloud, the sun, symbol of God’s providence (...) and I have no more ideas.”
Cluster 2 	“I will go on the easy way and combine all these lines between me (...) and leave such a drawing to the connoisseurs of modern art to guess what came out.”
Cluster 3 	“I will draw a monkey (...) head, eyes, ears (...) here will be a tree because this monkey must sit on something... and the river on which this monkey looks (...)”
Cluster 4 	“(...) I will draw such an obstacle course (...) the ball will fall into all points, or it will bounce off from them... Finally, it lands here [shows on the little square outside the frame] (...) but how to bounce the ball? At the beginning, the ball will fall off the block that will bounce it really high (...)”

An ANOVA was applied to examine whether the clusters differ in the total TCT-DP score. As illustrated on **Figure 3**, the last, fourth cluster was characterized by not only significantly, but also robustly higher total scores than the three remaining clusters (which did not differ from each other – see **Figure 3**). There was a substantial amount of variance in TCT-DP explained by cluster membership, $F(3,96) = 10.13$, $p < 0.001$, $\omega^2 = 0.22$, thus indicating that even relatively simple information about visual activity while solving the test is able to robustly predict the scores obtained by test participants.

Metacognition While Solving the Test

In an effort to examine the more dynamic subprocesses engaged in the creation of the drawing, two coders (first and third authors) independently scored a number of relevant characteristics of this process among a subset of 40 participants – those who obtained the lowest scores in TCT-DP ($n = 20$, $M = 9.55$, $SD = 1.91$) and those with the highest scores ($n = 20$, $M = 32.30$, $SD = 5.19$). Although we classified participants solely based on their total scores, there was a significant and substantial overlap with the clusters described above. In the low-TCT-DP group, none out of 20 participants came from the fourth cluster, while 11 people (55%) were previously assigned to the first cluster. Half of the high TCT-DP group came from the fourth-cluster members ($n = 10$) and this difference in distributions was highly significant, $\chi^2(df = 3, N = 40) = 16.37$, $p < 0.001$, Cramer’s $\Phi = 0.64$.



The description of coded categories is provided in **Table 3** together with descriptive statistics and reliabilities. The coders watched short, recorded movie clips and coded participants’ statements accordingly with the number of proposed segments of behaviors, categorized into three larger groups – exploratory activities, decision-making and control activities, and affective-evaluatory activities. Although in general these three groups of

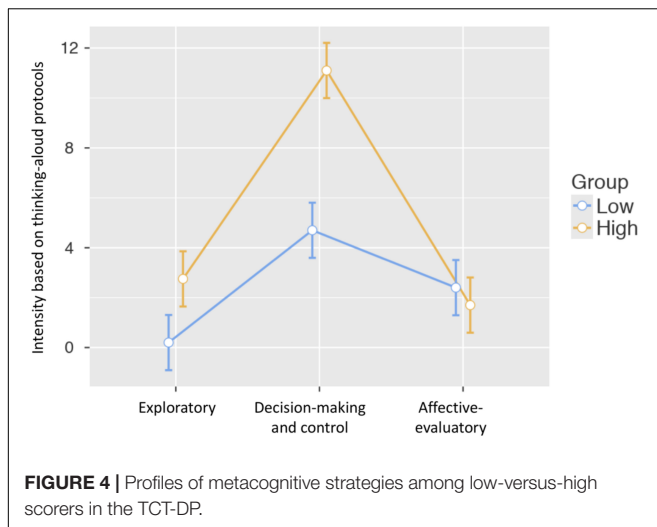
TABLE 3 | Meta-regulation during the creative process – examples of coded segments of participants thinking aloud statements with reliability and descriptive statistics.

Segment code		Example	Overview	Reliability	<i>M</i> (<i>SD</i>)
Exploratory activities	Strategic exploration	"I'm not sure what could it be... maybe a flower?"	Talks to oneself before drawing, analyzing the graphic elements placed on the test sheet	$\alpha = 0.97$	0.56 (0.94)
	Exploration in hand	'I add lines to the semicircle and now I know that it will be a guitar' [combines the semicircle with the curve line].	Discovers the relationship between graphic elements placed on the test sheet while drawing	$\alpha = 0.94$	0.93 (2.06)
Decision-making and control activities	Planning	"Now maybe I'll draw something that will look like helix. This will make it a little more scientific"	Provides justification for planned activity	$\alpha = 0.97$	1.58 (1.93)
	Correction	[Improves drawing of the sun and says] "maybe it will make the moon and the night."	Introduces amendments to the proposed solution	$\alpha = 0.83$	0.45 (0.96)
	Reporting and control	"Could this be a square? [wonders]. Ok, I'll come back to it later."	In a controlled way, postpones the execution of some activities related to the task completion	$\alpha = 0.95$	4.38 (2.77)
	Elaboration	"[Returns to the drawn face] I'll finish drawing a face and a cap I will do on his head... with a visor."	Returns to the proposed solution and elaborates it	$\alpha = 0.99$	1.45 (2.11)
Affective-evaluatory activities	Expressing emotions of a challenge	"I'll think of something... I can draw whatever I want."	Treats the test task as a challenge	$\alpha = 0.77$	0.23 (0.53)
	Expressing emotions of hesitance and uncertainty	"Can it be so simple? I think that it is impossible to do anything about it anymore. . ."	Seeking approval for proposed solutions	$\alpha = 0.94$	0.88 (1.20)
	Positive self-evaluation of skills	"The simplest solutions are the best, I cleverly outplayed it."	Positively evaluates own competencies associated with the performed test task and expresses satisfaction	$\alpha = 0.72$	0.08 (0.27)
	Negative self-evaluation of skills	"I cannot draw. I don't have such imagination!"	Negatively evaluates oneself and own competencies needed to solve the test task	$\alpha = 0.91$	0.30 (0.61)
	Global evaluation	[At the end of solving the test says] "probably it is not too original but... it is simple."	Evaluates proposed comprehensive test solution	$\alpha = 0.73$	0.08 (0.27)
	Evaluation of partial solutions	"It does not look like a butterfly... Oh, well! Let's say it's a butterfly."	Evaluates the partial solution while solving the test	$\alpha = 0.97$	1.50 (1.06)

meta-regulators were indeed observed in a rather subsequent manner, i.e., in most cases, exploratory activities preceded decision-making / control and affective-evaluatory activities, we note that several exceptions from such a step-by-step pattern were observed. Therefore, even if later on we analyze the differences between groups and categories in a processual manner, we emphasize that the process was not necessarily linear, and the phases should be treated in a much more dynamic and reciprocal way.

To explore potential differences between participants with the highest and the lowest scores in TCT-DP, a mixed 3×2

ANOVA was used. Three groups of meta-cognitive strategies (exploratory, decision-making, affective-evaluatory) served as within-person factors, while the group (low versus high TCT-DP scores) became a between-person factor. There were significant and strong differences between the intensity of different meta-cognitive processes, $F(2,76) = 97.3$, $p < 0.001$, $\omega^2 = 0.60$. As illustrated on **Figure 4**, there were significantly more expressions that focused on control and decision-making during the process ($M = 7.90$, $SE = 0.53$) than those focused on the exploratorily-generative and evaluatory talk ($M = 1.48$, $SE = 0.30$ and $M = 2.05$, $SE = 0.32$, respectively). We also observed a robust main effect



of the group, $F(1,38) = 27.4$, $p < 0.001$, $\omega^2 = 0.40$, which demonstrated that participants who obtained high scores in TCT-DP were those who outperformed their counterparts with low scores ($M_{\text{low}} = 1.48$, $SE_{\text{low}} = 0.30$, $M_{\text{high}} = 7.90$, $SE_{\text{high}} = 0.53$). Finally, there was a significant *Process \times Group* interaction, $F(2,76) = 24.3$, $p < 0.001$, $\omega^2 = 0.15$. Although the profiles looked similar (Figure 4), exploratory ($M_{\text{low}} = 0.20$, $M_{\text{high}} = 2.75$, both $SEs = 0.42$) and decision-making/control activities were much more profound within the group that obtained high scores in TCT-DP ($M_{\text{low}} = 4.70$, $M_{\text{high}} = 11.10$, $SEs = 0.75$), while the level of affective-evaluatory activities was similar in both groups ($M_{\text{low}} = 2.40$, $M_{\text{high}} = 1.70$, $SEs = 0.45$). Although we did not observe between-group differences in terms of emotion-based statements during the process, a more detailed analysis showed that there was marginal difference in favor of high-scorers in terms of treating the task as a challenge [Welch's $t(df = 26.66) = 1.52$, $p = 0.07$, one-tailed, Cohen's $d = 0.48$], and a significantly higher level of uncertainty related statements in the low-TCT-DP group [$t(df = 21.69) = 3.79$, $p < 0.001$, Cohen's $d = 1.20$].

Meta-Cognitive Strategies and Eye-Tracking During the Creative Process

The last step of our analyses examined the extent to which the observed metacognitive strategies are related to visual activity during the creative process, as measured by eye-tracking scores (see Table 4). We used a correlational analysis to examine bivariate relationships and regression analysis to control for the covariance among ET indices. As dwell time was almost perfectly correlated with the number of fixations ($r = 0.98$), we excluded the fixation count from our regression models to avoid multicollinearity.

As illustrated in Table 4, there were robust, but also diverse correlations between the intensity of metacognitive strategies and ET scores. The level of exploratory behavior during the creative process was excellently predicted by the total dwell time

spent on exploring all AOIs ($r = 0.80$, $\beta = 0.86$, $p < 0.001$). Interestingly, although bivariate correlations demonstrated that exploratory statements were also related to a number of revisits ($r = 0.67$, $p < 0.001$), this effect disappeared in regression analysis ($\beta = -0.05$).

A number of statistically significant correlations was observed between decision-making and control activities and ET indices – the intensity of this metacognitive strategy was linked to a later entry time ($r = 0.54$) and longer dwell time overall ($r = 0.59$), the number of revisits ($r = 0.57$) and fixations ($r = 0.58$, all $ps < 0.001$), while being negatively correlated with the percentage of time spent on single element ($r = -0.36$). However, when we controlled for the covariance between ET indices, only the entry time marginally predicted decision-making and control activity ($\beta = 0.30$, $p = 0.056$). Affective-evaluatory behaviors during the process were significantly related only to entry time – the later it was, the higher the affective-evaluatory behaviors were ($r = 0.35$, $\beta = 0.48$).

DISCUSSION

How does the creative process look when people are struggling with a psychometric creativity test? Is it possible to explore the process using the test of creativity – an instrument routinely used to capture individual differences rather than creative processes? Are thinking-aloud protocols and eye-tracking glasses able to inform our understanding of this process? These three broad questions largely informed our endeavors presented in this exploratory study. Below, we discuss the main findings and their theoretical consequences, with a special focus on promises and risks related to a blended methodology-based analysis of the creative process.

Our results may be summarized with two broad points. First, even very basic scores obtained thanks to the use of eye-tracking methodology were able to explain quite a substantial portion of the variance of the total score in creativity tests. Not only were such parameters as entry time, dwell time, number of revisits between different elements of the tests, and number of fixations on test's elements, robustly correlated with the total TCT-DP score, but a clear “creative” group emerged when we put eye-tracking scores into hierarchical cluster analysis. This group was characterized by a distinct profile of gaze functioning while solving the test: exploratory on the one hand, but also highly strategic on the other. In short, this cluster combined those who spent a relatively long time while dealing with the test's material, but also very dynamically switched between its main elements, with many fixations overall, but relatively little time spent on a single element. An illustration provided suggested that individuals who solved the TCT-DP in this way looked for a more complex and interpretable solution rather than simply continuing the drawing. And although this line of reasoning is speculative, it is supported by our subsequent analysis of the meta-regulators during the process.

The second main observation refers to the reports from the thinking-aloud protocols during the process. We categorized them into a wide range of specific categories that described

TABLE 4 | Metacognition and eye-tracking – a summary of correlation and regression analyses with intensity of metacognitive strategies regressed onto eye-tracking scores.

ET scores	Metacognitive strategies during the process					
	Exploratory		Decision-making and control		Affective-evaluatory	
	<i>r</i>	β	<i>R</i>	β	<i>r</i>	β
Entry time (ms)	0.23 (–0.09–0.50)	–0.14	0.54** (0.28–0.73)	0.30	0.35* (0.04–0.59) ^b	0.48*
Dwell time (ms)	0.80** (0.65–0.89)	0.86**	0.59** (0.34–0.76)	0.26	0.03 (–0.29–0.33)	–0.01
First fixation duration (ms)	0.11 (–0.21–0.41)	0.10	0.28 (–0.04–0.54)	0.18	–0.01 (–0.32–0.30)	–0.08
Revisits	0.67** (0.45–0.81)	–0.05	0.57** (0.31–0.75)	0.24	–0.06 (–0.37–0.25)	–0.12
Fixation count	0.76** (0.58–0.87)	a	0.58** (0.32–0.75)	a	0.02 (–0.29–0.33)	a
Dwell time within AOI (%)	–0.20 (–0.48–0.12)	–0.15	–0.36* (–0.60–0.06) ^b	–0.12	–0.02 (–0.33–0.29)	0.18
Average fixation duration (ms)	0.04 (–0.28–0.35)	–0.04	0.15 (–0.17–0.44)	0.05	0.07 (–0.25–0.37)	–0.01
<i>R</i> ²		0.66		0.55		0.17

N = 40, a = due to the multicollinearity, fixation count was excluded from the regression model; b = correlation is no longer significant after applying the Holm–Bonferroni correction. **p* < 0.05, ***p* < 0.001.

different behaviors and strategies observed across the different phases of the process. More synthetically, however, all these detailed categories were classified into exploratory activities, identifiable especially during the initial phases of the process, decision-making and control activities – the most severe during the whole process, as well as affective-evaluatory activities – visible not only during the final stages, but in fact, dynamically present during the entire process as well. When we compared individuals, who scored the highest and the lowest in the TCT-DP, it became apparent that the differences between groups were primarily related to the first two groups of strategies and activities. High TCT-DP scorers explored the possible ways of solving the test more intensively, but also put much more energy into the continuous assessment whether their initial drawings fit into the goal. Importantly, though, this goal was not always clear in advance. In other words, for many participants who created the most creative drawings it was not necessarily obvious what should be drawn from the very beginning. Therefore, although their activity was goal-directed, the goal was quite general (“to create something interesting”) rather than specific in terms of the actual theme of their drawings. In a sense, initially this process was blind, spontaneous, and chaotic (Simonton, 1999, 2010), but thanks to the executive strategies to control and order it, it became quite analytical and effective (Benedek, 2018). Individuals who scored highly also demonstrated higher challenge-related statements that suggest their higher self-efficacy and engagement (e.g., Beghetto and Karwowski, 2017; Karwowski and Beghetto, 2018).

On a theoretical as well as methodological level, our results may open new avenues of investigation for the creative process. On the one hand, the presented approach may be promising for researchers who are still looking for more dynamic, accurate, and ecologically valid creativity process assessment (Plucker and Renzulli, 1999). On the other, we attempt to conduct a micro-analysis of attentional patterns during a drawing-production process using cluster analysis and demonstrating various patterns of attentional processing of visual information.

Could these findings inform our theorizing about solving creativity tests or creative process in a more general way? We posit that although it is likely more challenging than the traditional psychometric approaches, such a blended methodology holds the potential to enrich our understanding of the dynamics of creative processes in a wide range of spheres – from solving creativity tests, all the way to a more general process of creative learning (see Karwowski, 2018). We do not suggest that conclusions that stem from mixed methods applications are always straightforward or consistent across methods. Quite the opposite, very often they seem a little chaotic and contradictory. Even if this is true, however, in our perception the blended methodology holds the promise of enriching our understanding of the creative process, especially in comparison with static, output-based assessment of creativity tests.

When interpreting the present findings, a number of strengths and limitations should be considered. Among its strengths, we see that by measuring the creativity process in real-time (Schwarz, 2012), we may potentially reduce recall and retrospective biases and thus collect more valid and reliable data. Moreover, we emphasize the potential of dynamizing creativity test for capturing the ongoing and shifting nature of creativity process and its nuances. For instance, such a triangulated and combined approach may be successfully applicable to exploring the interplay between attentional processes and regulatory self-beliefs (i.e., creative metacognition: Kaufman and Beghetto, 2013) during creative thought. Moreover, we suggest that simultaneously tracking eye movements or physiological responses and examining whether it corresponds with patterns of self-regulation or self-learning strategies uses allows us to investigate the creativity process via a more complex and holistic design.

However, our research had certain limitations as well. First, to analyze the creative process we used only one type of creativity measurement, TCT-DP (Urban and Jellen, 1996). Thus, it is possible that the specific structure of this drawing test may evoke a specific profile of gaze distribution or self-regulation strategies. To test the generalizability of our findings beyond

this *bottom-up* interpretation, other creativity tests or open tasks should be used in future studies. What is more, although in the presented study the think-aloud method did not influence participant performance, we still keep in mind that, specifically during drawing activity, simultaneous verbalization may interfere with the creative process (Lloyd et al., 1995).

As the present study is exploratory, future research is necessary to incorporate relevant moderating and mediating factors, such as creative self-beliefs (Karwowski and Beghetto, 2018) or experience in drawing. This latter factor was unrelated to the results of the TCT-DP in previous studies (Urban and Jellen, 1996), yet it may be important for metacognition. Indeed, as previous studies demonstrated depending on the expertise people differ on organization (meta-regulation) of the creative process (Kay, 1991).

Despite the fact that the overall scores in the TCT-DP test did not differ from those achieved in previous studies conducted in Poland (Karwowski et al., 2016b), it is important to note the potential influence of the instruction we used. As we encouraged the respondents to be “natural about their reactions,” it could be interpreted in different ways by different participants. Previous studies showed that the type of instruction is related to the level of task performance (e.g., O’Hara and Sternberg, 2001; Chua and Iyengar, 2008), and it is possible that while for some of our participants “natural” meant “to be very creative” for others it might have meant quite the opposite – for example esthetically appealing, logical, etc. Although we consider it unlikely that this instruction influenced our findings heavily, future studies should explore this possibility as well.

REFERENCES

- Agnoli, S., Franchin, L., Rubaltelli, E., and Corazza, G. E. (2015). An eye-tracking analysis of irrelevance processing as moderator of openness and creative performance. *Creat. Res. J.* 27, 125–132. doi: 10.1080/10400419.2015.1030304
- Amabile, T. (1996). *Creativity in Context*. Boulder, CO: Westview.
- Baer, J. (1994). Why you shouldn’t trust creativity tests. *Edu. Leadersh.* 51, 80–83.
- Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., and Schooler, J. W. (2012). Inspired by distraction: mind wandering facilitates creative incubation. *Psychol. Sci.* 23, 1117–1122. doi: 10.1177/0956797612446024
- Beaty, R. E., Benedek, M., Silvia, P. J., and Schacter, D. L. (2016). Creative cognition and brain network dynamics. *Trends Cognit. Sci.* 20, 87–95. doi: 10.1016/j.tics.2015.10.004
- Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., et al. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1087–1092. doi: 10.1073/pnas.1713532115
- Beaty, R. E., and Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychol. Aesthet. Creat. Arts* 6, 309–319. doi: 10.1037/a0029171
- Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., and Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Mem. Cogn.* 42, 1186–1197. doi: 10.3758/s13421-014-0428-8
- Beghetto, R. A., and Karwowski, M. (2017). “Toward untangling creative self-beliefs,” in *Creative Self: Effects of Beliefs, Self-Efficacy, Mindset, and Identity*, eds M. Karwowski and J. C. Kaufman (San Diego, CA: Academic Press), 4–24.
- Benedek, M. (2018). “Internally directed attention in creative cognition,” in *The Cambridge Handbook of the Neuroscience of Creativity*, eds R. E. Jung and O.

CONCLUSION

This exploratory investigation examined the possibilities of integrating the psychometric approach to studying creativity with an eye-tracking methodology and thinking-aloud protocols, while studying the creative process. Although primarily methodological, we believe that it also illustrates how such blended approaches may inform more substantial theorizing on the process; theorizing that involves not only cognitive, but also metacognitive aspects of the process.

AUTHOR CONTRIBUTIONS

DJ contributed to the conceptualization of the study, coding the data, and approval of the final version of the manuscript. MC drafted the manuscript and approved the final version of the manuscript. IL contributed to coding the data and approval of the final version of the manuscript. MK contributed to the conceptualization of the study, data analyses, drafting the manuscript, and approval of the final version of the manuscript.

FUNDING

This study was supported by the funding obtained from National Science Centre (UMO-2016/22/E/HS6/00118) for MK.

- Vartanian (Cambridge: Cambridge University Press), 180–194. doi: 10.1017/9781316556238.011
- Benedek, M., Stoiser, R., Walcher, S., and Körner, C. (2017). Eye behavior associated with internally versus externally directed cognition. *Front. Psychol.* 8:1092. doi: 10.3389/fpsyg.2017.01092
- Botella, M., Glăveanu, V., Zenasni, F., Storme, M., Myszkowski, N., Wolff, M., et al. (2013). How artists create: creative process and multivariate factors. *Learn. Individ. Differ.* 26, 161–170. doi: 10.1016/j.lindif.2013.02.008
- Bourgeois-Bougrine, S., Glăveanu, V., Botella, M., Guillou, K., De Biasi, P. M., and Lubart, T. (2014). The creativity maze: exploring creativity in screenplay writing. *Psychol. Aesthet. Creat. Arts* 8, 384–399. doi: 10.1037/a0037839
- Chua, R. Y. J., and Iyengar, S. S. (2008). Creativity as a matter of choice: prior experience and task instruction as boundary conditions for the positive effect of choice on creativity. *J. Creat. Behav.* 42, 164–180. doi: 10.1002/j.2162-6057.2008.tb01293.x
- Csikszentmihalyi, M. (1988). Motivation and creativity: toward a synthesis of structural and energistic approaches to cognition. *New Ideas Psychol.* 6, 159–176. doi: 10.1016/0732-118X(88)90001-3
- Csikszentmihalyi, M. (1996). *Creativity. Flow and the Psychology of Discovery and Invention*. New York, NY: Harper Collins.
- Csikszentmihalyi, M., and Getzels, J. W. (1971). Discovery-oriented behavior and the originality of creative products: a study with artists. *J. Pers. Soc. Psychol.* 19, 47–52. doi: 10.1037/h0031106
- Dunbar, K. (1997). “How scientists think: on-line creativity and conceptual change in science,” in *Creative Thought: An Investigation of Conceptual Structures and Processes*, eds T. B. Ward, S. M. Smith, and J. Vaid (Washington, DC: American Psychological Association), 461–494.
- Finke, R. A., Ward, T. B., and Smith, S. S. (1992). *Creative Cognition: Theory, Research, and Applications*. Cambridge, MA: MIT Press.

- Gabora, L., and Ranjan, A. (2013). "How insight emerges in a distributed, content addressable memory," in *The Neuroscience of Creativity*, eds A. Bristol, O. Vartanian, and J. C. Kaufman (Oxford: Oxford University Press), 19–44.
- Gajda, A., Beghetto, R. A., and Karwowski, M. (2017). Exploring creative learning in the classroom: a multi-method approach. *Think. Skills Creat.* 24, 250–267. doi: 10.1016/j.tsc.2017.04.002
- Gilhooly, K. J., Fioratou, E., Anthony, S. H., and Wynn, V. (2007). Divergent thinking: strategies and executive involvement in generating novel uses for familiar objects. *Br. J. Psychol.* 98, 611–625. doi: 10.1111/j.2044-8295.2007.tb00467.x
- Glăveanu, V. P. (2013). Rewriting the language of creativity: the five A's framework. *Rev. Gen. Psychol.* 17, 69–81. doi: 10.1037/a0029528
- Glăveanu, V. P. (2014). The psychology of creativity: a critical reading. *Creat. Theor. Res. Appl.* 1, 10–32.
- Glăveanu, V. P., Gillespie, A., and Karwowski, M. (2018). Are people working together inclined towards practicality? A process analysis of creative ideation in individuals and dyads. *Psychol. Aesthet. Creat. Arts.* (in press).
- Gralewski, J., Lebeda, I., Gajda, A., Jankowska, D. M., and Wiśniewska, E. (2016). Slumps and jumps: another look at developmental changes in creative abilities. *Creat. Theor. Res. Appl.* 3, 152–177. doi: 10.1016/j.actpsy.2015.09.001
- Hao, N., Wu, M., Runco, M. A., and Pina, J. (2015). More mind wandering, fewer original ideas: be not distracted during creative idea generation. *Acta Psychol.* 161, 110–116. doi: 10.1016/j.actpsy.2015.09.001
- Hargrove, R. A., and Nietfeld, J. L. (2015). The impact of metacognitive instruction on creative problem solving. *J. Exp. Edu.* 83, 291–318. doi: 10.1207/s15327752jpa4505_1
- Hocevar, D. (1981). Measurement of creativity: review and critique. *J. Pers. Assess.* 45, 450–464.
- Hocking, I., and Vernon, D. (2017). A bridge too far: conceptual distance and creative ideation. *Creat. Theor. Res. Appl.* 4, 333–352.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.
- Howard-Jones, P. A. (2002). A dual-state model of creative cognition for supporting strategies that foster creativity in the classroom. *Int. J. Technol. Des. Edu.* 12, 215–226. doi: 10.1023/A:1020243429353
- Isaksen, S. G., Dorval, K. B., and Treffinger, D. J. (2000). *Creative Approaches to Problem Solving*, 2nd Edn. Dubuque, IA: Kendall/Hunt.
- Ivcevic, Z., and Nusbaum, E. (2017). "From having an idea to doing something with it: self-regulation for creativity," in *Creative Self: Effects of Beliefs, Self-Efficacy, Mindset, and Identity*, eds M. Karwowski and J. C. Kaufman (San Diego, CA: Academic Press), 536–576.
- Jauk, E., Benedek, M., Dunst, B., and Neubauer, A. C. (2013). The relationship between intelligence and creativity: new support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence* 41, 212–221. doi: 10.1016/j.intell.2013.03.003
- Karwowski, M. (2018). The flow of learning. *Eur. J. Psychol.* 14, 291–295. doi: 10.5964/ejop.v14i2.1660
- Karwowski, M., and Beghetto, R. A. (2018). Creative behavior as agentic action. *Psychol. Aesthet. Creat. Arts* doi: 10.1037/aca0000190
- Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D. M., Gajda, A., et al. (2016a). Is creativity without intelligence possible? A necessary condition analysis. *Intelligence* 57, 105–117. doi: 10.1016/j.intell.2016.04.006
- Karwowski, M., Jankowska, D. M., Gralewski, J., Gajda, A., Wiśniewska, E., and Lebeda, I. (2016b). Greater male variability in creativity: a latent variables approach. *Think. Skills Creat.* 22, 159–166. doi: 10.1016/j.tsc.2016.10.005
- Kaufman, J. C., and Beghetto, R. A. (2013). In praise of clark kent: creative metacognition and the importance of teaching kids when (not) to be creative. *Roeper Rev.* 35, 155–165. doi: 10.1080/02783193.2013.799413
- Kay, S. (1991). The figural problem solving and problem finding of professional and semiprofessional artists and nonartists. *Creat. Res. J.* 4, 233–252. doi: 10.1080/10400419109534396
- Krampen, G. (2012). Cross-sequential results on creativity development in childhood within two different school systems: Divergent performances in Luxembourg versus German kindergarten and elementary school students. *Eur. J. Psychol.* 8, 423–448. doi: 10.5964/ejop.v8i3.468
- Kulkarni, D., and Simon, H. A. (1988). The processes of scientific discovery: the strategy of experimentation. *Cogn. Sci.* 12, 139–175. doi: 10.1016/j.coph.2009.08.004
- Langley, P., Simon, H. A., Bradshaw, G. L., and Zytkow, J. M. (1987). *Scientific Discovery: Computational Explorations of the Creative Process*. Cambridge, MA: MIT Press.
- Lloyd, P., Lawson, B., and Scott, P. (1995). Can concurrent verbalization reveal design cognition? *Des. Stud.* 16, 237–259. doi: 10.1016/0142-694X(94)00011-2
- Mumford, M. D., Baughman, W. A., Maher, M. A., Costanza, D. P., and Supinski, E. P. (1997). Process-based measures of creative problem-solving skills: IV. Category combination. *Creat. Res. J.* 10, 59–71. doi: 10.1207/s15326934crj1001_7
- Mumford, M. D., Mobley, M. I., Uhlman, C. E., Reiter-Palmon, R., and Doares, C. (1991). Process-analytic models of creative capabilities. *Creat. Res. J.* 4, 91–122. doi: 10.1080/10400419109534380
- O'Hara, L. A., and Sternberg, R. J. (2001). It doesn't hurt to ask: effects of instructions to be creative, practical, or analytical on essay-writing performance and their interaction with students' thinking styles. *Creat. Res. J.* 13, 197–210. doi: 10.1207/S15326934CRJ1302_7
- Perry, S. K. (1999). *Writing in Flow: Keys to Enhanced Creativity*. Cincinnati, OH: Writer's Digest Books.
- Pesut, D. J. (1990). Creative thinking as a self-regulatory metacognitive process: a model for education, training, and further research. *J. Creat. Behav.* 2, 105–110. doi: 10.1002/j.2162-6057.1990.tb00532.x
- Plucker, J. A., and Renzulli, J. S. (1999). "Psychometric approaches to the study of human creativity," in *Handbook of Creativity*, ed. R. J. Sternberg (New York: Cambridge University Press), 35–60.
- Plucker, J. A., and Runco, M. A. (1998). The death of creativity measurement has been greatly exaggerated: current issues, recent advances, and future directions in creativity assessment. *Roeper Rev.* 21, 36–39. doi: 10.1080/02783199809553924
- Pringle, A., and Sowden, P. T. (2016). The mode shifting index (MSI): a new measure of the creative thinking skill of shifting between associative and analytic thinking. *Think. Skills Creat.* 23, 17–28. doi: 10.1016/j.tsc.2016.10.010
- Pringle, A., and Sowden, P. T. (2017). Unearthing the creative thinking process: fresh insights from a think aloud study of garden design. *Psychol. Aesthet. Creat. Arts* 11, 344–358. doi: 10.1037/aca0000144
- Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behav. Res.* 14, 57–74. doi: 10.1207/s15327906mbr1401_4
- Root-Bernstein, R., and Root-Bernstein, M. (2001). *Sparks of Genius: The 13 Thinking Tools of the World's Most Creative People*. Boston, MA: Houghton Mifflin.
- Rubenstein, L. D., Callan, G. L., and Ridgley, L. M. (2017). Anchoring the creative process within a self-regulated learning framework: inspiring assessment methods and future research. *Edu. Psychol. Rev.* 30, 921–945. doi: 10.1007/s10648-017-9431-5
- Runco, M. A., and Jaeger, G. J. (2012). The standard definition of creativity. *Creat. Res. J.* 24, 92–96. doi: 10.1080/10400419.2012.650092
- Salvi, C., Bricolo, E., Franconeri, S. L., Kounios, J., and Beeman, M. (2015). Sudden insight is associated with shutting out visual inputs. *Psychon. Bull. Rev.* 22, 1814–1819. doi: 10.3758/s13423-015-0845-0
- Sawyer, K. (2012). *Explaining Creativity: The Science of Human Innovation*, 2nd Edn. Oxford: Oxford University Press.
- Schwarz, N. (2012). "Why researchers should think 'real-time': a cognitive rationale," in *Handbook of Research Methods for Studying Daily Life*, eds M. R. Mehl and T. S. Conner (New York, NY: Guilford Press), 22–42.
- Silvia, P. J., Nusbaum, E. C., and Beaty, R. E. (2017). Old or new? Evaluating the old/new scoring method for divergent thinking tasks. *J. Creat. Behav.* 51, 216–224. doi: 10.1002/jocb.101
- Simonton, D. K. (1999). Creativity as blind variation and selective retention: is the creative process Darwinian? *Psychol. Inq.* 10, 309–328.

- Simonton, D. K. (2010). Creative thought as blind-variation and selective-retention: combinatorial models of exceptional creativity. *Phys. Life Rev.* 7, 156–179. doi: 10.1016/j.plrev.2010.02.002
- Ueda, Y., Tominaga, A., Kajimura, S., and Nomura, M. (2015). Spontaneous eye blinks during creative task correlate with divergent processing. *Psychol. Res.* 80, 652–659. doi: 10.1007/s00426-015-0665-x
- Urban, K. K., and Jellen, H. G. (1996). *Test for Creative Thinking-Drawing Production (TCT-DP)*. Lisse: Swets and Zeitlinger.
- Vallee-Tourangeau, F., Anthony, S. H., and Austin, N. G. (1998). Strategies for generating multiple instances of common and ad hoc categories. *Memory* 6, 555–592. doi: 10.1080/741943085
- Vartanian, O. (2009). Variable attention facilitates creative problem solving. *Psychol. Aesthet. Creat. Arts* 3, 57–59. doi: 10.1037/a0014781
- Walcher, S., Körner, C., and Benedek, M. (2017). Looking for ideas: eye behavior during goal-directed internally-focused cognition. *Conscious. Cogn.* 53, 165–175. doi: 10.1016/j.concog.2017.06.009
- Walker, W. H., and Kintsch, W. (1985). Automatic and strategic aspects of knowledge retrieval. *Cognit. Sci.* 9, 201–223. doi: 10.1016/j.neuropsychologia.2016.05.003
- Wallas, G. (1926). *The Art of Thought*. New York, NY: Harcourt, Brace, & World.
- Ward, T. B. (2001). Creative cognition, conceptual combination and the creative writing of Stephen R. Donaldson. *Am. Psychol.* 56, 350–354. doi: 10.1037/0003-066X.56.4.350
- Weisberg, R. W. (1986). *Creativity: Genius and Other Myths*. New York: Freeman.
- Yim, O., and Ramdeen, K. T. (2015). Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. *Q. Methods Psychol.* 11, 8–21. doi: 10.20982/tqmp.11.1.p008

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 Jankowska, Czerwonka, Lebuda and Karwowski. 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.



Creativity in the Here and Now: A Generic, Micro-Developmental Measure of Creativity

Elisa Kupers^{1,2*}, Marijn Van Dijk² and Andreas Lehmann-Wermser³

¹ Department of Special Needs Education and Youth Care, University of Groningen, Groningen, Netherlands, ² Department of Developmental Psychology, University of Groningen, Groningen, Netherlands, ³ Institut für Musikpädagogische Forschung, Hochschule für Musik, Theater und Medien Hannover, Hanover, Germany

OPEN ACCESS

Edited by:

Philip A. Fine,
University of Buckingham,
United Kingdom

Reviewed by:

Ana-Maria Olteteanu,
Freie Universität Berlin, Germany
Guido Bugmann,
Plymouth University, United Kingdom
Frank Loesche,
Plymouth University, United Kingdom,
in collaboration with reviewer GB

*Correspondence:

Elisa Kupers
w.e.kupers@rug.nl

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 19 April 2018

Accepted: 11 October 2018

Published: 08 November 2018

Citation:

Kupers E, Van Dijk M and
Lehmann-Wermser A (2018) Creativity
in the Here and Now: A Generic,
Micro-Developmental Measure of
Creativity. *Front. Psychol.* 9:2095.
doi: 10.3389/fpsyg.2018.02095

Creativity is a relevant yet elusive concept, and consequently there is a large range of methods to assess creativity in many different contexts. Broadly speaking, we can differentiate between creativity measures on the level of the person (such as the Torrance tests), the level of the creative product (consensual assessment), and the level of the creative process. In the recent literature on children's creativity, 80% of the studies employed measures on either the person or the product level (Kupers et al., submitted). However, for parents, teachers, and employers who wish to stimulate creativity, insight in the (often socially embedded) creative process is badly needed. This move from the inter-individual to the intra-individual level of assessment is furthermore in line with research in many other domains in psychology. Although there is some research focusing more on detailed descriptions of creative processes, the studies are usually purely qualitative and therefore highly context-specific, making generalization difficult. In this paper, we present a newly developed coding frame as a systematic, generic, micro-level measure of creativity. What is unique about this coding frame is that it can be applied to observations of creative processes in many different contexts, and for different kinds of creative tasks. The core of the instrument is that it allows us to assess the two core components of creativity - novelty and appropriateness on an ordinal 4-point scale, at each moment during the creative process. The coding frame can be applied in three steps. The first step is to determine the unit of analysis, that is, the level of detail in which the creative process is assessed. The second step and third steps are coding the units on two ordinal scales of novelty and appropriateness, respectively. In order to illustrate the versatility of our instrument, we apply it to two cases of very different creative processes: a musical composition task (open-ended) and a scientific reasoning task (closed-ended). Last, we demonstrate the possibilities for analyzing this type of dense intra-individual measurements of creativity (time series analysis and state space grids) and discuss the future research that is needed to fully validate the instrument.

Keywords: creativity, microgenetic theory, process research, observational methods, scientific reasoning, musical composition

INTRODUCTION

Creativity is the human capacity to use your imagination and create to create solutions for complex problems (Welch and McPherson, 2012). Therefore, it is essential for our survival and prosperity. Creativity has been recognized as one of the most important “twentieth century skills,” which should be leading in shaping current and future educational policy and practice. For teachers, managers, and others who wish to stimulate creativity, it is therefore important to gain an understanding of how creative processes unfold in the here and now.

In the study of human behavior, there is currently an increasing interest in real-time processes relating to fundamental issues such as intra-individual variability as a mechanism of change. These developments are being further enhanced by technological advancements that make the collection of dense intra-individual data more feasible, such as EMA (ecological momentary assessment: Shiffman et al., 2008), eye-tracking, (e.g., Odean et al., 2015), and wireless heart rate monitoring (e.g., Vickhoff et al., 2013; Gregersen et al., 2014). Within empirical research on creativity, however, processes in the here and now are often overlooked—possibly due to a lack of systematic, quantitative measurement instruments that can be used for measuring creativity across a variety of contexts. In this article, we will explain both why such an instrument is indispensable and which criteria it needs to meet. We will present a basic coding frame for assessing the key elements of creative processes (novelty and appropriateness), and will use two empirical examples to illustrate how this framework can be applied. Furthermore, we will describe the steps involved in applying the framework to two particular cases of creative behavior (during an open-ended musical composition task and during a closed-ended scientific reasoning task). We conclude with implications for creativity research and the next steps needed in order to further validate the instrument.

CREATIVITY AS NOVELTY AND APPROPRIATENESS

Creativity is defined as “the interaction among aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined in a certain social context” (Plucker et al., 2004, p.90). On the one hand, creativity is something unexpected; something beyond what is already known at a certain point. On the other hand, the definition implies that creativity requires more than just novelty; the response or product must also be useful or appropriate (Cropley, 2006; Runco and Jaeger, 2012). It must be a fitting solution to the task or problem at hand. The characteristics of novelty and appropriateness relate to the two distinct processes that together make up creativity. The first process is *divergent thinking*, which is the skill to generate as many possible solutions to a problem as possible. Divergent thinking requires a person to be able to associate quickly, make unexpected links between components, and transform information into

unexpected forms (Guilford, 1967a; Runco, 2010). Three features of divergent thinking are usually assessed: fluency, flexibility, and originality (Guilford, 1957; Sternberg, 2006; Baas et al., 2008). Fluency refers to the amount of unique ideas a person is able to generate within a fixed amount of time. Flexibility is the capacity to be able to quickly switch between approaches to and characteristics of the problem at hand. Consider the example where a child is asked to come up with as many uses of a paperclip as possible. One child may respond: a paper binder, a necklace, a tool to open a lock. Another child may respond: a necklace, a bracelet, earrings. Although the fluency of these two sets of responses is the same, the second child demonstrates a lower level of flexibility as each solution stems from the same overall semantic category (jewelry). The third component, originality, refers to the uniqueness of an idea or solution. When comparing children’s responses to the “paperclip problem,” some responses might be very common (such as the “paper binder” response) while others are more uncommon (such as the “tool to open a lock” response). In creativity research, divergent thinking is often equated with creative thinking. However, as previously mentioned, divergent thinking entails more than just novelty; usefulness or appropriateness is also important. For true creativity, we need to evaluate whether the many solutions generated contribute in any way to solving the problem or finishing the task at hand. This involves *convergent thinking*. While divergent thinking is the generation of as many solutions to a problem as possible, convergent thinking is defined as “oriented toward deriving the single best (or correct) answer to a clearly defined problem” (Cropley, 2006, p. 391). Convergent thinking is closely connected to using prior knowledge; in order to arrive at the best solution to a problem, one must know what is already known about the problem and build on that existing knowledge. A problem has certain aspects or constraints, and being able to deal with these task constraints is what eventually determines whether an idea is actually creative (Cropley, 2006; Glăveanu, 2013b, 2014). This applies across domains, from clearly defined scientific problems, to literature, poetry, or music (Cropley, 2006). In cognitive models of creativity, divergent, and convergent thinking are often closely interlinked. Within the theory of blind variation and selective retention (BVSr) for example, divergent thinking plays a role in ideation or the generation of possible ideas, convergent thinking mainly in the selection of fruitful ideas (Simonton D. K., 1999; Simonton K., 1999; Simonton, 2015).

ASSESSING CREATIVITY

When looking at how creativity can be assessed, a distinction can be made between three levels on which creativity can be measured: the level of the person, the level of the product, and the level of real-time actions. A similar distinction is made in Rhodes’ “4P” model of creativity (see Rhodes, 1961), where he distinguishes between creativity on the levels of the Person, Product, Process, and Press (the latter referring to environmental influences). These levels of measurement differ in the extent to which they see creativity as an aggregated construct—for

instance, as the average across moments, products or even a person's lifetime (Kupers et al., submitted). The highest level of measurement is the level of the **person**. Here, creativity is seen as a personal characteristic that may or may not change over time. Assessments of creativity on this level can answer questions about differences between groups of people—for instance between men and women (Baer and Kaufman, 2008), between cohorts of different generations (Kim, 2011), or between children with and without developmental disorders (Healey and Rucklidge, 2006; Tafti et al., 2009; Kim and VanTassel-Baska, 2010). Alternatively, questions can be answered about the relation between creative thinking and other personal variables, such as IQ. The most frequently used assessments on this level are tests for divergent thinking, such as the “Guilford Alternative Uses test” (Guilford, 1967b) or the “Torrance Tests of Creative Thinking” (Torrance, 1966). These types of tests come in many different forms, but they all involve asking someone to come up with many different responses to a single problem. This can be a verbal task—such as when someone is asked to come up with as many alternative uses of a brick as possible—or a non-verbal task—for instance, completing a drawing based on one shape. The extent to which a person is then considered creative depends on how their responses score for flexibility, fluency, and originality. Some of these divergent thinking tasks (such as the Torrance Test of Creative Thinking) also take into account a score for elaboration (see for instance Torrance, 1966).

In the past few decades of creativity research, the most prominent way of assessing creativity has been through divergent thinking tasks (Long, 2014; Kupers et al., submitted). Another type of creativity test is formed by problem-solving tasks, in which one specific way of solving the problem tends to be considered the “correct” response. For this reason, these types of tests mainly assess convergent thinking. Some less commonly used measures of creativity on the level of the person include personality tests or interviews, either self- or other assessments (e.g., Runco et al., 2001; Butcher and Niec, 2005; Kaufman et al., 2010; Putwain et al., 2012). In the domain of self-report questionnaires, a distinction can be made between self-reported creative achievements or behaviors on the one hand (participants rating whether they wrote a book, achieved success in an artistic domain, etc.), questionnaires or interviews of creative self-concept (participants' ideas about whether they view themselves as creative) on the other. Both types of self-reported creativity can be assessed in a reliable and valid way (Silvia et al., 2012). Creativity is also assessed by having others evaluate creative **products**, such as written poems or stories, musical compositions and paintings. This type of assessment acknowledges that the decision regarding “what is truly creative” is inherently intersubjective; something is creative when people who are familiar in the domain judge it as creative. These types of assessments are commonly known as “consensual assessments,” based on the work of Amabile (1983, 1996). Similarly to assessments on the person level, assessments on the product level can be used to answer questions about group differences in creativity—but they are also used to measure the effect of (educational) interventions (e.g., Patera et al., 2008).

On the level of real-time actions, studies zoom in on the creative **process** as it occurs in the behaviors of individuals

from moment to moment. These types of studies aim to get more insight into things like how the creative process unfolds, whether a distinction can be made between different “stages” of the creative process, etc. (e.g., Burnard and Younker, 2004). The creative processes that are studied can be either individual or more socially situated. Studies on social creativity are focused on questions of whether and how social interactions, such as interactions between peers or with a teacher, help to shape creativity (e.g., Vass, 2007; Fernández-Cárdenas, 2008; Chappell and Craft, 2011; Glăveanu, 2013a). In the “Four Ps” model of creativity, these environmental influences are referred to as “**press**” (Rhodes, 1961). The data in these studies on socially situated creativity are almost always qualitative—such as video observations or field notes, which are coded “bottom-up” to make sense of the data. In a systematic review of empirical literature on children's creativity published in the last decade (Kupers et al., submitted), we found that the vast majority of papers (80%) assessed creativity either on the person level or on the product level, as described above. This is in line with earlier work by Long (2014). Although his categorization system is slightly different, we can conclude that in the last two decades creativity research has shifted—from largely qualitative process descriptions of creativity, toward largely quantitative descriptions of creativity being quantitatively by means of creativity tests. This type of quantitative research, which assesses creativity on a more aggregated level (the level of the person), has provided valuable insights into group differences in overall creativity (e.g., Baer and Kaufman, 2008; Cheung and Lau, 2010). Moreover, these measures are often used to evaluate the effect of (educational) interventions targeting creativity (e.g., Hu et al., 2013; Dziedziewicz et al., 2014). Then again, the danger of focusing on creativity on these aggregated levels is that the core of creativity, namely the creative process (Glăveanu, 2013b; Kupers et al., submitted), is overlooked. Qualitative studies on the process level of creativity have offered rich, detailed descriptions of many different types of creative processes. However, due to the type of analysis used—which is intrinsically qualitative, ethnographic, and “bottom-up”—it is very difficult to generalize any findings beyond their original context, or to test hypotheses regarding different kinds of processes or conditions.

In order to measure creativity in real time, there must be a focus on (real-time) behavior in the “here and now” in a specific context. Such a measure would enable us to describe the “microdevelopment” of creativity: the development of creativity that unfolds during a short time span (days, hours, minutes). A micro-developmental study takes the changing individual—together with his or her immediate social and physical environment, such as the interaction between a child, teacher and task—as the fundamental unit of analysis (Granott and Parziale, 2002; Lavelli et al., 2005). For this purpose, micro-developmental studies use dense observations and employ intensive analyses to capture the processes of change (Siegler and Crowley, 1991; Granott et al., 2002). Many studies that look into micro-developmental changes during a task stem from the domain of cognitive development (such as Siegler and Chen, 1998), but the term is also used in studies within other domains—such as problem-solving (Chen and Siegler, 2000), mother-infant communication (Lavelli and Fogel, 2002), early emotional

development (de Weerth et al., 1999), and second language acquisition (Sun et al., 2016). Micro-developmental data are more detailed than data collected through other methods, and can be used to analyze trial-by-trial variability, detect transitions, and analyze instructional manipulations (Siegler, 2002). The idea behind this approach is to examine changes as they are occurring (Siegler and Crowley, 1991). Gaining this type of knowledge about creative processes is of crucial importance for theory building, and indispensable for anyone who wishes to stimulate creativity. In order to take the field of creativity research a step further, an instrument is needed that enables researchers to assess creativity on the level of the creative process as it unfolds from moment to moment, in the here and now. This instrument should preferably be applicable to many different contexts, thereby making it possible for researchers to compare contrasting processes and to draw conclusions about individual differences. In the remainder of this article, we present such an instrument. The method we propose has its roots in qualitative methodology of systematic coding (Gläser and Laudel, 2013) and qualitative research into individual and social creativity. However, the proposed method is new in the sense that it *quantifies* qualitative data on two ordinal scales. This enables the micro-developmental analysis of patterns within creative behavior. Some specific options regarding this type of analysis are also presented in the remainder of this article.

A GENERIC MICRO-DEVELOPMENTAL CODING FRAMEWORK OF CREATIVITY

If we aim to measure creativity on a “real-time” level—that is, as creativity occurs in the here and now—we need to focus on both aspects of its definition: novelty and appropriateness. In the next section, we will describe three necessary steps toward constructing a generic coding scheme tailored to meet the needs of specific contexts in which creativity is measured. It is important to note that what we are offering here is a *framework* (including guidelines) for coding creative processes, which researchers can use to construct their own coding schemes. For that reason, we will present a detailed illustration of how the coding framework can be applied and tailored to specific data.

Step 1: Determine the Unit of Analysis

When assessing creativity from moment to moment, the first step is to determine what those “moments” or units are. It is important to note that this decision depends on the nature of the particular creative processes being studied. For instance, when a professional artist is making a painting or a sculpture, every small variation or new idea is likely to take considerable time to prepare and execute. In this case, each “turn” can take minutes. However, in a situation where two students have to write a poem together, they may well think out loud, trying out different combinations of words, sounds and meanings in rapid succession. In this case, each turn may only take a matter of seconds. Since this can differ for different creative processes, a unit of analysis is always based on observable behavior of the individual.

In order to determine a valid unit of analysis, it is important to consider the following criteria. First, to be able to analyze trends over time within the execution of an assignment or the making of a product, the codes (which will result from the coding scheme as a whole) must be sufficiently detailed. Second, units of analysis should be on the level of ideas or variations. Again, what these ideas or variations are depends on the nature of the creative process. If the process is primarily verbal, or if the product is in written language (stories, poems, scientific reasoning, etcetera), a straightforward choice for the unit of analysis would be each verbal (spoken) turn or utterance. In this case, transcripts can be based on an existing language transcription system such as CHILDES (MacWhinney, 2000), which offers guidelines for determining utterance boundaries and turns. If the creative process is primarily non-verbal (dance, arts), the unit of analysis could be any meaningful action (within dance this might be each movement, turn or step; within visual arts it could be adding new lines or figures to a drawing or painting; within musical composition it could be each musical motive). If the creative process is primarily non-verbal, but also includes verbal elements (for instance, students working together on a musical composition and negotiating which musical motives to add to the overall composition), then turns can be either verbal (spoken turns), non-verbal (meaningful actions), or a combination of both (i.e., proposing an idea and executing it at the same time would be coded as one turn).

Determining the lowest level categories is crucial for *any* coding system, and this level should be defined both conceptually and operationally. It is recommended that researchers describe the units of analysis in conceptual terms, provide prototypical examples, and also describe non-units and examples that would not be coded as a unit (Yoder and Symons, 2010). For instance, in a specific study on creativity in the building of a tower, researchers may decide to code each time a child picks up a block, each time a block is placed, and each time a block is taken away—but *not* each time the child scratches his nose, or merely touches a block. As with any specific coding system, researchers should be trained in determining units of analysis. To make the procedure more transparent, it is also recommended for inter-observer reliability to be established on the level of unit segmentation before codes are assigned (Srijbos et al., 2006). Any disagreement between researchers can be used to refine the decision guidelines concerning unit segmentation, until reliability is satisfactory.

Step 2: Code Each Unit for Novelty

Once the units of analysis have been determined, and the segmentation of the data has proven to be reliable, the next steps consist of coding each unit. This must be done on both the novelty and appropriateness dimensions (see **Figure 1**). These dimensions are summarized in **Tables 1,2** below. Since novelty and appropriateness are relative terms (something is novel compared to what?), it is important to bear in mind that novelty is assessed on an intra-individual level—that is, *something is assessed as novel or less novel compared to what has happened up until that moment*. Importantly, this is in contrast to common divergent thinking tasks, which assess how

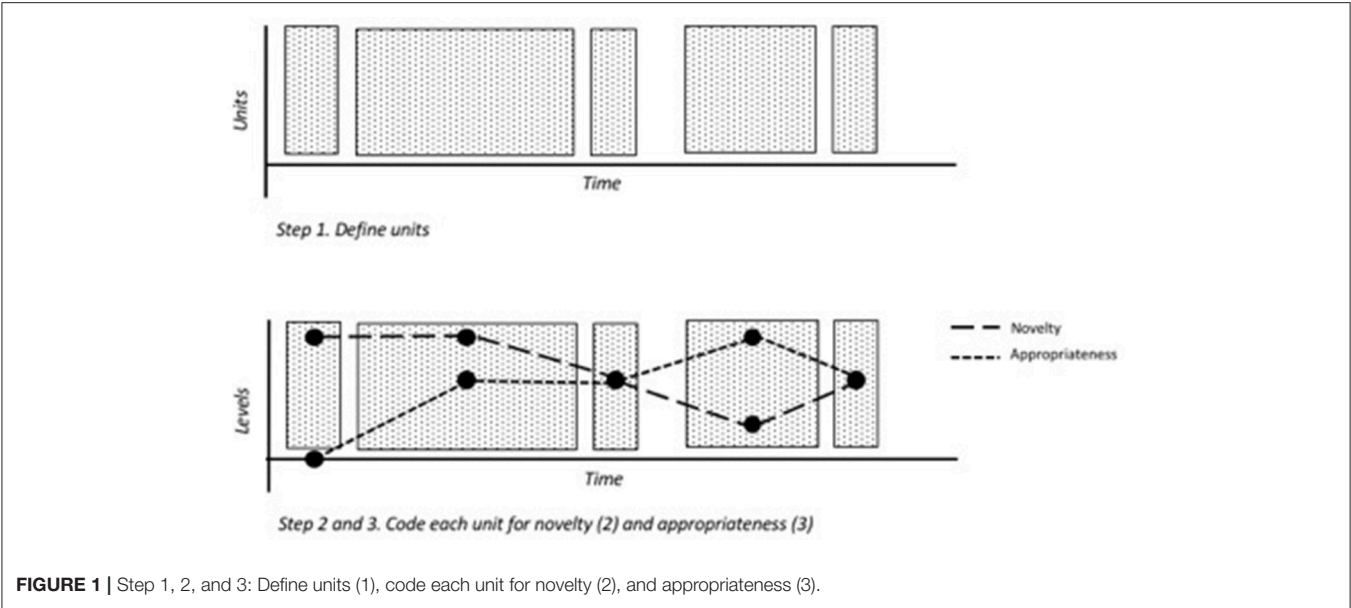


TABLE 1 | Coding frame for the novelty dimension.

Level	Description
0	The current turn/idea is a repetition of the previous turn. No new information is added. Also: a confirmation of the previous idea (agreeing with previous, listening back, evaluating, summarizing the process so far). A rejection of an idea without proposing an alternative.
1	The current turn is related to the previous turn, but also adds something new. Small elaboration: The current idea contains subtle differences compared to the previous idea, small variations on one main idea
2	Large elaboration: The current idea contains more substantial differences, or multiple elements are added
3	The current turn is a new idea , initiative or proposition. The idea has no common elements compared to the previous turns. The child verbalizes a new component that has not been mentioned before, uses new material or performances completely novel actions upon the existing material.

novel a response is compared to the responses of a norm group.

The core of the novelty dimension, as described in **Table 1**, is assessing how much the current idea or turn has in common with the previous ideas that have already been observed. The categories are loosely based on the coding scheme of Miell and Macdonald (2000), in which a distinction is made between transactive turns (elaborating on what has previously been said or done) and non-transactive turns (either adding no new information or going in a completely new direction). When we translate this to the construct of novelty, three to four

TABLE 2 | Coding frame for the appropriateness dimension.

Level	Description
0	Off-task behavior: The child walks away from the task, talks about completely unrelated subjects, or does something unrelated to the task
1	Somewhat related to task: The child uses task materials in a way not obviously related to task, or talks about something only remotely related to the task
2	On-task behavior: The child shows focused, concentrated work related to the task
3	Explicit reference to task elements: The child makes a link (verbal or non-verbal) with elements that are embedded in the task

ordinal categories can be distinguished conceptually: a turn with no novel elements, a turn with partially prior elements, and partially novel elements (possibly with subcategories), and a turn with only novel elements. On the lowest level (0), the current turn adds no new elements to the turns before; it is simply a repetition or confirmation of the ideas up until this point. Regarding verbal responses, saying “I don’t know” or disapproving of an idea without offering another suggestion also fall into this category. Level 1 and 2 are both elaborations, meaning the current turn builds upon previous turns (it has some common elements compared to the previous turns, but also adds something new). In most cases, a distinction can be made between small elaborations (level 1, in which only one element is added or more subtle changes are made) and large elaborations (level 2, in which more elements are added or more substantive changes are made). Level 3 means the current idea does not contain any elements that had already been mentioned.

Again, the specific descriptions of which behaviors (units of analysis) belong to which of the four categories should be described in any specific coding scheme. At this step, conceptual and operational guidelines should be established again, as should inter-observer reliability.

Step 3: Code Each Unit for Appropriateness

The core of the appropriateness dimension is to assess how much the current turn fits the overall task or assignment. As is the case with novelty, the exact number of categories on the ordinal scale can be adjusted to the nature of the task—but on conceptual grounds, we propose an ordinal scale of at least three or, if possible, four categories. The lowest level (level 0) is off-task behavior, such as talking about unrelated topics or walking away from the task. Level 1 codes are assigned to behaviors that have some relation to the task, but use task elements in a way that is not clearly related to the task. Examples of level 1 codes would be dancing to the music when composing a musical piece on a computer, or talking about hospital syringes during a linking syringes task at school. Level 2 is assigned to behaviors that are focused and on-task, such as (in case the task is making a musical composition) browsing through a library of music loops and clicking on several to see whether they sound appealing. Finally, Level 3 is assigned to behaviors that explicitly refer to specific task elements or how to complete a task (it can contain metacognitive elements). An example of level 3 behaviors in a task where the aim is to link two syringes and make the air go from one syringe to the other, would be when the child pushes one of two connected syringes or says “Now the air goes from here to here!” As in the previous two steps, researchers of any topic should define beforehand which behaviors belong to which of the levels, train coders, and establish reliability.

TWO EMPIRICAL EXAMPLES

In the following section, the coding framework is applied to two case studies. These were selected as representative cases out of larger samples of video data, taken respectively from a study on musical creativity (Kupers, 2013) and a study on scientific reasoning and problem-solving (Guevara-Guererro, 2015). We present these examples here to demonstrate the steps that need to be taken in order to construct a coding scheme for coding creative behaviors in a specific task, on the basis of the framework presented in this article. For this demonstration, we chose two contexts in which the generation of new ideas plays an important role, but that were very different in other regards. The first example concerns an open-ended task in the context of music education, and the second one concerns a closed-ended task in elementary school science. The case studies serve to illustrate the potential for applying a generic measure in many different (educational) contexts. Due to the illustrative nature of these case studies, and since determining the inter-observer reliability of a coding scheme quantitatively generally requires more data than just two short cases, calculating inter-observer reliability

is not appropriate in this phase of developing the framework. However, discussions did take place between the first and second author(s) regarding the segmentation and assigned codes of all data.

Example 1: A Musical Composition Task Participants

The data of the first case study were selected from a larger study on teacher-student and peer interactions during a musical composition task in primary education (Kupers, 2013). From the six teacher-student dyads, interactions that were dominated by the teacher were considered not suitable, as sufficient student actions and utterances are required in order to fully illustrate the application of the coding framework. One of the dyads (“John” and his teacher) was selected as a case study for the current article. The video of this particular dyad gave the overall impression that the student was very much an active participant in the creative process, which is why we picked this case as most appropriate for illustrating the coding procedure. The student, a Dutch boy (native Dutch-speaking), was 9 years old at the time of data collection. The teacher was an undergraduate student in music education, who was doing a teaching internship at the student’s school. The teacher and the student’s parents gave their consent for participating in the study—in line with the guidelines of the Ethical Committee of the University of Groningen, department of Pedagogy and Educational Sciences.

Task and Procedure

The assignment was to compose a short musical composition on the basis of a scene from a movie or book, using composition software. The students first had a short introduction class in which the role of music in telling a story was explained (illustrated by scenes from a Harry Potter movie), and in which the basis of the musical composition software (Magix Music Maker) was discussed. Furthermore, the teachers attended a short workshop about the basics of teaching for creativity, after which they had the opportunity to practice using the composition software. After this introduction, the student and teacher worked on the task for 30 min (using a laptop), in a room separate from the normal classroom. The software, Magix Music Maker, works with an extensive library of “loops”: short fragments of music, beats or sounds that can be selected and dragged onto a “canvas,” where the loops can be put together and edited (for instance, adjusting the dynamics or length of the loop) in order to compose a piece of music. Two video cameras were installed to record the composition process: one in front of them (facing the teacher and student) and one behind them (recording the actions on the computer screen). Participants were aware that they were being filmed. Afterwards, the spoken language was transcribed (at the level of interpretation) in F4 (transcription software), then exported to Excel where descriptions for non-verbal behavior were added with time stamps. We converted the time to timepoints of half seconds (meaning time point 10 occurs 5 s after the start of the video). These turns were then coded by using our coding frame. Both the segmentation of the data into turns (step 1)

and the coding of the turns (steps 2 and 3) were extensively discussed by the first author (who coded the data) and the co-authors.

Application of the Coding Scale

Step 1: Determine the unit of analysis

In this context—a student working on a musical composition task, supported by a teacher—we chose to only code the student turns for novelty and appropriateness (teacher turns could still be coded on other dimensions at a later point; see “Further analyses”). A turn could be either verbal, non-verbal or a combination of both, because the task entails both constructing something (a product) as well as reflecting on the actions and thinking out loud. For verbal units, each time the student made a remark, answered a question, etcetera, this was defined as a turn. In this case, turns are more suitable as units of analysis than utterances, because answers, ideas, and elaborations often encompass multiple utterances. Non-verbal turns were defined as “meaningful actions,” in the sense that they were part of the creative process, compared to merely procedural ones (e.g., saving the document, restarting the program after an error). In this context, examples of meaningful non-verbal turns were playing and selecting a loop, adjusting the volume or length of loops that were already in the composition, deleting parts of the composition, and playing back a composed piece of music. If a meaningful action was accompanied by a verbal turn (e.g., saying “I’ll put this at the beginning” while dragging a loop to the beginning of the piece), they were coded together as one turn since the action and verbal turn together make up one meaningful unit. If the student voiced a new general idea that took multiple actions to execute, these “minor actions” were coded as one turn (e.g., saying “I’ll make all of these very loud” and then adjusting the volume of multiple loops). Verbal turns and actions that referred to technical errors of the software or that were strictly procedural (e.g., “This loop doesn’t work,” “How do you adjust the volume?”) were excluded from the analysis. In cases of doubt, the segmentation of turns was discussed by the authors. This procedure resulted in 68 turns in the first 10 min of the assignment, which were then coded for novelty and appropriateness.

Step 2: Code each unit for novelty

The next step is to code all turns by dividing them into one of the four levels of novelty. In **Table 3** below, the first part of the coded transcript is presented, accompanied with explanations for each given code.

Step 3: Code each unit for appropriateness

Step 2 was repeated, only now coding all turns for appropriateness. All turns were coded on high levels of appropriateness (level 2 and 3), meaning the student was engaged in the task during the entire fragment. Since the assignment in this case was to compose a piece of music to go with a scene from a story, level 3 was assigned when the student verbally made a link between elements of the story (events occurring in the scene, the atmosphere of the scene, etcetera) and

TABLE 3 | First 17 turns (9 student turns) of the musical composition task coded for novelty, translated to English by the first author.

Turn	Transcript	Level novelty	Motivation
1	<i>T: Do you have a scene in mind?</i>		
2	S: A scene, no not really.	0	Student doesn't come up with an idea yet.
3	<i>T: No?</i>		
4	S: Uh, it's hard. A scene from a movie...	0	Student doesn't come up with an idea yet.
5	<i>T: Or a scene from a book.</i>		
6	S: That's also possible. Yes, uh, well... The Grey Hunter.	3	The book “The Grey Hunter” is a new idea, it has not been mentioned before.
7	<i>T: The Grey Hunter.</i>		
8	S: Yes, that's very exciting.	1	The term “exciting” adds a small part of new information about the book; it refers to a nuance in the atmosphere of the book.
9.	<i>T: Very exciting, okay. Do you know something, anything that happened [in the book] or something?</i>		
10.	S: Yes, they lit a bridge on fire and then they fell off the bridge.	2	The student now goes from the general idea of the book to describing one specific scene in some detail. This is a large change compared to the previous general talk about the book.
11.	<i>T: Ah, well. What kind of music would go with that?</i>		
12.	S: Something with guitar, I think.	3	The idea that guitar could go well with the scene from the book is new.
13.	<i>T: Okay.</i>		
14.	S: And drums.	3	The idea of adding drums is new.
15.	<i>T: Go ahead.</i>		
16.	S: Guitar... [plays a guitar loop]	1	Small elaboration (trying one specific guitar loop) on the previous general idea.
17.	S: [plays another guitar loop]	1	Small elaboration (trying another specific guitar loop) on the previous general idea.

S, Student; T, Teacher. Teacher turns are in italics. [...] marks excluded, procedural turns.

the music. Level 2 was assigned when the student was working on-task, but without explicit referral to the task.

Example 2: A Problem-Solving Task About Air Pressure Participants

The data of the second case study came from a larger study on peer interaction and scientific reasoning (Guevara et al., 2016), for which permission of the Ethical Committee of

Psychology was received (ppo-011-128). For the current case we chose the interaction between one 6-year-old girl (who we will refer to as “Sarah”) and a researcher. Although Sarah was living in the Netherlands, she went to an international school and her native language was English. Therefore, the experiment was conducted in English. The researcher was a trained PhD-student. This specific case study was selected for solely pragmatic reasons: the researcher and the parents of the child gave informed consent for the use of this video, and the recording was of good technical quality.

Task and Procedure

The task consisted of a set of tubes and syringes that had to be connected to each other in order to reach a certain goal (on one of the syringes, the plunger had get to a red mark). The overarching theme of the task was the understanding of air pressure. The syringes had different sizes and the tubes had different shapes. The child was asked to use the materials (connect materials and push the syringes) to reach the goal, and also to describe, predict and explain what was happening. The task consisted of a sequence of steps that introduced different elements. In this example, we used three elements: a first in which two equally sized syringes had to be connected, a second in which one small and one large syringe were used, and a third in which a Y-shaped tube had to be connected to two syringes. In total, this part of the task took roughly 4 min. The experiment was video-recorded, and participants were aware that they were being filmed. All spoken language and any actions involving the materials were transcribed in Excel, in which descriptions of those actions were added with time stamps (manually). It should be noted that any spoken language was transcribed at the interpretation level (meaning we corrected for grammatical errors, false starts, unintelligible parts, etcetera.).

Application of the Coding Framework

Step 1: Determine the unit of analysis

The child's utterances and actions were considered the units of analysis, and coded for novelty and appropriateness. Considering the units of analysis, we followed CHILDES guidelines with regard to determining utterance boundaries (MacWhinney, 2000). The reason for using utterances as verbal units instead of turns is that, in this task, the aim was to form an understanding of the principles of air pressure, and separated utterances might already contain information about this understanding without being a completely formed “idea.” Actions were separated on the basis of meaningful chunks of movements: (attempts at) pushing/pulling the plunger (attempts at), connecting two elements with each other, turning an object in another direction, pointing toward an object, blowing into a tube, etcetera. In this context, meaningful units were any manipulations of the materials, such as connecting or disconnecting syringes, pushing or pulling plungers. All verbal utterances were also considered meaningful. Unintelligible language and giggling were considered not meaningful.

TABLE 4 | First 15 turns of transcripts and codes for the problem-solving task.

Turn	Transcript	Level novelty	Motivation
1	<i>T: So, you connected the two syringes. Can you explain how we can make this red mark inside go to here?</i>		
2	S: You do like this... [pulls]	3	Student refers to pulling, which is new.
3	<i>T: Yes, what will happen?</i>		
4	S: When you push down, the wind will go to there, and then to there. (points finger).	3	Student mentions pushing and wind moving through the tube, which are both new
5	<i>T: The wind will go to there to here?</i>		
6	S: And then it will blow up.	1	Small elaboration because blowing up is the target of the task, she mentions the consequence of the action described above.
7	<i>T: Blow up? To where?</i>		
8	S: To here. (points at red mark)	1	Small elaboration because she just adds exact location, which was the task assignment.
9.	<i>T: Now we will try it. One, two, three.</i>		
10.	S: (Pushes syringe to the red mark)	0	No new elements because pushing was mentioned.
11.	<i>T: Great job. Can you explain to me what happened?</i>		
12.	S: It went up.	0	No new elements.
13.	<i>T: Yes exactly, and what made that happen? What made it go to the red mark?</i>		
14.	S: Because the wind was... the wind was blowing that up.	0	No new elements, “wind” and “blow up” where already mentioned before.
15.	<i>T: Okay, because the wind blew this up. Great.</i>		
16.	S: I like this game.	3	Affective aspect of task was not mentioned before and is therefore new, but is slightly less appropriate in this task context.
17.	S: (pushes the syringe)	0	Repetition of action

S, Student; T, Teacher. Teacher turns are in italics. (...) marks non-verbal turns.

Step 2: Code each unit for novelty

Next, all turns were coded on one of the four levels of novelty. In **Table 4** below, the first part of the coded transcript is presented, along with explanations for each given code.

Step 3: Code each unit for appropriateness

In the case of Sarah, all units were highly appropriate. The child was clearly immersed with the task and all actions and verbalizations in the 4-min fragment were related to task

elements. There were several slight deviations, in which Sarah expressed that she liked the task. As these utterances did not contain a description, prediction, or explanation, and did not refer to specific task elements either, we decided to score them at level 2 on the appropriateness scale.

POSSIBILITIES FOR DATA ANALYSIS

Time Series

A first inspection of the data can be obtained by plotting the levels of novelty and appropriateness over time. In **Figure 2**, the levels of novelty in the case of John were plotted over time. Looking at this graph, we can see that John frequently switches between different novelty levels. Both at the beginning of the assignment (between time point 108 and 330) and later on (time point 856 to 1077), we see an episode where a new idea (level 3) is followed by a dense series of small elaborations (level 1). This seems to be quite characteristic for this student working on this task. The time series of Sarah's novelty levels is shown in **Figure 3**. Here it is clearly visible that Sarah also frequently switches between different novelty levels. In Sarah's task behavior, we also observe that high levels of novelty occur across the session, and often alternate with less novel ideas or actions. Between turn 40 and turn 60 there seems to be a temporal "dip" in her creative behavior, with many repetitions of the same idea. After time point 60, actions and ideas with a relatively high level of novelty re-emerge (and when observing the video, it becomes clear that

point 60 is exactly when a new task element is introduced, in the form of the Y-shaped tube).

State Space Grids of Novelty and Appropriateness

After having coded the two dimensions of creativity, we can now combine them. One particularly useful technique when analyzing the interactions between both dimensions is the *State Space Grid (SSG) method* (Hollenstein, 2013). This technique is based on the idea that combinations of behaviors can be described in terms of their movements across the range of all behavioral possibilities for a given system. The data are described according to two ordinal variables that define the behavior of interest—in this case the novelty and the appropriateness dimension. The child's actions are seen as a *state* of the system, and they are represented by dots. Consequently, all *movements* between states are presented by lines. The advantage of using SSGs is not only that it offers a powerful visual analysis of the behavior in qualitative terms, but also that the software computes a set of *measures* that express the global flexibility or stability of the child's repertoire as shown in a specific task setting (for an example of this, see "Further analyses").

Figure 4 displays the SSG of novelty and appropriateness for John, with appropriateness on the vertical axis and novelty on the horizontal axis. Each node represents one unit or event. Since we coded units as discrete events (due to our choice to code

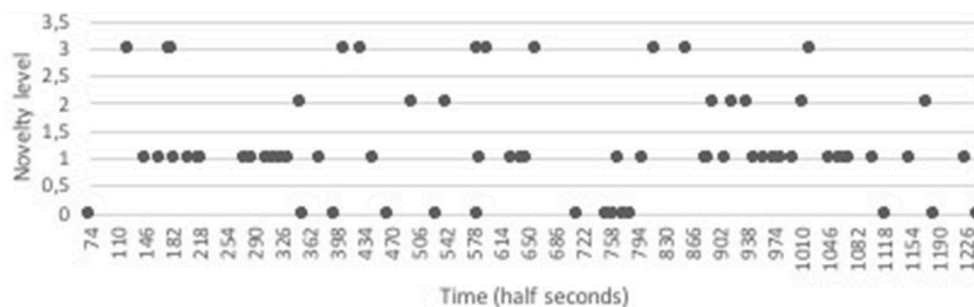


FIGURE 2 | Time series of Novelty levels for John.

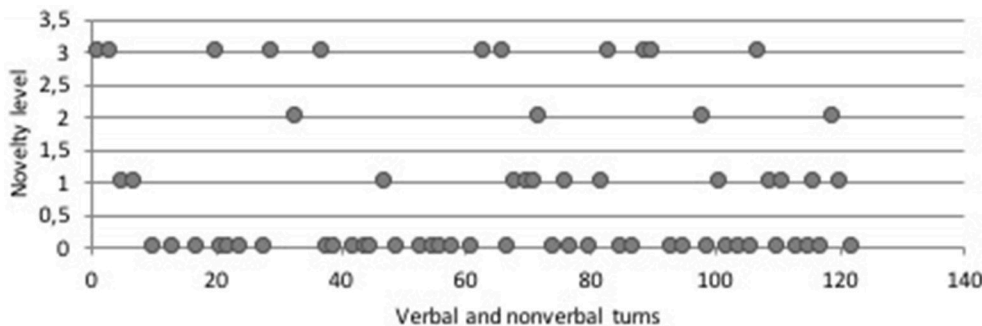


FIGURE 3 | Time series of Novelty levels for Sarah.

utterances, turns and short-lived actions), the duration is not taken into account and all nodes are therefore the same size. For instance, a node in the bottom left corner represents an event that was both low in novelty and low in appropriateness (for the sake of visibility, the exact locations of nodes within a cell are random). The open node represents the first event. Overall, we see in **Figure 4** that there is a cluster of ideas with a lower level of novelty (small elaborations) and a high level (2) of appropriateness. This corresponds with the observation that John is frequently engaged in a series of small elaborations on the same overarching idea, while staying focused on the task (for instance, the novel idea that a guitar should be added to the piece is followed by John trying out many different guitar loops before selecting one).

It can be observed in **Figure 5** (the SSG for Sarah) that there are many changes in the novelty dimension of the scale, with actions and ideas constantly moving from left to right and back. Appropriateness is much less variable in that regard. However, three out of four instances that show that the child drops slightly in appropriateness occur when novelty is also low.

So far, we have seen in both illustrations that the assignment of the codes could be done in a relatively straightforward way. The resulting analyses showed that in each case the actions and verbalizations were quite variable regarding the novelty dimension. We also observed in both cases that novel elements were often introduced by the child after repeated turns with no or low levels of novelty. A clearly observable difference between the tasks was that, in comparison with the syringes task, the music task elicited more behaviors that were variations on previous actions than the syringes task. An open-ended task like a music composition might involve more new “big ideas” that are then further developed through elaborations on those ideas. Although in both cases the task led to many highly appropriate actions and verbalizations, the music task elicited more level 2 actions and fewer level 3 actions than the syringes task. The reason for this is that the music assignment is more demanding, as it involves relating music to a storyline—compared to simply connecting task elements, such as in the syringes task.

State Space Grids of Novelty and Teacher Behavior

In order to further demonstrate the potential of the micro-developmental measure of creativity, we will show how it can be combined with another real-time variable. In the context of this study, we could for instance relate it to the utterances of the teacher—which is what we will do in the next example. In this case, we chose to code each teacher utterance into one of the following categories: Instruction, Feedback, Information, Repetition, Closed Question, Open Question, Encouragement, None, or Other. The categories were ordered according to the underlying dimension of how much room the teacher leaves for student initiative in each case (based on previously developed and validated scales of “openness” Meindertsma, 2014 and “autonomy support” Kupers, 2014). For instance, giving

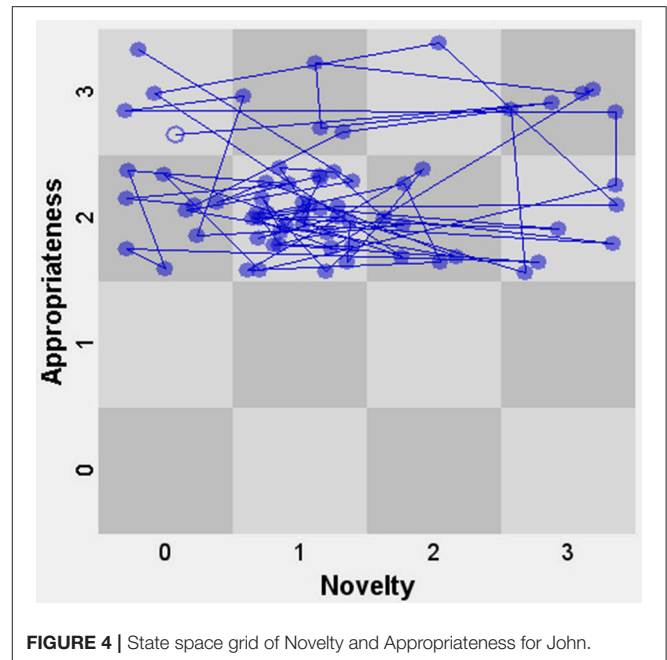
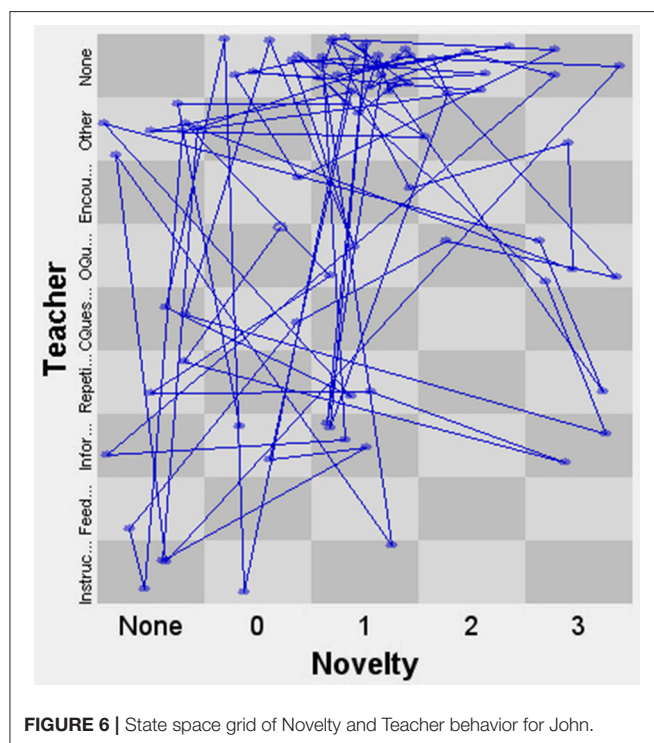
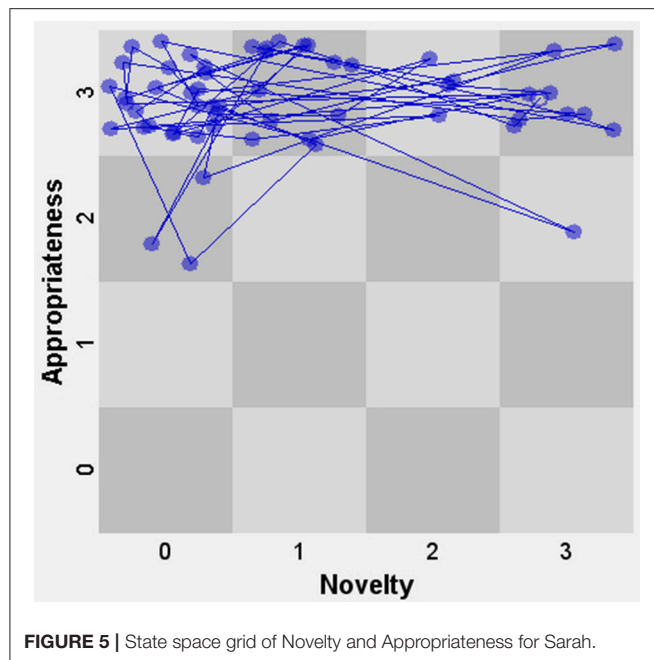


FIGURE 4 | State space grid of Novelty and Appropriateness for John.

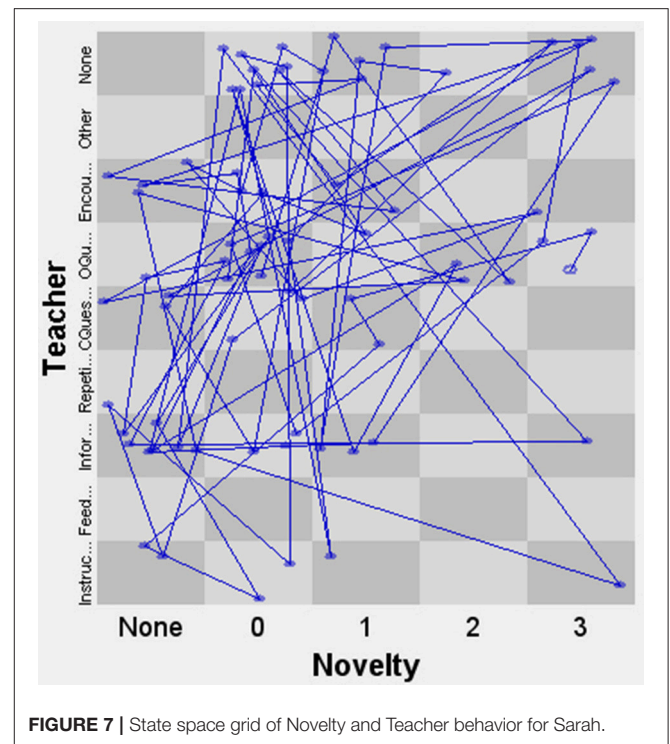
directive instructions leaves less room for student initiative than asking an open-ended question or providing encouragement. The advantage of linking these two variables is that it is possible to use the SSG-technique to plot the interactions on both dimensions. We show the application of this in **Figures 6, 7** (these state space grids represent, respectively, the interactions of John and Sarah during the music and linked syringes tasks described above). Each blue dot in the graph represents a teacher utterance followed by a student action. For instance, a dot in the bottom left corner means that the teacher gave an instruction that was followed by a student turn with the lowest level of novelty (0). The last category on both axes, None, indicates a student turn that was not preceded by a teacher utterance (in other words, a student self-iteration) or vice versa (a teacher self-iteration). In this way, we can analyze which teacher utterances are followed by higher or lower levels of student novelty. Furthermore, we can also inspect whether the dyadic interaction is characterized by strong attractor states or high variability over time and across states. These applications are similar to what is used in Menninga et al. (2017) and van Vondel et al. (2016), but in this case they feature a measure of the student’s creativity instead of the student’s level of cognitive performance.

In **Figure 6**, we see that although the data of John are scattered broadly across the grid, most data points are in the top half of the grid. The amount of student self-iterations shows that quite often there are sequences where the student proposes ideas without a specific prompt from the teacher, indicating that the creative process is—at times, at least—student-led. We also see teacher utterances with a high level of openness, which lead to student turns with varying degrees of novelty. As for the interaction between Sarah and her teacher in **Figure 7**, we observe a lot of variability in both teacher behavior and the novelty of student



responses. Though most interactions lead to relatively low-novelty responses, the responses that are high in novelty seem either to be preceded by an open question or to be a self-iteration.

In both cases, there are no clear attractor states in the interaction dynamics between the children and the teachers. The quantitative measurements show a high level of variability over time, especially for Sarah. Dispersion, which can vary between 0



(all events in one state) and 1 (all events spread out evenly over the grid), was 0.91 for John and 0.96 for Sarah—suggesting that Sarah's interactions may be slightly more variable than John's.

This illustration shows that combining the micro-developmental coding of creativity with a second variable (such as teachers' verbalizations) offers many possibilities for analyzing their dynamic interactions on a more advanced level. These interactions can be visually inspected and quantified by means of the measures offered by the technique. These measures can be used to make, for instance, a comparison between different teacher-student dyads working on the same task, or on different (versions of) tasks, etc. Another option is to analyze the interactions between peers as they work together on a task, investigating how creative behaviors emerge from their collaboration.

GENERAL DISCUSSION

Creativity research has flourished in the last decades. When it comes to empirical research, creativity is mostly measured either at the level of the person (by means of divergent or creative thinking tests) or at the level of the product (by means of consensual assessment) (Kupers et al., submitted). Studies on creative processes are usually purely qualitative. These qualitative studies provide thorough descriptions of creative processes in a certain domain (such as dance, music, or scientific research), but their domain specificity makes it hard to generalize their findings to other contexts. For this reason, our aim was to develop a quantitative measure of creativity that on the one hand is focused on measuring creativity in the here-and-now

of the creative process, and on the other hand is systematic and generic in the sense that it can be applied to many different contexts. We have illustrated the potential of this micro-developmental measure by applying it to an open-ended musical composition task as well as a closed-ended scientific reasoning task.

The framework we propose is rooted in the sociocultural tradition of studying creativity most prominently represented by Csikszentmihalyi and his Systems model of Creativity (Csikszentmihalyi, 1988), and since then developed by Sawyer (1999, 2007) and Glăveanu (2010a,b, 2014), amongst others. The micro-developmental nature of the framework allows us to zoom in on the *interaction* between the person being studied, other persons, and the creative product or task (Glăveanu, 2013b). This is in line with a recent movement within psychology—originating from cultural, ecological perspectives and Complex Dynamical Systems approaches—toward reinterpreting psychological constructs as dynamic, embodied, embedded and enacted (Granic, 2005; Lichtwarck-Aschoff et al., 2008; Rowlands, 2010; Borsboom and Cramer, 2013; de Ruiter et al., 2017). These new theoretical developments ask for new approaches to measuring creativity as well, and our instrument forms an important step in further developing these ideas. Central to a process approach to creativity is the idea that creativity emerges from moment to moment in interaction between a person, the immediate social environment (teachers, peers etc.), and the task (Glăveanu, 2013b; Kupers et al., submitted). However, one domain that remains relatively unattended in creativity research is that of the task. From a dynamic, enactment perspective, any task has certain affordances. Affordances are characteristics of the task that provide opportunities in the interaction with that task (Gibson, 1977; Withagen et al., 2012, 2017). For instance, a task that requires children to copy a drawing by their teacher provides very little opportunity for students to come up with their own ideas, while the assignment to design and draw your own dream house gives students much more opportunities to come up with new ideas. With the framework we present in this article, it is possible to look in detail at creative affordances of different kinds of tasks in many different settings.

Our coding framework is based on the two core components of creativity: novelty and appropriateness. Although the importance of both elements is underlined theoretically (e.g., Cropley, 2006), psychological tests of creativity usually only assess “divergent thinking,” which is basically the ability of a person to come up with many (fluency) ideas that are original (novelty) and unrelated to each other (flexibility). The more novel, unrelated and appropriate an idea is, the greater is its underlying trait of creative thinking—that is the assumption of these tests. The question is whether “more is better” also applies to measuring the creative process, which our coding frame aims to address. Is a creative process more “successful” if it features more ideas with the highest level of novelty, given a high level of appropriateness? More research is necessary—especially on the micro level—to unravel the ways in which appropriateness and novelty interact from moment to moment.

LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

This paper presents a general framework for the coding of creativity on the level of micro-development, in the interaction between a person, task, and the direct social environment. While an advantage of the proposed method (and an aim of the authors) was that the instrument is applicable to many different contexts, this also poses limitations. We have stressed throughout this paper that, for each dataset, the general coding framework presented here needs to be adjusted in order to form an actual coding *scheme*—which involves specific decision rules regarding the segmentation of the data in units of analysis and the coding of those units. Though we have provided an illustration of how to do this in the case of two different creative tasks, this should not be seen as an attempt to validate the method but rather as a demonstration of applying the coding frame to specific data—which is an important first step. Any coding schemes that future researchers construct on the basis of our general coding framework need to be validated on larger datasets, as is generally the case with observational coding schemes.

Important theoretical foundations of this coding framework have been social-constructivist approaches to creativity (Csikszentmihalyi, 1988; Amabile, 1996; Sawyer, 2007; Glăveanu, 2010a). This automatically raises two questions: a. Is the coding scheme only applicable to creativity in social interactions, as demonstrated here? b. When the coding scheme is indeed applied to social interactions, how should the actions of the “other” (in this case, the teacher) be coded? With regard to the first question: the coding scheme can also be applied to individual creativity (of children as well as adults), as long as all the steps in the creative behavior are observable. In order to get a better understanding of individual creativity as an emergent property from real-time interactions, it could be interesting to follow individuals over longer periods of time as they engage in different creative tasks. With regard to the second question: we have provided an example of how teacher behaviors can be coded, but many different options are possible. In peer interactions, it is possible to code novelty and appropriateness of both interaction partners. In teacher-student interactions, one promising construct is “teaching for creativity.” By translating this construct into observable behavior, we can get a direct analysis of which aspects of the theoretical construct actually lead to student creativity.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the ethical guidelines of the University of Groningen; respectively, the Ethical Committee of Pedagogical and Educational Sciences and the Ethical Committee of Psychology. The protocols were approved by the Ethical Committee of Pedagogical and Educational Sciences and the Ethical Committee of Psychology (University of Groningen). All (parents of) subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

EK designed method, discussed application of method with MVD and AL-W, wrote first draft of article (and took the lead in further drafts), collected and analyzed data for

the first case study. MVD co-designed method, wrote parts of the first and second draft, gave comments on several drafts, analyzed data for the second case study. AL-W co-designed method, gave comments on analyses data and several drafts.

REFERENCES

- Amabile, T. (1996). *Creativity in Context*. Boulder: Westview Press.
- Amabile, T. M. (1983). The social psychology of creativity: a componential conceptualization. *Pers. Process. Individ. Differ.* 45, 357–376.
- Baas, M., De Dreu, C. K. W., and Nijstad, B. A. (2008). A meta-analysis of 25 years of mood-creativity research: hedonic tone, activation, or regulatory focus? *Psychol. Bull.* 134, 779–806. doi: 10.1037/a0012815
- Baer, J., and Kaufman, J. C. (2008). Gender differences in creativity. *J. Creat. Behav.* 42, 75–105. doi: 10.1002/j.2162-6057.2008.tb01289.x
- Borsboom, D., and Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annu. Rev. Clin. Psychol.* 9, 91–121. doi: 10.1146/annurev-clinpsy-050212-185608
- Burnard, P., and Younker, B. A. (2004). Problem-solving and creativity: insights from students' individual composing pathways. *Int. J. Music Educ.* 22, 59–76. doi: 10.1177/0255761404042375
- Butcher, J. L., and Niece, L. N. (2005). Disruptive behaviors and creativity in childhood: the importance of affect regulation. *Creativity Res. J.* 17, 181–193. doi: 10.1080/10400419.2005.9651478
- Chappell, K., and Craft, A. (2011). Creative learning conversations: producing living dialogic spaces. *Educational Research.* 53, 363–385. doi: 10.1080/00131881.2011.598663
- Chen, Z., and Siegler, R. S. (2000). "Intellectual development in childhood," in *Handbook of intelligence*, ed R. J. Sternberg (New York, NY: Cambridge University Press), 92–116. doi: 10.1017/CBO9780511807947.006
- Cheung, P. C., and Lau, S. (2010). Gender differences in the creativity of hong kong school children: comparison by using the new electronic Wallace-Kogan creativity tests. *Creativity Res. J.* 22, 194–199. doi: 10.1080/10400419.2010.481522
- Cropley, A. (2006). In praise of convergent thinking. *Creat. Res. J.* 18, 391–404. doi: 10.1207/s15326934crj1803
- Csikszentmihalyi, M. (1988). "Society, culture, and person: a systems view of creativity," in *Conceptions of Giftedness*, eds R. Sternberg and J. E. Davidson. (New York, NY: Cambridge University Press), 325–339.
- de Ruiter, N. M. P., van Geert, P. L. C., and Kunnen, S. (2017). Explaining the "How" of self-esteem development: the self-organizing self-esteem model. *Rev. Gen. Psychol.* 21, 49–68. doi: 10.1037/gpr0000099
- de Weerth, C., van Geert, P., and Hoijtink, H. (1999). Intraindividual variability in infant behavior. *Dev. Psychol.* 35, 1102–1112. doi: 10.1037/0012-1649.35.4.1102
- Dziedzieciewicz, D., Gajda, A., and Karwowski, M. (2014). Developing children's intercultural competence and creativity. *Thinking Skills Creativ.* 13, 32–42. doi: 10.1016/j.tsc.2014.02.006
- Fernández-Cárdenas, J. M. (2008). The situated aspect of creativity in communicative events: how do children design web pages together? *Thinking Skills Creativ.* 3, 203–216. doi: 10.1016/j.tsc.2008.09.007
- Gibson, J. (1977). "The theory of affordances," in *Perceiving, Acting, and Knowing: Towards an Ecological Psychology*. eds. R. E. Shaw and J. D. (Routledge: Bransford Abingdon-on-Thames).
- Glaser, J., and Laudel, G. (2013). Life with and without coding: two methods for early-stage data analysis in qualitative research aiming at causal explanations. *Forum Qualitative Soc. Res.* 14. Available online at: <http://www.qualitative-research.net/index.php/fqs/article/view/1886/3528>
- Glăveanu, V. P. (2010a). Paradigms in the study of creativity: introducing the perspective of cultural psychology. *New Ideas Psychol.* 28, 79–93. doi: 10.1016/j.newideapsych.2009.07.007
- Glăveanu, V. P. (2010b). Principles for a cultural psychology of creativity. *Culture Psychol.* 16, 147–163. doi: 10.1177/1354067X10361394
- Glăveanu, V. P. (2013a). Creativity development in community contexts: the case of folk art. *Thinking Skills Creativ.* 9, 152–164. doi: 10.1016/j.tsc.2012.11.001
- Glăveanu, V. P. (2013b). Rewriting the language of creativity: the five A's framework. *Rev. Gen. Psychol.* 17, 69–81. doi: 10.1037/a0029528
- Glăveanu, V. P. (2014). *Distributed Creativity: Thinking Outside the Box of the Creative Individual*. Cham: Springer.
- Granic, I. (2005). Timing is everything: developmental psychopathology from a dynamic systems perspective? *Develop. Rev.* 25, 386–407. doi: 10.1016/j.dr.2005.10.005
- Granott, N., Fischer, K. W., and Parziale, J. (2002). "Bridging the unknown: a transition mechanism in learning and development," in *Microdevelopment: Transition Processes in Development and Learning*, eds N. Granott and J. Parziale. (New York, NY: Cambridge University Press), 131–156.
- Granott, N., and Parziale, J. (Eds.). (2002). *Microdevelopment: Transition Processes in Development and Learning*. Cambridge, MA: Cambridge University Press.
- Gregersen, T., Macintyre, P. D., and Meza, M. D. (2014). The motion of emotion: idiodynamic case studies of learners' foreign language anxiety. *Modern Lang. J.* 98, 574–588. doi: 10.1111/j.1540-4781.2014.12084.x
- Guevara, M., Van Dijk, M., and Van Geert, P. (2016). Microdevelopment of peer interactions and scientific reasoning in young children. *Infancia y Aprendizaje.* 39:4. doi: 10.1080/02103702.2016.1215083
- Guevara-Guerro, M. (2015). *Peer Interaction and Scientific Reasoning in Preschoolers: An Intra-Individual Approach*. Unpublished doctoral dissertation. University of Groningen, Groningen, the Netherlands.
- Guilford, J. P. (1957). Creative abilities in the arts. *Psychol. Rev.* 64, 110–118.
- Guilford, J. P. (1967a). Creativity: yesterday, today and tomorrow. *J. Creat. Behav.* 1, 3–14. doi: 10.1002/j.2162-6057.1967.tb00002.x
- Guilford, J. P. (1967b). *The Nature of Human Intelligence*. New York, NY: McGraw-Hill.
- Healey, D., and Rucklidge, J. J. (2006). An investigation into the relationship among ADHD symptomatology, creativity, and neuropsychological functioning in children. *Child Neuropsychol.* 12, 421–438. doi: 10.1080/09297040600806086
- Hollenstein, T. (2013). *State Space Grids: Depicting Dynamics Across Development*. New York, NY: Springer Science + Business Media.
- Hu, W., Wu, B., Jia, X., Yi, X., Duan, C., Meyer, W., and Kaufman, J. C. (2013). Increasing students' scientific creativity: the "Learn to Think" intervention program. *J. Creative Behav.* 47, 3–21. doi: 10.1002/jocb.20
- Kaufman, J. C., Evans, M. L., and Baer, J. (2010). The American Idol Effect: are students good judges of their creativity across domains? *Emp. Stud. Arts* 28, 3–17. doi: 10.2190/EM.28.1.b
- Kim, K. H. (2011). The creativity crisis: the decrease in creative thinking scores on the torrance tests of creative thinking. *Creat. Res. J.* 23, 285–295. doi: 10.1080/10400419.2011.627805
- Kim, K. H., and VanTassel-Baska, J. (2010). The relationship between creativity and behavior problems among underachieving elementary and high school students. *Creat. Res. J.* 22, 185–193. doi: 10.1080/10400419.2010.481518
- Kupers, E. (2013). *Creativity in Children's Compositions: A Microgenetic Approach*. Unpublished research proposal, Universität Bremen, Bremen, Germany.
- Kupers, E. (2014). *Socially Situated Learning in Individual Music Lessons*. University of Groningen.
- Lavelli, M., and Fogel, A. (2002). Developmental changes in mother-infant face-to-face communication: birth to 3 months. *Dev. Psychol.* 38, 288–305. doi: 10.1037/0012-1649.38.2.288
- Lavelli, M., Pantoja, A. P. F., Hsu, H.-C., Messinger, D., and Fogel, A. (2005). "Using microgenetic designs to study change processes," in *Handbook of Research Methods in Developmental Psychology*, eds D. G. Teti (Hoboken, NJ: John Wiley and Sons), 1–37.
- Lichtwarck-Aschoff, A., van Geert, P., Bosma, H., and Kunnen, S. (2008). Time and identity: a framework for research and theory formation. *Develop. Rev.* 28, 370–400. doi: 10.1016/j.dr.2008.04.001

- Long, H. (2014). An empirical review of research methodologies and methods in creativity studies (2003-2012). *Creativity Res. J.* 26, 427–438. doi: 10.1080/10400419.2014.961781
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk: Transcription Format and Programs, Vol 1, 3rd Ed.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers. Available online at <http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=psyh&AN=2000-03630-000&site=ehost-live&scope=site>
- Meindertsma, H. B. (2014). *Predictions and Explanations*. University of Groningen.
- Menninga, A., van Dijk, M., Steenbeek, H., and van Geert, P. (2017). Language use in real-time interactions during early elementary science lessons: the bidirectional dynamics of the language complexity of teachers and students. *Lang. Learn.* 67, 284–320. doi: 10.1111/lang.12219
- Miell, D., and Macdonald, R. (2000). Children's creative collaborations: the importance of friendship when working together on a musical composition. *Soc. Dev.* 9, 346–369. doi: 10.1111/1467-9507.00130
- Odean, R., Nazareth, A., and Pruden, S. M. (2015). Novel methodology to examine cognitive and experiential factors in language development: combining eye-tracking and LENA technology. *Front. Psychol.* 6:1266. doi: 10.3389/fpsyg.2015.01266
- Patera, M., Draper, S., and Naef, M. (2008). Exploring *Magic Cottage*: a virtual reality environment for stimulating children's imaginative writing. *Interact. Learn. Environ.* 16, 245–263. doi: 10.1080/10494820802114093
- Plucker, J. A., Beghetto, R. A., and Dow, G. T. (2004). Why Isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educ. Psychol.* 39, 83–96. doi: 10.1207/s15326985ep3902_1
- Putwain, D. W., Kearsley, R., and Symes, W. (2012). Do creativity self-beliefs predict literacy achievement and motivation? *Learn. Individ. Differ.* 22, 370–374. doi: 10.1016/j.lindif.2011.12.001
- Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan* 42, 305–310. Available online at: <https://www.jstor.org/stable/20342603>
- Rowlands, M. (2010). *The New Science of the Mind*. Cambridge, MA: The MIT Press.
- Runco, M. (2010). "Divergent thinking, creativity, and ideation," in *The Cambridge Handbook of Creativity*. eds. J. C. Kaufman and R. J. Sternberg (Cambridge: Cambridge University Press).
- Runco, M. A., and Jaeger, G. J. (2012). The standard definition of creativity. *Creativ. Res. J.* 37–41. doi: 10.1080/10400419.2012.650092
- Runco, M. A., Plucker, J. A., and Lim, W. (2001). Development and psychometric integrity of a measure of ideational behavior. *Creativ. Res. J.* 13, 393–400. doi: 10.1207/S15326934CRJ1334
- Sawyer, R. K. (1999). The emergence of creativity. *Philos. Psychol.* 12, 447–469.
- Sawyer, R. K. (2007). *Group Genius: The Creative Power of Group Collaboration*. New York, NY: Basic Books.
- Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Ann. Rev. Clin. Psychol.* 4, 1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415
- Siegler, R. S. (2002). "Microgenetic studies of self-explanation," in *Microdevelopment; Transition Processes in Development and Learning*, eds N. Granott and J. Parziale (New York, NY: Cambridge University Press), 31–58. doi: 10.1017/CBO9780511489709.002
- Siegler, R. S., and Chen, Z. (1998). Developmental differences in rule learning: a microgenetic analysis. *Cogn. Psychol.* 36, 273–310. doi: 10.1006/cogp.1998.0686
- Siegler, R. S., and Crowley, K. (1991). The microgenetic method. *Am. Psychol.* 46, 606–620. doi: 10.1037/0003-066X.46.6.606
- Silvia, P. J., Wigert, B., Reiter-Palmon, R., and Kaufman, J. C. (2012). Assessing creativity with self-report scales: a review and empirical evaluation assessing creativity with self-report scales: a review and empirical evaluation. *Psychol. Aesthet. Creativ. Arts.* 6, 19–34. doi: 10.1037/a0024071
- Simonton, D. K. (1999). Creativity as blind variation and selective retention: is the creative process darwinian? *Psychol. Inq.* 10, 309–328.
- Simonton, D. K. (2015). On praising convergent thinking: creativity as blind variation and selective retention. *Creat. Res. J.* 27, 262–270. doi: 10.1207/s15327965PLI1004_4
- Simonton, K. (1999). *Origins of Genius: Darwinian Perspectives on Creativity*. New York, NY: Oxford University Press.
- Sternberg, R. J. (2006). The nature of creativity. *Creat. Res. J.* 18, 87–98. doi: 10.1207/s15326934crj1801
- Strijbos, J. W., Martens, R., Prins, F., Jochems, W., Strijbos, J. W., et al. (2006). Content analysis: what are they talking about? *Comp. Educ.* 46, 29–48. doi: 10.1016/j.compedu.2005.04.002
- Sun, H., Steinkrauss, R., van der Steen, S., Cox, R., and de Bot, K. (2016). Foreign language learning as a complex dynamic process: a microgenetic case study of a Chinese child's English learning trajectory. *Learn. Individ. Differ.* 49, 287–296. doi: 10.1016/j.lindif.2016.05.010
- Tafti, M. A., Hameedy, M. A., and Baghal, N. M. (2009). Dyslexia, a deficit or a difference: comparing the creativity and memory skills of dyslexic and nondyslexic students in Iran. *Soc. Behav. Pers.* 37, 1009–1016. doi: 10.2224/sbp.2009.37.8.1009
- Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking - Norms - Technical Manual Research Edition - Verbal Tests. Forms A and B- Figural Tests*. Princeton, NJ: Personell Press.
- van Vondel, S., Steenbeek, H., van Dijk, M., and van Geert, P. (2016). "Looking at educational interventions: surplus value of a complex dynamic systems approach to study the effectiveness of a science and technology educational intervention," in *Complex Dynamical Systems in Education: Concepts, Methods and Applications*. eds. M. Koopmans and D. Stamovlasis New York, NY: Springer.
- Vass, E. (2007). Exploring processes of collaborative creativity-The role of emotions in children's joint creative writing. *Thinking Skills Creativ.* 2, 107–117. doi: 10.1016/j.tsc.2007.06.001
- Vickhoff, B., Malmgren, H., Åström, R., Nyberg, G. F., Ekström, S. R., et al. (2013). Erratum: music structure determines heart rate variability of singers. *Front. Psychol.* 4:599. doi: 10.3389/fpsyg.2013.00599
- Welch, G. F., and McPherson, G. E. (2012). "Introduction and commentary: music education and the role of music in people's lives," in *Oxford Handbook of Music Education*, eds G. E. McPherson and F. H. Welch (New York, NY: Oxford University Press), 5–20.
- Withagen, R., Araújo, D., and de Poel, H. J. (2017). Inviting affordances and agency. *New Ideas Psychol.* 45, 11–18. doi: 10.1016/j.newideapsych.2016.12.002
- Withagen, R., de Poel, H. J., Araújo, D., and Pepping, G. J. (2012). Affordances can invite behavior: reconsidering the relationship between affordances and agency. *New Ideas Psychol.* 30, 250–258. doi: 10.1016/j.newideapsych.2011.12.003
- Yoder, P., and Symons, F. (2010). *Observational Measurement of Behavior*. New York, NY: Springer Publishing Co. Available online at <http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=psyh&AN=2010-04202-000&site=ehost-live&scope=site>

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 Kupers, Van Dijk and Lehmann-Wermser. 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.



The Effect of Problem Construction on Team Process and Creativity

Roni Reiter-Palmon* and Vignesh Murugavel

Department of Psychology, University of Nebraska Omaha, Omaha, NE, United States

OPEN ACCESS

Edited by:

Philip A. Fine,
University of Buckingham,
United Kingdom

Reviewed by:

Matthew A. Cronin,
George Mason University,
United States
Sorin Cristian Ionescu,
Politehnica University of Bucharest,
Romania

*Correspondence:

Roni Reiter-Palmon
rreiter-palmon@unomaha.edu

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 27 May 2018

Accepted: 11 October 2018

Published: 05 November 2018

Citation:

Reiter-Palmon R and Murugavel V
(2018) The Effect of Problem
Construction on Team Process
and Creativity.
Front. Psychol. 9:2098.
doi: 10.3389/fpsyg.2018.02098

Although research on the benefits of problem construction within the creative process is expanding, research on team problem construction is limited. This study investigates the cognitive process of problem construction and identification at the team level through an experimental design. Furthermore, this study explores team social processes in relation to problem construction instructions. Using student teams solving a real-world problem, the results of this study revealed that teams that engaged in problem construction and identification generated more original ideas than teams that did not engage in such processes. Moreover, higher satisfaction and lower conflict was observed among groups that engaged in problem construction compared to groups that did not engage in problem construction. These findings highlight the utility of problem construction for teams engaging in creative problem-solving.

Keywords: teams, creativity, creative process, problem construction, creative problem solving

INTRODUCTION

In the last two decades, interest in creativity and innovation has grown tremendously. Creativity and innovation have been suggested as important for organizational performance (Dess and Picken, 2000; Shalley et al., 2004; Mumford and Hunter, 2005). In addition, increased frequency and rapid changes in technology, globalization, and increased competition have all created an environment in which creativity and innovation are necessary for organizational survival (Mumford et al., 2002b; Shalley et al., 2004). Specifically, creativity has been defined in terms of the production of a “novel product, idea, or problem solution that is of value to the individual and/or the larger social group” (Hennessey and Amabile, 2010, p. 572). The implementation of a creativity idea or solution and application of a creative product is referred to as innovation (Amabile, 1988).

Interest and research on team creativity has increased in recent years as a result of the complexity of problems that face organization exceeding the capabilities of any single individual (Shalley et al., 2004; Kozlowski and Bell, 2008; Reiter-Palmon et al., 2012). In the past, team creativity research has focused on evaluating the role of the creative individual as part of the team (Taggar, 2002). This research identified the relationship between specific team variables such as team diversity, team social processes such as conflict, and social cognitive processes such as shared mental models to creativity exhibited by each individual within the team (Hulsheger et al., 2009). However, less research has directly evaluated the factors that influence team creativity as a construct, as opposed to individual creativity within the team context (Reiter-Palmon et al., 2008).

Creative problem solving is an aspect of creativity that has been researched extensively at the individual level (Merrifield et al., 1962; Basadur, 1982; Silverman, 1985; Sternberg, 1988; Mumford et al., 1991; Finke et al., 1992). While the specific phases and stages of these models differ to some extent, all of these models suggest that creative problem solving starts with problem identification

and construction, followed by idea generation, then idea evaluation and selection (Mumford et al., 1991; Reiter-Palmon, 2018). The problem construction process is of particular importance due to the nature of the problems that allow for or require creativity. Problems that allow for creative solutions tend to be novel, ambiguous and ill-defined (Schraw et al., 1995). Ill-defined problems are characterized by multiple possible goals, multiple possible approaches to solving the problem, and multiple possible and acceptable solutions (Mumford et al., 1991; Schraw et al., 1995). Idea generation or brainstorming focuses on the development of ideas or solutions to the problem, and has been the focus of much of the research on creativity (Reiter-Palmon et al., 2008). Finally, ideas are evaluated to determine which of the ideas should be implemented (Mumford et al., 2002a).

TEAM COGNITIVE PROCESSES: PROBLEM CONSTRUCTION

Problem identification and construction refers to the process in which a problem is identified by the problem solver, an ill-defined problem is structured, and the parameters of that problem are defined (Reiter-Palmon and Robinson, 2009). Problem construction allows individuals to develop and provide some structure and direction to an ambiguous, ill-defined problem. At the individual level, creative individuals have been shown to engage in the process more so than their less creative counterparts (Getzels and Csikszentmihalyi, 1975; Rostan, 1994). However, it has been suggested that for most individuals the process of problem construction occurs automatically, and individuals are not aware that they are defining a problem (Mumford et al., 1994). Past research has demonstrated that active engagement in problem construction, through the use of instructions, has increased the creativity of the solutions developed and that the quality and originality of how the problem is constructed is directly related to quality and originality of the solutions generated (Mumford et al., 1994; Reiter-Palmon et al., 1997; Arreola and Reiter-Palmon, 2016). Because problem construction provides structure and allows individuals to manage and organize an ambiguous, ill-defined problem, it is not surprising that problem construction has been found to have a significant effect on creative problem solving (Ma, 2009).

However, research evaluating this process in teams is sparse. Reiter-Palmon et al. (2008) suggested that it is likely that teams focus on discussing solutions rather than discussing various problem constructions. Consequently, individuals are not aware of how they construct the problem, and potential differences in how different individuals within the team understand and define the problem. Further, it has been suggested that conflict regarding solutions may be rooted in differences in how problems are structured and goals are understood (Cronin and Weingart, 2007; Reiter-Palmon et al., 2008). Research indicates that creative teams suffer when problem frameworks vary across team members, and the goal states identified through problem construction cannot be reconciled in a single solution (Cronin and Weingart, 2007; Goh et al., 2013). Cronin and Weingart

(2007) refer to these differences as *representational gap* or rGaps. Teams with larger rGaps tend to have difficulty during problem construction, leading to poor cognitive integration as a team and lower creativity (Weingart et al., 2005). However, research has also suggested that larger rGaps may increase team creativity when teams identify the discrepancies early and use them to communicate about alternative pathways to solving the problem (Weingart et al., 2008). Differences in cognitive representation among group members have also been linked to team processes beyond problem construction. Cronin et al. (2011) found that these differences affect the formation of subgroups within a team, which can lead to potentially negative outcomes such as a decrease in satisfaction or effectiveness.

Research has also supported the notion that individuals rely on education and past experiences when developing an understanding of the problem, and therefore team members may construct problems differently. Leonardi (2011) found that individuals from different departments structured and constructed problems differently; however, they were largely unaware that they had different ways of conceptualizing the problem. Leonardi further found that leaders were especially important in resolving these differences, such that when leaders encouraged teams to discuss problem features they were able to develop a shared framework or construction. This mutually understood structure in turn guided the innovation process. Similarly, Gish and Clausen (2013) found that prior knowledge influenced how individuals within teams constructed problems. These teams also suffered from team conflict and disagreements during idea generation and team members were unaware of these differences in problem constructions. This conflict, in turn, resulted in lowered creativity. However, when additional information that facilitated divergence in problem construction to identify multiple problem definitions was introduced, teams were more effective at generating an innovative solution.

The current limited research on problem construction in teams suggests that differences in how individuals think about the problem are related to conflict, and that when this conflict is not resolved, creativity suffers (Weingart et al., 2005, 2008; Leonardi, 2011; Gish and Clausen, 2013). The studies discussed all imply that team processes such as team conflict, directly influence the creative processes. In addition, the research described above was all conducted in natural settings with no experimental controls. It is therefore difficult to determine whether conflict was a result of differences in problem construction, was the cause of differences, or whether conflict and create processes co-occurred. Other work on social processes and facets of creative problem solving suggest that the social processes of psychological safety and conflict may limit the effectiveness of cognitive processes such as information sharing or information elaboration (Hoever et al., 2012; Qu and Liu, 2017). Similarly, the Motivated Information Processing in Groups (MIP-G) framework suggests that effective and deep information processing or cognition in teams that leads to team creativity will occur when social processes are effective (De Dreu et al., 2011). Supporting evidence to the effect of poor social processes such as conflict and low trust comes from work on team diversity and its effect on team creativity. Leung and Wang (2015) found that poor team social processes, resulting from

team diversity, hinder knowledge sharing and communication, which in turn result in lowered team creativity. Further, Cronin et al. (2011) suggest that for the group to take advantage of different points of view and the richness of information that is available to different individuals, team members must share that information. They further suggest that cognitive integration becomes more difficult when there are different subgroups within the team.

Research and theory to date have focused on the role of social processes and their effect on team cognition or how the two occur concurrently. That is, studies have suggested that poor communication, conflict, low trust and other less effective social processes *resulted in* less effective cognition and therefore reduced creativity. While social processes can have an effect on cognitive processes, the reverse question, of whether cognitive processes, such as problem construction, can have an effect on social process, has not been addressed. As the process of problem construction aims to provide structure to an ambiguous problem, team engagement in the process may facilitate information sharing and discussion, allowing for better communication and sharing of ideas. As the research described above suggests, team members that construct the problem differently may not be aware of these differences (Weingart et al., 2005, 2008; Leonardi, 2011; Gish and Clausen, 2013). We therefore expect that active engagement in the problem construction process may facilitate understanding of the different ways in which team members understand the problem, and therefore can also influence the effectiveness of social processes in teams.

CURRENT STUDY

Before understanding how active engagement in problem construction processes influence team social processes, however, it was important to determine how active engagement was to be manipulated at the team level. At the individual level, this is accomplished by asking the individual to restate the problem in many different ways, prior to solving the problem. At the team level, these instructions could be given to individuals or to the team as a whole. At this point, the theoretical models of individual or team cognition do not specify which approach may be best, or how these approaches may differ (Reiter-Palmon et al., 2008; Reiter-Palmon, 2018). As a result, it is possible that variations in the instructions to teams may influence the effectiveness of such instructions. Past research on instructions (focusing on divergent thinking tasks) has found that specific instructions can result in specific effects such that instructions to generate multiple ideas result in more ideas generated, whereas instructions to generate original ideas result in more original ideas being generated. Specifically, instructions can be given to individuals, facilitating problem construction at the individual level, but may or may not result in team discussion about problem construction. Instructions can be given at the team level, resulting in team discussion, but potentially limiting individual problem construction. Finally, instructions can be focused on both individual and then team

problem construction, potentially maximizing both. Therefore, the first aim of the study was to directly compare three different approaches to manipulate active engagement in team problem construction to determine whether they are equivalent or whether they result in different outcomes related to solution creativity.

The second aim was to determine whether explicitly engaging problem construction prior to solving a problem in a team context would result in increased solution quality and originality, replicating individual level findings. Finally, the third aim of the study was to determine whether engagement in problem construction influenced any team processes, particularly those that relate to conflict and satisfaction. As it has been speculated that difficulties in team social processes such as conflict may arise due to differences in how problems are constructed, and that team members may not realize that they are constructing problems in dissimilar ways, it was expected that instructions to engage in problem construction would result in less conflict and increased satisfaction.

MATERIALS AND METHODS

Participants

The study was conducted using 65 groups. Each group consisted of three individuals who signed up for the study in the same timeslot. If more than three participants were signed up, they were randomly assigned to groups. If only three participants were signed up, they comprised the group. The total number of participants was 195, of which 109 were female (57.1%) and 82 were male (42.9%), with participants not responding. Average age was 22.88 ($sd = 6.26$). Groups were randomly assigned to one of four conditions.

Procedure

In all conditions, groups were presented and asked to solve a real-life problem relevant to students in which a student is having trouble with his current academic and extracurricular workload. Groups are asked to provide a solution to the student about his plans for the upcoming semester. The first condition was a control condition in which the team did not engage in problem construction. The group only provided a solution to the problem. The other three conditions varied on their problem construction manipulations.

Problem Construction Manipulation

As problem construction has not previously been manipulated in a team setting, three different conditions were used. Manipulations differed in the instructions given to the participants, and whether the focus was on individual generation or team generation of problem constructions. The purpose of including the three conditions was to determine whether there were any differences in the effectiveness of these instructions. In the first manipulation of problem construction, participants were asked to generate as many restatements of the problem as they could individually before proceeding to solve the problem as a team. In the second manipulation, participants engaged

in both problem construction and solution generation as a team. Finally, in the third manipulation, participants were instructed to generate as many restatements as they could to the problem individually, then reach consensus on these as a team, and then move on to developing a solution. Once the team completed the solution generation task, participants completed a number of measures including satisfaction with the team process and team outcome, a measure of team conflict, and demographics.

Measures

Team Conflict

Conflict within the groups was measured using Jehn and Mannix (2001) nine-item scale. The scale contains three subscales of intragroup conflict. The first subscale pertains to task conflict (i.e., “How much conflict of ideas is there in your work group?”; $\alpha = 0.94$). The second subscale involves relationship conflict (i.e., “How much relationship tension is there in your work group?”; $\alpha = 0.94$). The third subscale relates to process conflict (i.e., “How often are there disagreements about who should do what in your work group?”; $\alpha = 0.93$). Group members indicated the degree to which they experienced what was on each item using a Likert-style scale ranging from 1 = *none* to 5 = *a great deal*.

Team Satisfaction

Satisfaction with the team processes and outcomes was measured using two subscales of a group satisfaction scale developed by Briggs et al. (2006). Participants indicated their degree of agreement with statements on a seven-point Likert scale ranging from 1 = *strongly disagree* to 7 = *strongly agree*. Items from the subscale of the team processes pertained to feeling of satisfaction with procedures and operations followed by the group (e.g., “I feel satisfied with the procedures used in today’s meeting”). A Cronbach’s alpha of 0.96 was observed for this subscale. Items from the team outcomes subscale involved feelings of satisfaction related to the achievements of the group (e.g., “When the meeting was finally over, I felt satisfied with the results”; $\alpha = 0.93$).

Problem Solving

Solutions were rated for creativity by trained raters using a modified Consensual Assessment Technique (Amabile, 1996). Raters were graduate and undergraduate students. Raters were also blind to the study’s conditions. Raters received extensive training which involved a review of creativity, an overview of the rating scale system, the problem used in this study, and aspects of creativity to rate. Two raters assessed originality and three raters assessed quality as aspects of creativity. Originality refers to the uniqueness of the solution, whereas quality refers to the appropriateness and viability of the solution. Both facets of creativity were evaluated on a 1 = *very low* to 5 = *very high* scale. The two raters’ scores for originality were averaged, resulting in a single originality score for each solution. The three raters’ scores for quality were also averaged, resulting in a single quality score for each solution. Interclass correlations of 0.88 among ratings of originality and 0.94 for quality ratings were observed, indicating acceptable rater agreement (Shrout and Fleiss, 1979).

RESULTS

To address the methodological issue of which instructions for problem construction are effective in terms of their effect on solution quality and originality, the three conditions of problem construction were compared. One-way ANOVAs were conducted to compare the three problem construction conditions on originality and quality separately. No group differences in originality, $F(2,43) = 0.56$, $p = 0.578$, or quality, $F(2,43) = 0.68$, $p = 0.513$, were found based on the instructions for problem construction. Therefore, the three conditions were collapsed into one condition, allowing for control group to general problem construction manipulation comparisons. As a result, the following analysis reflected 19 groups in the control condition and 46 groups in the problem construction condition.

The second set of analyses was conducted to determine whether differences exist between teams that were asked to construct the problem and teams that were not asked in terms of the originality and quality of the solutions generated. Two ANOVAs were conducted to compare solution originality and quality, respectively, in problem construction and no problem construction conditions. Results indicated that there were marginal differences in solution originality for the problem construction condition and no problem construction condition $F(1,63) = 2.06$, $p = 0.078$; $\eta^2 = 0.03$), see **Table 1** and **Figure 1**. Teams that engaged in problem construction generated marginally significantly more original solutions compared to those that did not engage in problem construction. There were no differences in solution quality for the problem

TABLE 1 | ANOVA results comparing problem construction and no problem construction groups on solution quality and originality ratings.

		Sum of Squares	df	Mean Square	F	p	partial η^2
Quality	Contrast	0.33	1	0.33	0.27	0.304	0.00
	Error	76.33	63	1.21			
	Total	546.92	65				
Originality	Contrast	2.24	1	2.24	2.06	0.078	0.03
	Error	68.40	63	1.09			
	Total	520.50	65				

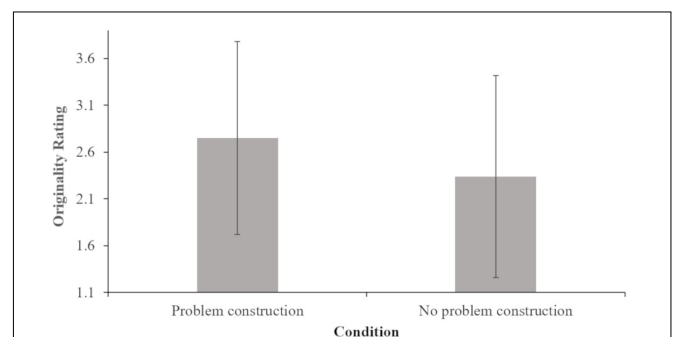


FIGURE 1 | Mean originality ratings of solutions from problem construction and no problem construction instruction conditions.

construction condition and no problem construction condition $F(1,65) = 0.272$, $p = 0.302$. The quality of solutions from teams that engaged in problem construction did not differ from the quality of solutions from teams that did not engage in problem construction.

To address the question of whether teams that engaged in problem construction were different than teams that did not engage in problem construction in terms of satisfaction and conflict, mean comparisons were used. As there were two different subscales for satisfaction and three for conflict, MANOVA was used for each one of the constructs, utilizing all the subscales. As both MANOVAs were significant, we are presenting the follow up ANOVAs on each subscale. There was a significant difference in process satisfaction between the problem construction condition and no problem construction condition $F = 3.2$, $p = 0.040$, $\eta^2 = 0.05$. Results for outcomes satisfaction indicated that there was a significant difference in outcome satisfaction between the problem construction condition and no problem construction condition $F = 2.10$, $p = 0.020$, $\eta^2 = 0.07$. That is, both process and outcome satisfaction was higher when teams engaged in problem construction compared to when teams did not engage in problem construction. See **Table 2** and **Figure 2**.

To evaluate whether there were differences between the problem construction condition and no problem condition for conflict, three ANOVAs were conducted. The first analysis of conflict involved task conflict. Results indicated that there was a significant difference in task conflict between the

problem construction condition and no problem construction condition $F = 5.09$, $p = 0.014$, $\eta^2 = 0.08$. There was a significant difference in relationship conflict between the problem construction condition and no problem construction condition $F = 3.9$, $p = 0.027$, $\eta^2 = 0.03$. Finally, to compare process conflict in problem construction and no problem construction conditions, a third ANOVA was conducted. Results indicated that there was a significant difference in process conflict between the problem construction condition and no problem construction condition $F = 3.21$, $p = 0.039$, $\eta^2 = 0.05$. These results indicated that all three measures of conflict, task, outcomes, and process conflict were lower when teams engaged in problem construction compared to when teams did not engage in problem construction. See **Table 3** and **Figure 3**.

DISCUSSION

This study provides the first empirical research in which team engagement in problem construction is manipulated through instructions. The findings above suggest that team problem construction can potentially benefit creativity. Although marginal, the apparent effect of problem construction on solution originality provides some initial support that team problem

TABLE 2 | ANOVA results comparing problem construction and no problem construction groups on satisfaction measures.

		Sum of Squares	df	Mean Square	F	p	partial η^2
Process	Contrast	0.85	1	0.85	3.20	0.040	0.05
	Error	16.68	63	0.27			
	Total	1192.30	65				
Outcome	Contrast	1.18	1	1.18	4.41	0.020	0.07
	Error	16.78	63	0.27			
	Total	1120.19	65				

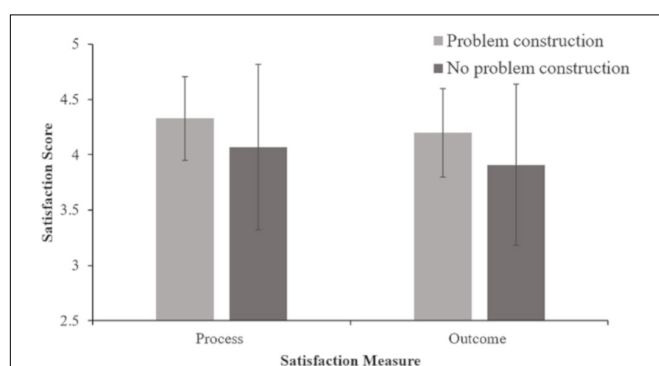


FIGURE 2 | Mean satisfaction scores in problem construction and no problem construction instruction conditions.

TABLE 3 | ANOVA results comparing problem construction and no problem construction group on conflict subscales.

		Sum of Squares	df	Mean Square	F	p	partial η^2
Task	Contrast	1.82	1	1.82	5.09	0.014	0.08
	Error	22.50	63	0.36			
	Total	225.04	65				
Relationship	Contrast	0.32	1	0.32	3.90	0.027	0.06
	Error	5.10	63	0.08			
	Total	94.74	65				
Process	Contrast	0.38	1	0.38	3.21	0.039	0.05
	Error	7.41	63	0.12			
	Total	104.62	65				

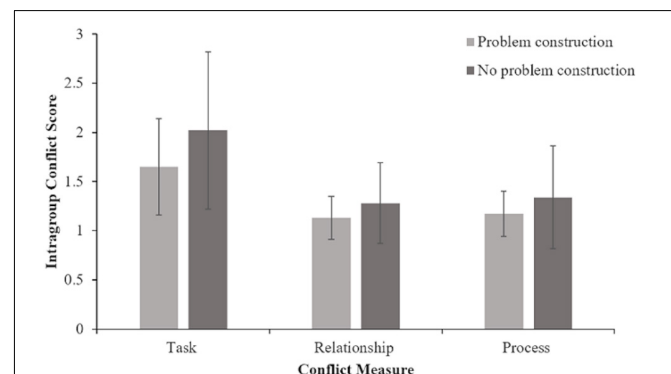


FIGURE 3 | Mean intragroup conflict scores in problem construction and no problem construction instruction conditions for each conflict measure.

construction leads to creative problem solving at the team level. Limited power, as a result of a relatively small number of teams in the control condition, offers some explanation for the observed bordering significance value of originality differences between groups. Nonetheless, the role of problem construction at the team level is further elucidated through the analyses.

More importantly, team problem construction may facilitate some of the social processes that can then help in effective problem solving. Taken together, the final set of analyses show that problem construction at the team level resulted in lower conflict and higher satisfaction. Past research focused on the effect of social processes on team cognition such as information sharing and elaboration or evaluated the concurrent nature of these relationships (Hoever et al., 2012; Qu and Liu, 2017). This study, however, evaluated the effect of team cognitive processes on social processes by manipulated instructions for problem construction. This experimental design allows us to directly evaluate whether cognitive processes can have an effect on social processes. Since problem construction was a manipulated variable, and occurred prior to the measurement of social processes, the causal inference that problem construction is the cause of improved social processes is appropriate. This study, therefore, addresses the call by Reiter-Palmon et al. (2012) to further elucidate the relationships between social processes and cognitive processes. As problem construction has been suggested to provide some basic structure for creative problem solving, this reduction in conflict and increase to satisfaction might result from a reduction in the uncertainty associated with ill-defined problems (Mumford et al., 1994). Furthermore, problem construction at a team level may counter disagreement and conflict, while also promoting group satisfaction, as a result of discussions early in the process while thinking about ideas and solutions is still more malleable. This research demonstrates that this cognitive process has implications beyond the individual level, denoting the broad utility of problem construction.

Finally, it is interesting to note that there do not seem to be differences among the various instructions provided for problem construction in terms of creative performance. This lack of condition differences hints that the manner in which a team engages in problem construction may not be as important to creativity as the act of a team engaging in problem construction in and of itself.

LIMITATIONS AND FUTURE RESEARCH

This study provides a first step in the study of manipulating problem construction in teams. One important limitation of this study was the fact that the sample size for the control group was somewhat low. This may have had a role in the marginal effect found for the originality of solutions. Future research should not only strive to replicate this research, but should include a larger number of teams to allow for more power and hopefully a significant effect of problem construction on creativity.

While we have speculated that problem construction caused a reduction in conflict and increased satisfaction due to the

structure that developed from the process, the exact nature of these relationships is still unclear. Future research should not only replicate the current findings but also add to them by identifying the process by which problem construction operates on these team processes, and whether indeed increased structure is what facilitated the benefits of problem construction. Further, while we expect that on average team composition variables and other relevant variables were equivalent in this sample, due to random assignment, this cannot be fully determined, and should be investigated. It is important to note that effective social processes can be more difficult to attain when teams are diverse (Leung and Wang, 2015). It would be therefore important to study whether the positive effect of problem construction on social processes found here, operates equally on diverse and non-diverse teams.

Another limitation is the use of short-term student teams. While short-term teams exist in organizations, and therefore this research provides meaningful information, the relationships between problem construction and creativity as well as social processes may not operate in the same way in long-term teams in which members share a history and know that they will continue to work with one another. As such, it is important to assess these relationships in long-term teams as well.

To add, although direct information on whether team members have had prior experience with each other was not collected, the current study assumes that the members of the three-person teams were strangers to each other. Given the large size of the university and psychology department and the method used to create groups, we expect that most teams included students that were not familiar with one another. It is possible that some teams had team members that were familiar with each other. In these groups the social processes of conflict reduction and satisfaction may operate differently than in groups composed of strangers. Further research and analyses are needed to determine the extent to which this effects influences the conclusions of this study.

Additionally, future research should also seek to reproduce this study's findings using organizational contexts and samples. Although many of this study's claims were intended to translate to application in organizational settings, the current study's findings were derived from data obtained from a student sample, as opposed to employees. Research on the generalizability of undergraduate research participants suggests that university student samples can be used to represent non-student populations when testing psychological processes and behaviors (Lucas, 2003). Despite the testing of such a process, problem construction, the true generalizability of the student sample in this study is unknown and stands as a potential limitation.

Finally, future research should evaluate whether problem construction influences other social process such as psychological safety, trust, or team efficacy. While our choice of studying conflict was a result of past research on rGaps, it is possible that effective problem construction stemming from instructions can also facilitate the development of psychological safety, trust, and communication, contributing to reduce conflict and increase satisfaction.

CONCLUSION

This study explores the benefits of problem construction instruction in facilitating creativity in teams. Furthermore, by relating the social process of team satisfaction and conflict to problem construction, this study provides empirical evidence that helps explain the role of team problem construction processes in team productivity. Although much more research is needed, this study contributes an initial look into team level creative cognition using an experimental design. As organizations continue to experience complex problems that surpass an individual's capacity, a more thorough understanding of the specific components of the creative process including and beyond problem construction at the team level is required.

REFERENCES

- Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Res. Organ. Behav.* 10, 123–167.
- Amabile, T. M. (1996). “A consensual technique for creativity assessment,” in *Creativity in Context*, ed. T. A. Amabile (Boulder, CO: Westview Press).
- Arreola, N., and Reiter-Palmon, R. (2016). The effect of problem construction creativity on solution creativity across multiple real-world problems. *Psychol. Aesthet. Creat. Arts* 10, 287–295. doi: 10.1037/a0040389
- Basadur, M. S. (1982). “Research in creative problem-solving training in business and industry,” in *Proceedings of the Creativity Week IV* (Greensboro, NC: Center for Creative Leadership).
- Briggs, R., Reining, B., and de Vreede, G. J. (2006). Meeting satisfaction for technology supported groups: an empirical validation of a goal-attainment model. *Small Group Res.* 37, 585–611. doi: 10.1177/1046496406294320
- Cronin, M. A., Bezrukova, K., Weingart, L. R., and Tinsley, C. H. (2011). Subgroups within a team: the role of cognitive and affective integration. *J. Organ. Behav.* 32, 831–849. doi: 10.1002/job.707
- Cronin, M. A., and Weingart, L. R. (2007). Representational gaps, information processing, and conflict in functionally diverse teams. *Acad. Manag. Rev.* 32, 761–773. doi: 10.5465/amr.2007.25275511
- De Dreu, C. K. W., Nijstad, B. A., Bechtoldt, M. N., and Baas, M. (2011). Group creativity and innovation: a motivated information processing perspective. *Psychol. Aesthet. Creat. Arts* 5, 81–89. doi: 10.1037/a0017986
- Dess, G. G., and Picken, J. C. (2000). Changing roles: leadership in the 21st century. *Organ. Dyn.* 28, 18–34. doi: 10.1016/S0090-2616(00)88447-8
- Finke, R. A., Ward, T. B., and Smith, S. M. (1992). *Creative Cognition: Theory, Research, and Applications*. Cambridge, MA: The MIT Press.
- Getzels, J. W., and Csikszentmihalyi, M. (1975). “From problem solving to problem finding,” in *Perspectives in Creativity*, eds I. A. Taylor and J. W. Getzels (Chicago, IL: Aldine), 90–115.
- Gish, L., and Clausen, C. (2013). The framing of product ideas in the making: a case study of the development of an energy saving pump. *Technol. Anal. Strateg. Manag.* 25, 1085–1101. doi: 10.1080/09537325.2013.832746
- Goh, K. T., Goodman, P. S., and Weingart, L. R. (2013). Team innovation processes: an examination of activity cycles in creative project teams. *Small Group Res.* 44, 159–194. doi: 10.1177/1046496413483326
- Hennessey, B. A., and Amabile, T. M. (2010). Creativity. *Annu. Rev. Psychol.* 61, 569–598. doi: 10.1146/annurev.psych.093008.100416
- Hoever, I. J., van Knippenberg, D., van Ginkel, W. P., and Barkema, H. G. (2012). Fostering team creativity: perspective taking as key to unlocking diversity's potential. *J. Appl. Psychol.* 97, 982–996. doi: 10.1037/a0029159
- Hulsheger, U., Anderson, N., and Salgado, J. (2009). Team-level predictors of innovation at work: a comprehensive meta-analysis spanning three decades of research. *J. Appl. Psychol.* 94, 1128–1145. doi: 10.1037/a0015978

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the IRB of the University of Nebraska Medical Center and University of Nebraska at Omaha. The protocol was approved by the IRB. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

RR-P was responsible for the conceptualization of the study, data analysis, and writing. VM was responsible for helping with data and writing.

- Jehn, K. A., and Mannix, E. A. (2001). The dynamic nature of conflict: a longitudinal study of intragroup conflict and group performance. *Acad. Manag. J.* 44, 238–251.
- Kozlowski, S., and Bell, B. (2008). *Team Learning, Development, and Adaptation. Work Group Learning: Understanding, Improving and Assessing How Groups Learn in Organizations*. New York, NY: Taylor & Francis Group, 15–44.
- Leonardi, P. M. (2011). Innovation blindness: culture, frames, and cross-boundary problem construction in the development of new technology concepts. *Organ. Sci.* 22, 287–539. doi: 10.1287/orsc.1100.0529
- Leung, K., and Wang, J. (2015). Social processes and team creativity in multicultural teams: a socio-technical framework. *J. Organ. Behav.* 36, 1008–1025. doi: 10.1002/job.2021
- Lucas, J. W. (2003). Theory-testing, generalization, and the problem of external validity. *Sociol. Theory* 21, 236–253. doi: 10.1111/1467-9558.00187
- Ma, H. (2009). The effect size of variables associated with creativity: a meta-analysis. *Creat. Res. J.* 21, 30–42. doi: 10.1080/10400410802633400
- Merrifield, P. R., Guilford, J. P., Christensen, P. R., and Frick, J. W. (1962). The role of intellectual factors in problem solving. *Psychol. Monogr.* 76, 1–21. doi: 10.1037/h0093850
- Mumford, M. D., and Hunter, S. T. (2005). Innovation in organizations: a multi-level perspective on creativity. *Multi Level Issues Strategy Methods* 4, 11–73. doi: 10.1016/S1475-9144(05)04001-4
- Mumford, M. D., Loneragan, D. C., and Scott, G. (2002a). Evaluating creative ideas: processes standards and context. *Inq. Crit. Think. Across Disciplines* 22, 21–30. doi: 10.5840/inquiryctnews20022213
- Mumford, M. D., Scott, G. M., Gaddis, B., and Strange, J. M. (2002b). Leading creative people: orchestrating expertise and relationships. *Leadersh. Q.* 13, 705–750. doi: 10.1016/S1048-9843(02)00158-3
- Mumford, M. D., Mobley, M. I., Uhlman, C. E., Reiter-Palmon, R., and Doares, L. M. (1991). Process analytic models of creative capacities. *Creat. Res. J.* 4, 91–122. doi: 10.1080/10400419109534380
- Mumford, M. D., Reiter-Palmon, R., and Redmond, M. R. (1994). “Problem construction and cognition: applying problem representations in ill-defined domains,” in *Problem Finding, Problem Solving, Creativity*, ed. M. Runco (Norwood, NJ: Ablex), 3–39.
- Qu, X., and Liu, X. (2017). Informational faultlines, integrative capability, and team creativity. *Group Organ. Manag.* 42, 767–791. doi: 10.1177/105960117716008
- Reiter-Palmon, R. (2018). “Creative cognition at the individual and team level: what happens before and after idea generation,” in *Nature of Creativity*, eds R. Sternberg and J. Kaufman (New York, NY: Cambridge Press), 184–208. doi: 10.1017/9781108185936.015
- Reiter-Palmon, R., Herman, A. E., and Yammarino, F. (2008). “Creativity and cognitive processes: a multi-level linkage between individual and team cognition,” in *Multi-level Issues in Creativity and Innovation*, Vol. 7, eds M. D. Mumford, S. T. Hunter, and K. E. Bedell-Avers (Stamford, CT: JAI Press), 203–267.

- Reiter-Palmon, R., Mumford, M. D., Boes, O. J., and Runco, M. A. (1997). Problem construction and creativity: the role of ability, cue consistency, and active processing. *Creat. Res. J.* 10, 9–24. doi: 10.1207/s15326934crj1001_2
- Reiter-Palmon, R., and Robinson, E. J. (2009). Problem identification and construction: what do we know, what is the future? *Psychol. Aesthet. Creat. Arts* 3, 43–47. doi: 10.1037/a0014629
- Reiter-Palmon, R., Wigert, B., and Vreede, T. D. (2012). “Team creativity and innovation: the effect of group composition, social processes, and cognition,” in *Handbook of Organizational Creativity*, ed. M. D. Mumford (Amsterdam: Elsevier Inc.), 295–326. doi: 10.1016/B978-0-12-374714-3.00013-6
- Rostan, S. (1994). Problem finding, problem solving, and cognitive controls: an empirical investigation of critically acclaimed productivity. *Creat. Res. J.* 7, 97–110. doi: 10.1080/10400419409534517
- Schraw, G., Dunkle, M. E., and Bendixen, L. D. (1995). Cognitive processes in well-defined and ill-defined problem solving. *Appl. Cogn. Psychol.* 9, 523–538. doi: 10.1002/acp.2350090605
- Shalley, C., Zhou, J., and Oldham, G. (2004). The effects of personal and contextual characteristics on creativity: where should we go from here? *J. Manag.* 30, 933–958. doi: 10.1016/j.jm.2004.06.007
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420
- Silverman, B. G. (1985). The use of analogs in the innovation process: a software engineering protocol analysis. *IEEE Trans. Syst. Man Cybern.* 15, 30–44.
- Sternberg, R. J. (1988). “A three-facet model of creativity,” in *The Nature of Creativity: Contemporary Psychological Perspectives*, ed. R. J. Sternberg (New York, NY: Cambridge University Press), 125–147.
- Taggar, S. (2002). Individual creativity and group ability to utilize individual creative resources: a multilevel model. *Acad. Manag. J.* 45, 315–330.
- Weingart, L., Cronin, M., Houser, C., Cagan, J., and Vogel, C. (2005). “Functional diversity and conflict in cross-functional product development teams: considering representational gaps and task characteristics,” in *Understanding Teams*, eds L. Neider and C. Schriesheim (Greenwich, CT: IAP), 89–100.
- Weingart, L. R., Todorova, G., and Cronin, M. A. (2008). Representational gaps, team integration and team creativity. *Paper Presented at the Academy of Management 2008 Annual Meeting: The Questions We Ask, AOM 2008*, Briarcliff Manor, NY.

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 Reiter-Palmon and Murugavel. 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.



A Growth-Curve Analysis of the Effects of Future-Thought Priming on Insight and Analytical Problem-Solving

Monica Truelove-Hill*, Brian A. Erickson, Julia Anderson, Mary Kossoyan and John Kounios

Department of Psychology, Drexel University, Philadelphia, PA, United States

OPEN ACCESS

Edited by:

Amory H. Danek,
Universität Heidelberg, Germany

Reviewed by:

Saskia Jaarsveld,
Technische Universität Kaiserslautern,
Germany

Mark Beeman,
Northwestern University,
United States

Kristin Grunewald,
Northwestern University,
United States, in collaboration with
reviewer MB.

*Correspondence:

Monica Truelove-Hill
mlh349@drexel.edu

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 16 January 2018

Accepted: 09 July 2018

Published: 30 July 2018

Citation:

Truelove-Hill M, Erickson BA,
Anderson J, Kossoyan M and
Kounios J (2018) A Growth-Curve
Analysis of the Effects
of Future-Thought Priming on Insight
and Analytical Problem-Solving.
Front. Psychol. 9:1311.
doi: 10.3389/fpsyg.2018.01311

Research based on construal level theory (CLT) suggests that thinking about the distant future can prime people to solve problems by insight (i.e., an “aha” moment) while thinking about the near future can prime them to solve problems analytically. In this study, we used a novel method to elucidate the time-course of temporal priming effects on creative problem solving. Specifically, we used growth-curve analysis (GCA) to examine the time-course of priming while participants solved a series of brief verbal problems. Participants were tested in two counterbalanced sessions in a within-subject experimental design; one session featured near-future priming and the other featured far-future priming. Our results suggest high-level construal may temporarily enhance analytical thinking; far-future priming caused transient facilitation of analytical solving while near-future priming induced weaker, transient facilitation of insightful solving. However, this effect is short-lived; priming produced no significant differences in the total number of insights and analytical solutions. Given the fleeting nature of these effects, future studies should consider implementing methodology that allows for aspects of the time-course of priming effects to be examined. A method such as GCA may reveal mild effects that would be otherwise missed using other types of analyses.

Keywords: creativity, problem solving, temporal construal, growth curve analysis, insight

INTRODUCTION

Construal level theory (CLT) proposes that psychological distance from the self determines the way that one represents an object or event through mental construal (Trope and Liberman, 2003, 2010). High-level construals encompass the abstract, general features of an event or object. They omit the fine details about an object in favor of a broader representation of the object’s features (Trope and Liberman, 2010). Conversely, low-level construals include the context-dependent, concrete features of events or objects. For example, moving from ‘animal’ to ‘mammal’ to ‘canine’ to ‘dog’ represents a gradual shift from high-level to low-level construal. According to CLT, events that are psychologically distant will be represented by high-level, abstract construals, while those that are psychologically proximal will be represented by low-level concrete construals. Temporal distance reflects psychological distance in time of an event from the individual.

In line with CLT, thinking about the distant future requires more high-level construals than thinking about the near future, the latter requiring more low-level construals. In other words, individuals will form a more abstract mental representation of an event in the distant future than of an event in the near future. Because the near future is relatively proximal to the present, one has a more concrete idea of what to expect of events that occur in this time period. The distant future, on the other hand, requires more imagination—the context is unknown, and factors that are relevant to the present may change in the meantime. For example, when someone is planning a trip in the near future, there are very specific deadlines that must be met. Tickets must be booked, accommodations must be arranged, and even minor details such as the upcoming weather are known and may be incorporated in one's decisions. If a trip is taking place in the distant future, the planning is much more abstract. General ideas such as where to go and what to do may be identified, but the concrete details cannot be considered until the trip is much closer.

Research by Liberman et al. (2002) supports this idea. In one study, participants were asked to think about completing everyday life tasks in either the distant future (1 year from the present) or the near future (1 week from the present). Participants in the distant future condition rated their ability to cope with a wide variety of everyday life tasks more similarly than those in the near future condition, suggesting less nuance in the way that distant-future tasks are conceptualized compared to near future tasks. Additionally, participants who underwent distant future priming implemented broader categories when sorting objects than those who underwent near future priming, suggesting that the more abstract mindset promoted by distant future thought can be generalized to other tasks. Other research has substantiated this idea—inducing a more abstract mindset may influence, for example, how consumers perceive advertisements (Martin et al., 2009) and how individuals deploy self-presentation strategies (Carter and Sanna, 2008). Another area which may be influenced by temporal construal priming is the method by which someone solves problems.

One of the methods people commonly use when confronted with a problem is to consciously manipulate the elements of the problem until a solution is derived. In this *analytical* approach, one works through a problem, step by step, and gradually comes to a solution. For example, one typically uses analytical problem solving when faced with an arithmetic problem. Another method by which one may solve a problem is through *insight*, commonly considered a form of creative cognition (for a discussion of the relationship between creativity and insight, see Kounios and Beeman, 2015). To solve by insight involves a sudden restructuring of the problem so that the solution is immediately clear. Unlike analytical solving (DeWall et al., 2008), insight solving is largely the result of unconscious processing (Kounios and Beeman, 2014); one's subjective experience is that the solution came from nowhere (Schooler and Melcher, 1995). Indeed, research has shown that participants are able to rate their nearness to a solution in the case of analytical solving, but not for insight solving (Metcalfe and Wiebe, 1987). In this study, we tested whether these two problem-solving styles would be differentially affected by temporal construal priming.

Research has already shown that problem-solving style may be affected by a person's prior internal state (see review by Kounios and Beeman, 2014). For example, neural activity immediately preceding the presentation of a problem predicts whether participants will solve that problem insightfully or analytically (Kounios et al., 2006). Subramaniam et al. (2009) showed that mood may also influence one's brain state; in their study, a positive mood facilitated insightful solving, while an anxious mood enhanced analytical solving. Furthermore, resting-state brain activity predicts individual differences in problem-solving strategies: Participants who tend to rely more on insight exhibit different patterns of prior resting-state electroencephalogram (EEG) brain activity than those who tend to rely on analysis (Kounios et al., 2008). In sum, neuroimaging findings are consistent with the idea that mindset changes via temporal construal priming could have a significant influence on cognitive style.

A behavioral study by Förster et al. (2004) suggested that temporal construal priming influences problem-solving style. Specifically, they hypothesized that high-level construals utilized to imagine the distant future would promote insightful problem solving and that low-level construals utilized to imagine the near future would promote analytical solving. In a series of experiments, participants were asked to both imagine their life in general and imagine solving the subsequent task either in the distant future (1 year from the present day) or the near future (the next day). They reported that individuals asked to think about the distant future solved more insight problems, performed better on a creativity task, and performed worse on an analytical task. Often in creative problem-solving, one must overcome a cognitive fixation on how they assume the problem should be solved to restructure the problem in a novel manner (Smith, 1995). This fixation would be more difficult to overcome if a problem is presented in a greater level of detail, as might be expected for concrete, low-level construal. Research has supported this—when individuals were given examples on how to solve a problem, they were less likely to produce novel solutions than participants who were not provided with examples (Marsh et al., 1999). Therefore, it seems intuitive that high-level, abstract construal would benefit insight, as Förster and colleagues hypothesized. Indeed, previous research has shown that approaching a problem in a more abstract manner leads to more novel solutions than when the task is approached more concretely (Ward et al., 2004).

However, studies in which a specific mindset is primed in order to observe its effect on subsequent behavior have proven difficult to replicate (e.g., Gong and Medin, 2012; Pashler et al., 2012; Shanks et al., 2013). This report examines the consequences of mindset priming for problem-solving style. In particular, we applied a new analytic approach to investigate the time-course of the effects of future thought priming on analytical solving versus solving by insight. The present study had two main goals. The first was to test whether distant prospection benefits insight while more proximal prospection benefits analytical thinking, as suggested by Förster et al. (2004). Because of recent concerns about the replicability of social priming studies (Kahneman, 2012) we deemed it worthwhile to examine this issue.

Second, we implemented several methodological refinements to better isolate and elucidate the effects of priming. Förster et al. (2004) tasked their participants with solving both verbal and visual insight problems but did not verify whether their participants actually solved these problems with insight. Insight research has shown that just because a person has solved a so-called “insight problem” does not mean that he or she solved it with insight (Kounios and Beeman, 2014; Danek et al., 2016). We used the insight judgment procedure developed by Bowden et al. (2005) to determine which problems were solved insightfully and which were solved analytically. Instead of using classic insight problems which take participants a considerable amount of time to solve (when they are able to solve them), we used compound remote associates (CRA) problems, verbal puzzles which can be solved in less than 15 s and which have a long history of use in studying creativity and insight (Bowden and Jung-Beeman, 2003). CRAs are well-defined, convergent problems. Each CRA problem consists of 3 stimulus words that can be combined with a single solution word to form 3 individual compound words or phrases (e.g., horse, plant, over; solution = power: horsepower, power plant, overpower). Importantly, CRA problems can be solved either by insight or analysis. Based on an individual participant’s trial-by-trial reports of their solution strategy, insightful and analytical solutions can be sorted and compared. One of the major benefits of this approach is that it allows the experimenter to compare solving strategies while holding constant the type of problem.

Another benefit of using short puzzles is that it allows researchers to trace the time-course of priming effects on solving strategy. One reason that priming effects are difficult to replicate may be because these effects are too short-lived to reliably influence a subsequent task. We were able to assess this possibility by adapting growth curve analysis (GCA) to examine the time-course of temporal construal priming. GCA is a type of multilevel regression that allows for the analysis of the trajectory of time-course data (Mirman, 2014) so that one can examine change in the data over time. Using more traditional statistical methods (e.g., *t*-tests), one can compare between individual time points. However, these methods provide no information about what is happening across those time points. Using GCA, one can observe the patterns of change that occur across time points.

Growth curve analysis models are developed based on the shape of the data, fixed effects (group-level predictor variables), and random effects (variables that represent individual variability). In many cases, a linear model is a suitable reflection of time-course data. Indeed, if the priming effect persists throughout the experiment, we would expect that a linear model would best fit the data as the primed behavior would remain relatively stable. However, a linear model would not accurately identify the deterioration of priming effects over the course of an experiment. Rather, a quadratic model would successfully reveal this pattern, as one would expect an initial increase in the primed behavior, followed by a decline as the effect decays. Therefore, GCA is a useful analysis that allows for the examination of the nature of the priming effect. If these effects deteriorate over the course of a short experiment, priming researchers should take that into account during future study development. This

is particularly important for those who utilize classic insight problems in their research, as these problems may require extensive solving time. Depending upon the number of problems used, the effect of priming may deteriorate before all problems have been solved.

Additionally, to maximize statistical power, we used a within-subject experimental design in which each participant was included in both a near-future and far-future thought condition (in separate counterbalanced sessions). This contrasts with the lower-power between-group design of Förster et al. (2004) and most other social priming studies.

Finally, given that one’s brain activity before a problem is presented is known to influence the strategy with which one solves the problem (Kounios and Beeman, 2014), we also measured participants’ resting-state electroencephalograms (EEG) between priming phases in order to ascertain how such priming affects ongoing brain activity.

In sum, we tested the effects of temporal-construal priming on problem-solving style (insightful versus analytic) and examined the time-course and neural correlates of the resulting effects.

MATERIALS AND METHODS

This study was carried out in accordance with the recommendations of the Drexel University IRB with written informed consent from all the subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Drexel University IRB. The data are available for download at: https://figshare.com/articles/Temporal_Priming_Creative_Insight/4007745.

Participants

Förster et al. (2004) reported large effects of temporal priming. Furthermore, based on past EEG studies with the insight judgment procedure and a within-subject design (e.g., Kounios et al., 2008), we expected that approximately 25 participants would yield good statistical power for analyses of both the behavioral and EEG data. Given expected participant exclusions due to EEG artifacts, low problem-solving accuracy, failure to follow instructions, and participant withdrawals, we recruited 38 participants.

All participants were right-handed, had no self-reported neurological disorders or psychiatric conditions, and refrained from taking substances that might affect cognition (i.e., alcohol, psychoactive medications, or recreational drugs) for 24 h prior to the experiment. We excluded 2 participants who did not produce at least 1 solution of each type (insight and analytic) because this suggested that they were responding stereotypically or were not following instructions. We also excluded 2 subjects who did not achieve an accuracy lower than 1.5 standard deviations below the sample mean (~15% accuracy) in solving the problems, 3 due to equipment problems, and 4 who chose not to complete the study. After these exclusions, our final sample included 27 Drexel University students ages 18–30 ($M = 22.15$, $SD = 3.28$, 13 females, 13 males, 1 declined to report) who were paid \$30 to participate.

Procedure

Participants completed 2 2-h experimental sessions on different days. During the first session, participants filled out demographic and handedness questionnaires and watched an instructional video during which the experimental procedure was explained and the differences between analytical and insightful problem solving were described. We recorded 5 min of eyes-closed baseline resting-state EEG data during which participants were instructed to let their minds wander. Then, participants were presented with 1 of 4 possible priming scenarios (2 in the near condition and 2 in the far condition, as described below) and asked to write about that scenario for 5 min. After this priming, we recorded 5 min of eyes-closed resting-state EEG data. Because of the documented effects of mood on insight (Subramaniam et al., 2009), participants then completed the Positive and Negative Affect Scale (PANAS). Participants completed another priming scenario (same time-frame) for 5 min to refresh the priming after the EEG recording and PANAS. Following this, participants attempted 72 CRA problems while recording EEG. The second session used the same procedure (**Figure 1**). Participants who received far-future priming scenarios in the first session

received near-future scenarios in the second session, and vice versa.

Materials

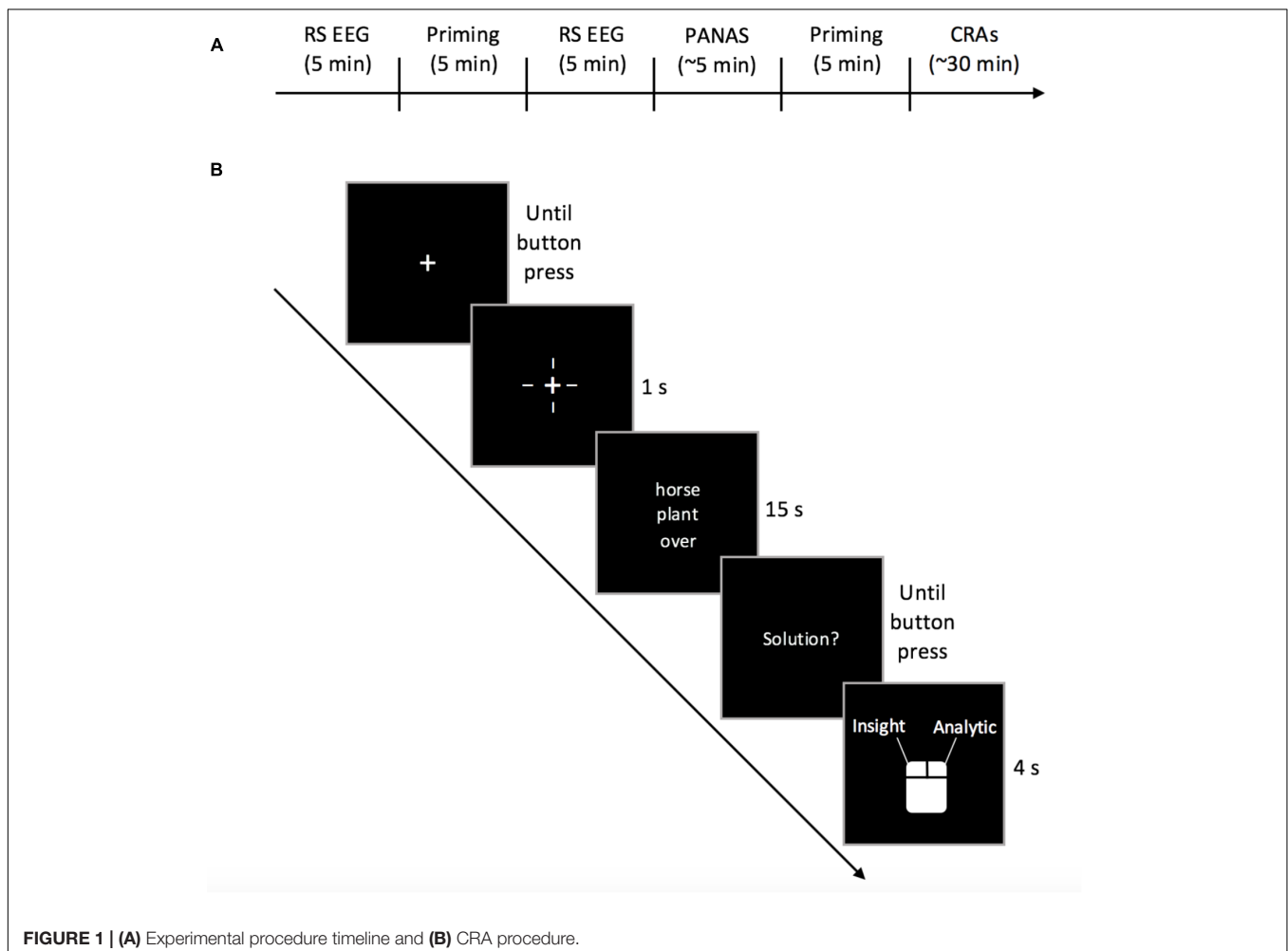
Priming Scenarios

We used 4 priming scenarios, differing both in content and temporal proximity. The scenarios were restricted to the Philadelphia area to control for potential spatial-distance priming effects. The scenarios are as follows:

- “Imagine that you will be finding a place to live in Philadelphia next week. You have 5 min to write about whatever comes to mind about this.”
- “Imagine that you will be finding a new job in Philadelphia next week. . .”
- “Imagine that you will be finding a place to live in Philadelphia in 10 years. . .”
- “Imagine that you will be finding a new job in Philadelphia in 10 years. . .”

Compound Remote Associates

Participants were presented with 144 CRA problems over the course of the study. The assignment of CRA problems to sets was



randomized and the sets were counterbalanced between groups. Each set of CRAs was presented to participants in a single random order. The problems were presented using e-Prime 2.0. Eight practice trials were presented before each session. Participants held a mouse in both hands with left and right thumbs placed on the corresponding buttons. A fixation cross was displayed in the center of the screen until participants initiated the presentation of a problem with a bimanual button press. Once participants initiated the problem, crosshairs appeared around the fixation cross for 1000 ms after which the problem appeared. The 3 words of each problem were displayed in a column for 15 s. If a participant was unable to reach a solution, the screen returned to the fixation cross and the trial was terminated. If a participant reached a solution, she or he indicated this with a bimanual button-press. Then, a prompt appeared on the screen, participants verbalized their solution, and the experimenter recorded solution accuracy. Participants were then prompted to press a button to indicate whether they had solved the problem insightfully (i.e., resulting from an “aha” moment in which the solution suddenly intrudes on ongoing thought) or analytically (i.e., in which the solution resulted from deliberate, conscious manipulation of the elements of the problem, as in hypothesis testing; Bowden et al., 2005). If participants were unable to come to a conclusion as to how the problem was solved, they refrained from pressing anything, and the program continued after 4 s.

EEG Recording and Data Processing

Eighty-four channel electroencephalographic data were recorded with tin electrodes embedded in a nylon cap (Electro-Cap International, Eaton, OH, United States) using the MANSCAN EEG recording system (SAM Technology, Inc., San Francisco, CA, United States) and extended 10–20 system locations referenced to digitally linked mastoid electrodes. Data were preprocessed using the EEGLAB toolbox in Matlab 7.14 (Mathworks, Inc., Natick, MA, United States). Bad channels were removed by visual inspection. Data were segmented and filtered using a 1-Hz high-pass and 55-Hz low-pass FIR filter. Movement artifacts were removed using an amplitude threshold ranging from -300 to 300 μ V (Hoffman and Falkenstein, 2008). ICA weights were calculated using EEGLAB's FASTICA algorithm and submitted to the ADJUST artifacting tool (Mognon et al., 2011). Previously removed channels were replaced by interpolation. Analyses were conducted in SPM 12's EEG toolbox (Litvak et al., 2011). Fast Fourier transforms (FFT) were calculated from 2 to 55 Hz in frequency steps of 2 Hz (Hamming windowed), robust averaged, and log transformed within session, then transformed into 3D Scalp \times Frequency images. Tests were performed with a $p < 0.001$ cluster-correction threshold.

TABLE 1 | Positive and negative affect scores by condition.

	PAS		NAS	
	Mean	SD	Mean	SD
Near-future condition	34.96	5.56	21.79	7.95
Far-future condition	33.96	5.63	22.46	6.58

Behavioral Data Analysis

Growth-curve analysis (Mirman, 2014) was used to analyze change over time in the relative accumulation of solutions over the course of the 72 CRA problems presented during each session. GCA offered information both about the influence of priming on solution type and the time course of this influence. All analyses were undertaken with R version 3.1.1 using the lme4 package (version 1.1-7).

Solution Difference (Insight – Analytical Solutions)

The time-course of changes in solving style (insight versus analysis) was modeled with second-order orthogonal polynomials using fixed effects of priming on all time terms (in all analyses in this report, this refers to the intercept, linear, and quadratic terms) and with participant and participant-by-condition (near versus far priming) random effects on all time terms. In this analysis, the intercept term refers to the average solution difference score, the linear term refers to the change in the solution difference score over time, and the quadratic term captures the curvature of the data—specifically, the increase and then subsequent decrease of the solution difference score over time, or vice versa. The far-priming condition was treated as baseline with parameters being estimated for the near-priming condition. Parameter-specific p -values were estimated using the normal approximation.

Solution Accumulation

The overall time-course for each condition (near versus far priming) was modeled with second-order orthogonal polynomials using fixed effects of solution type on all time terms and with participant and participant-by-solution type (analytical versus insight solution) random effects on all time terms. In this analysis, the intercept term refers to the average number of each solution type, the linear term captures the solution accumulation rate, and the quadratic term reflects the change the rate of solution accumulation over the course of the experiment. Insight solutions were treated as baseline with parameters being estimated for analytical solutions. Parameter specific p -values were estimated using the normal approximation.

RESULTS

Mean Performance

In the far-future priming condition, participants reported an average of 11.30 ($SD = 6.03$) correct insight solutions and 12.26 ($SD = 7.34$) correct analytical solutions. They solved 8.44 ($SD = 11.41$) problems incorrectly, and timed out in 39.44 ($SD = 11.61$) trials. In the near-future priming condition, participants reported an average of 10.78 ($SD = 5.44$) correct insight solutions and 10.81 ($SD = 5.76$) analytical solutions. They solved an average of 10.30 ($SD = 14.31$) problems incorrectly, and timed out in 38.85 ($SD = 14.34$) trials. Neither the positive affect ($t = -1.00$, $p = 0.329$) nor the negative affect ($t = 0.486$, $p = 0.632$) PANAS scores significantly differed between conditions (see Table 1).

There were no significant differences between priming conditions in terms of total correct solutions ($p = 0.106$), total incorrect solutions ($p = 0.429$), and total timeouts ($p = 0.897$). Data for all of the following models can be found in **Table 2**.

Solution Difference-Scores

The effect of priming significantly improved model fit on the quadratic term, $\chi^2 = 12.75$, $p < 0.001$, indicating that a curvilinear model best fits the data. Solution pattern and consistency differed significantly between conditions over the course of the experiment, as reflected by differences in the steepness of the quadratic curvature between the near- and far-future priming conditions. Specifically, participants in the far-future priming condition produced consistently more analytical solutions in the initial stages of the experiment, $Estimate = 2.80$, $SE = 1.12$, $p = 0.013$. Conversely, significance in the opposite direction in the near-future priming condition indicates that participants utilized more insightful solving immediately after priming, $Estimate = -5.54$, $SE = 1.51$, $p < 0.001$. This difference grew smaller as the experiment progressed (**Figure 2**).

Near-Future Priming Condition

The effect of solution type significantly improved model fit on the quadratic term, $\chi^2 = 7.14$, $p = 0.008$, indicating a curvilinear model as the best fit of the data. Solution type did not significantly affect the intercept or the linear terms, $p = 0.599$, indicating that there was no significant difference in solution type in the near-priming condition; overall, participants tended to apply analytical and insightful methods about equally often. However, the effect of solution type on the quadratic term reflects differences in the steepness of quadratic curvature between the two conditions. This can be related to solution-type accumulation over time. Specifically, with near-future priming, insights initially accumulated somewhat more rapidly than analytical solutions, $Estimate = -2.47$, $SE = 0.73$, $p = 0.001$. However, the curvature of analytical solutions was also significant, but in the opposite direction, $Estimate = 2.75$, $SE = 0.99$, $p = 0.006$, which suggests that they were mildly suppressed by near-future priming.

Although participants applied roughly equal numbers of insightful and analytical solving methods over the course of the experiment, the rate of accumulation of each solution type differed (**Figure 3**). Insightful solutions accumulated slightly more rapidly than analytical solutions in the initial portion of the experiment.

Far-Future Priming Condition

The effect of solution type significantly improved model fit on the quadratic term, $\chi^2 = 8.10$, $p = 0.004$, indicating a curvilinear model as the best fit of the data (**Figure 4**). As in the near-priming condition, there was no significant difference in solution type, $p = 0.306$, but, rather, there was a significant difference in the rate of solution accumulation over time, as indicated by the steepness of the curvature in the analytical condition, $Estimate = -2.80$, $SE = 0.95$, $p = 0.003$. Specifically, analytical solutions initially accumulated more rapidly than insights.

Similar to the near-future priming condition, participants utilized relatively equal numbers of insightful and analytical solutions over the course of the experiment. However, the rate of accumulation differed. In this condition, analytical solutions accumulated more rapidly than insightful solutions in the initial portion of the experiment.

Resting-State EEG Data

The resting-state EEGs were subjected to frequency-domain analyses. To test for priming differences across all frequency bands (2–50 Hz), a flexible factorial model was created with the factors *order* (of priming condition) and *priming-condition* (near-versus far-future conditions). The first contrast tested the main effect of priming-condition in an F -test. No clusters survived at a cluster-forming threshold of $p < 0.001$. Because in-preparation analysis of other resting state data that we have collected shows that differences in resting-state beta-band oscillations are the strongest predictor of subsequent problem-solving strategy, we performed a focused analysis of priming condition constrained to the beta band (13–30 Hz). Again, no clusters survived at a cluster-forming threshold of $p < 0.001$. In sum, these analyses revealed no significant brain-activity differences between the near-future and far-future priming conditions after 5 min of priming (**Figure 5**). Means of the logged beta EEG power values for selected representative electrodes are shown in **Table 3**.

DISCUSSION

Research by Förster et al. (2004) indicates that thinking about the distant future promotes both creative processes (such as insight) and creative outputs and suppresses analytical reasoning. Our data contradict this. Distant-future thought primed analytical problem solving while near-future thinking primed insightful solving. Moreover, the shapes of the fitted curves illustrate a deterioration of these priming effects over approximately 30 min

TABLE 2 | Model fit results for each analysis.

	Solution difference (I – A)			Near-future priming			Far-future priming		
	LL	χ^2	p	LL	χ^2	p	LL	χ^2	p
Base model	–5933.3	–	–	–3945.2	–	–	–4369.6	–	–
Intercept	–5931.9	2.75	0.097	–3943.9	2.54	0.111	–4368.6	1.85	0.174
Linear	–5872.6	118.46	<0.001*	–3943.9	0.08	0.771	–4368.6	0.16	0.687
Quadratic	–5866.2	12.75	<0.001*	–3940.3	7.14	0.008*	–4364.5	8.1	0.004*

*Significant improvement in model fit.

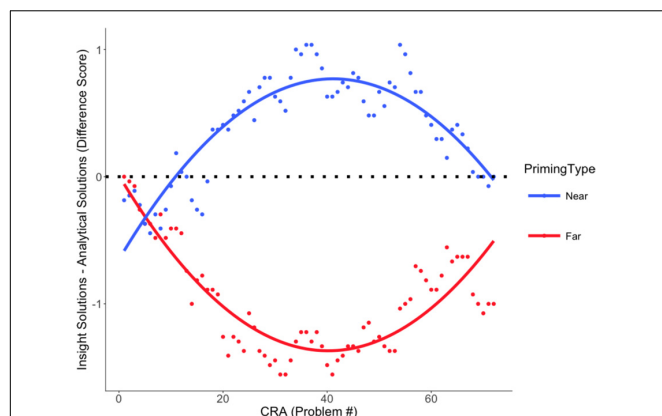


FIGURE 2 | Solution difference score by priming type. Model fit of the solution difference score (insight solutions – analytical solutions) by priming type (near-future versus far-future) over the series of CRA problems.

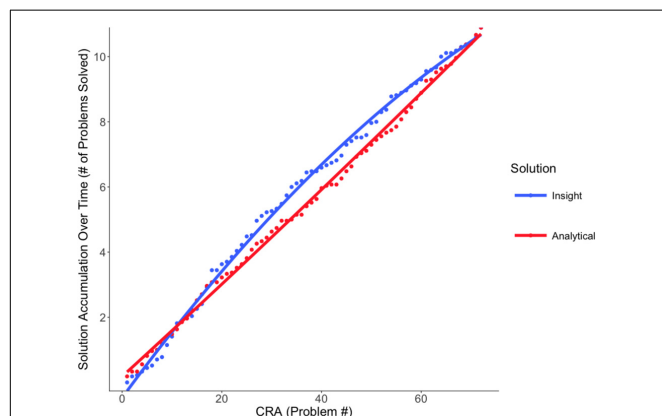


FIGURE 3 | Near-future priming solution accumulation. Model fit of the accumulated solutions (insight versus analytical) over the series of CRA problems in the near-future priming condition. Although participants used roughly the same number of each solution type, insightful solutions initially accumulated somewhat more rapidly than analytical solutions.

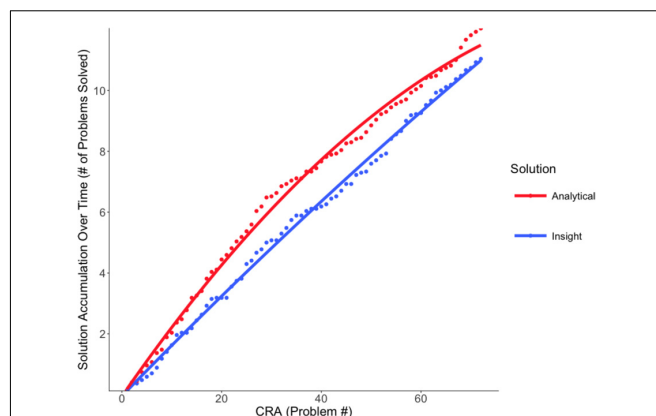


FIGURE 4 | Far-future priming solution accumulation. Model fit of the accumulated solutions (insight versus analytical) over the series of CRA problems in the far-future priming condition. Although participants used similar numbers of each solution type, analytical solutions initially accumulated more rapidly than insightful solutions.

(the time course of the stimulus presentation procedure). The priming effect was more pronounced in the far-future priming condition than in the near-future condition. This was not unexpected because the near future is similar to the present. Far-future thought would plausibly induce a greater change in mind-set and a more pronounced priming effect because the far future is comparatively dissimilar to the present.

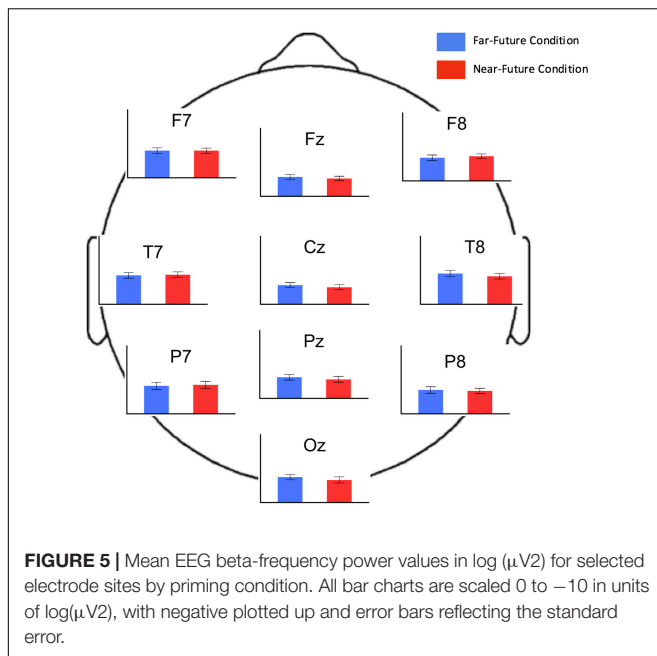
One possible explanation for the difference between our findings and Förster et al.'s (2004) is that future-thought priming effects may be highly dependent on the specific content of the priming scenarios. For example, our scenarios may have prompted more concrete construals, regardless of priming condition, than those used in the Förster et al. (2004) study. Thinking about detail-oriented tasks such as finding a place to live or finding a job may produce an inherently more low-level construal than thinking about life in general. However, if this were the case, then we might expect predominantly analytical

solutions in both priming conditions. This did not occur – near-future priming gave a small temporary boost to insightful solving. Another hypothesis is that the priming scenarios could have induced mood changes strong enough to override temporal-construal priming (Subramaniam et al., 2009). However, the absence of any significant priming effects on the PANAS mood questionnaire results weighs against this hypothesis. Finally, it is possible that the tasks that Förster et al. (2004) used did not tap creativity or insight and that their participants were using analytic thought to accomplish them. Because participants may solve so-called classic insight problems by using analytical methods (Danek et al., 2016), the present study used a method that revealed on a trial-by-trial basis the type of processing that each participant used to solve each CRA problem.

One potential explanation for our findings is that high-level construal, such as thinking about the distant future, may engage executive processes involved in working memory maintenance and inhibition of prepotent long-term memory representations more than low-level construal. Indeed, several studies indicate

TABLE 3 | EEG beta power values for selected electrodes (μV^2).

Electrode	Near-future		Far-future	
	Mean	SD	Mean	SD
Fz	–2.50	1.85	–2.46	2.43
F7	–3.98	1.97	–4.00	2.02
F8	–3.62	1.73	–3.40	2.12
Cz	–2.26	2.17	–2.20	2.59
T7	–4.79	2.32	–4.49	2.69
T8	–4.05	2.26	–4.51	2.34
Pz	–2.66	2.36	–2.62	2.69
P7	–4.19	2.72	–3.95	2.89
P8	–3.61	2.33	–3.62	3.05
Oz	–2.85	2.86	–2.73	3.27



that imagining a future event draws heavily on working memory and other executive processes required for analytical problem solving (e.g., D'Argembeau et al., 2010; Zavagnin et al., 2016). It is expected that far-future priming would draw more heavily upon these processes than near-future priming because an event in the near future is very similar to an event in the present. Specifically, imagining that you are looking for a job next week is not significantly different than imagining that you are currently looking for a job. The only details that must be retained in working memory are the few slight deviations from one's current situation; namely, that one has to find a job. In contrast, one is likely to assume that things will be quite different 10 years from the present. One may assume that they are married, possibly with children, and may have other family responsibilities or interests. They will likely expect to have different career options than they presently have. Thus, when imagining the distant future, one has to maintain in working memory all of these new features, while inhibiting some features of the present that conflict with those being imagined. In essence, imagining the distant future is likely a more computationally complex simulation than imagining the near-future. These findings, together with research indicating that enhanced analytical problem solving depends on working memory capacity more than insightful problem solving (Fleck, 2008; Wiley and Jarosz, 2012; DeCaro, 2016), lend credence to the idea that thinking about the distant future primes analytical thinking by activating these executive processes.

Interestingly, the temporary facilitation of analytical problem-solving in the distant-future condition did not produce a significant change in the total number of analytical solutions compared to insights. This suggests that not only does the priming's facilitating effect deteriorate, but analytical solving may actually be suppressed for a short time, as in a rebound effect. Because analytical problem-solving requires deliberate, focused

attention (Kahneman, 2011) and because executive processes are susceptible to resource depletion (e.g., van der Linden et al., 2003; Persson et al., 2007), it is plausible that a rebound effect may occur due to cognitive fatigue from sustained analytical thought. This rebound effect is not as robust in the near-future priming condition, which may be in keeping with the idea that insightful problem-solving is largely unconscious (Fleck, 2008), and would plausibly induce less cognitive fatigue. However, it may also be less robust because the effect of near-priming is weaker in general.

Regarding the temporary nature of the priming effect, there are two important implications. The relative brevity of such effects may be responsible for some previous failures to replicate social priming effects if the test phases of those experiments were either too long or too delayed after a weak priming phase. Indeed, had we examined behavioral priming effects averaged over the session rather than analyzing the time-courses of these priming effects, we could have missed them altogether. This is consistent with other recent research that suggests that priming effects of future thought may not be as robust as previously suggested (Stins et al., 2016). Thus, the dynamic properties of priming should be taken into consideration in future studies.

Furthermore, though we observed temporary priming effects on behavior after participants received two 5-min priming sessions, the first 5-min priming phase was insufficient to cause any detectable changes in resting-state brain activity, the likely mediator of priming effects (Kounios et al., 2008). This is likely because the priming duration was too brief. Although behavioral effects could be observed after 10 min of priming, these were short-lived. This indicates that the effect of priming on problem-solving is relatively weak. Some prior priming studies have used short periods of even less-immersive priming, thus decreasing the likelihood of obtaining even subtle effects.

To summarize, growth-curve analysis showed that high-level construals engaged by distant-future thought transiently primed analytical solving while low-level construals engaged by near-future thought transiently primed insightful solving. Further research should investigate whether the direction, duration, and intensity of such priming effects are determined by specific features of the priming scenarios and whether other types of priming are similarly fleeting.

AUTHOR CONTRIBUTIONS

JK devised the conceptual idea for the study. All authors contributed to study design. JA and MK collected pilot data. MT-H and BE collected the data and performed the data processing and analysis. MT-H wrote the manuscript with support from JK.

FUNDING

This work was supported by the National Science Foundation grant 1144976.

REFERENCES

- Bowden, E. M., and Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behav. Res. Methods Instrum. Comput.* 35, 634–639. doi: 10.3758/BF03195543
- 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
- Carter, S. E., and Sanna, L. J. (2008). It's not just what you say but when you say it: self-presentation and temporal construal. *J. Exp. Soc. Psychol.* 44, 1339–1345. doi: 10.1016/j.jesp.2008.03.017
- Danek, A. H., Wiley, J., and Öllinger, M. (2016). Solving classical insight problems without Aha! experience: 9 dot, 8 coin, and matchstick arithmetic problems. *J. Probl. Solving* 9, 47–57. doi: 10.7771/1932-6246.1183
- D'Argembeau, A., Ortaleva, C., Jumentier, S., and Van der Linden, M. (2010). Component processes underlying future thinking. *Mem. Cogn.* 38, 809–819. doi: 10.3758/MC.38.6.809 doi: 10.3758/MC.38.6.809
- DeCaro, M. S. (2016). When higher working memory capacity hinders insight. *J. Exp. Psychol. Learn. Mem. Cogn.* 42, 39–49. doi: 10.1037/xlm0000152
- DeWall, C. N., Baumeister, R. F., and Masicampo, E. J. (2008). Evidence that logical reasoning depends on conscious processing. *Conscious Cogn.* 17, 628–645. doi: 10.1016/j.concog.2007.12.004
- Fleck, J. I. (2008). Working memory demands in insight versus analytical problem solving. *Eur. J. Cogn. Psychol.* 20, 139–176. doi: 10.1080/09541440601016954
- Förster, J., Friedman, R. S., and Liberman, N. (2004). Temporal construal effects on abstract and concrete thinking: consequences for insight and creative cognition. *J. Pers. Soc. Psychol.* 87, 177–189. doi: 10.1037/0022-3514.87.2.177
- Gong, H., and Medin, D. L. (2012). Construal levels and moral judgment: some complications. *Judgm. Decis. Mak.* 7, 628–638.
- Hoffman, S., and Falkenstein, M. (2008). The correction of eye blink artefacts in the EEG: a comparison of two prominent methods. *PLoS One* 3:e3004. doi: 10.1371/journal.pone.0003004
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus, and Giroux.
- Kahneman, D. (2012). A proposal to deal with questions about priming effects. *Nature*. Available at: www.nature.com/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20Letter.pdf
- Kounios, J., and Beeman, M. (2014). The cognitive neuroscience of insight. *Annu. Rev. Psychol.* 65, 71–93. doi: 10.1146/annurev-psych-010213-115154
- Kounios, J., and Beeman, M. (2015). *The Eureka Factor: Aha Moments, Creative Insight, and the Brain*. New York, NY: Random House.
- Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, M., et al. (2008). The origins of insight in resting-state brain activity. *Neuropsychologia* 46, 281–291. doi: 10.1016/j.neuropsychologia.2007.07.013
- 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
- Liberman, N., Sagristano, M. D., and Trope, Y. (2002). The effect of temporal distance on level of mental construal. *J. Exp. Soc. Psychol.* 38, 523–534. doi: 10.1016/S0022-1031(02)00535-8
- Litvak, V., Mattout, J., Kiebel, S. J., Phillips, C., Henson, R. N. A., Kilner, J., et al. (2011). EEG and MEG data analysis in SPM8. *Comput. Intell. Neurosci.* 2011:852961. doi: 10.1155/2011/852961
- Marsh, R. L., Ward, T. B., and Landau, J. D. (1999). The inadvertent use of prior knowledge in a generative cognitive task. *Mem. Cogn.* 27, 94–105. doi: 10.3758/BF03201216
- Martin, B. A. S., Gnoth, J., and Strong, C. (2009). Temporal construal in advertising. *J. Advert.* 38, 5–19. doi: 10.2753/JOA0091-3367380301
- Metcalfe, J., and Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Mem. Cogn.* 15, 238–246. doi: 10.3758/BF03197722
- Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Boca Raton, FL: CRC Press.
- Mognon, A., Jovicich, J., Bruzzone, L., and Buiatti, M. (2011). Adjust: an automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* 48, 229–240. doi: 10.1111/j.1469-8986.2010.01061.x
- Pashler, H., Coburn, N., and Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLoS One* 7:e42510. doi: 10.1371/journal.pone.0042510
- Persson, J., Welsh, K. M., Jonides, J., and Reuter-Lorenz, P. A. (2007). Cognitive fatigue of executive processes: interaction between interference resolution tasks. *Neuropsychologia* 45, 1571–1579. doi: 10.1016/j.neuropsychologia.2006.12.007
- Schooler, J. W., and Melcher, J. (1995). “The ineffability of insight,” in *The Creative Cognition Approach*, eds S. M. Smith, T. B. Ward, and R. A. Finke (Boston, MA: MIT).
- Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., and Cenac, Z. (2013). Priming intelligent behavior: an elusive phenomenon. *PLoS One* 8:e56515. doi: 10.1371/journal.pone.0056515
- Smith, S. M. (1995). “Fixation, incubation, and insight in memory and creative thinking,” in *The Creative Cognition Approach*, eds S. M. Smith, T. B. Ward, and R. A. Finke (Boston, MA: MIT).
- Stins, J., Habets, L., Jongeling, R., and Cañal-Bruland, R. (2016). Being (un)moved by mental time travel. *Conscious. Cogn.* 42, 374–381. doi: 10.1016/j.concog.2016.04.014
- Subramaniam, K., Kounios, J., Parrish, T. B., and Jung-Beeman, M. (2009). A brain mechanism for facilitation of insight by positive affect. *J. Cogn. Neurosci.* 21, 415–432. doi: 10.1162/jocn.2009.21057
- Trope, Y., and Liberman, N. (2003). Temporal construal. *Psychol. Rev.* 110, 403–421. doi: 10.1037/0033-295X.110.3.403
- Trope, Y., and Liberman, N. (2010). Construal-level theory of psychological distance. *Psychol. Rev.* 117, 440–463. doi: 10.1037/a0018963
- van der Linden, D., Frese, M., and Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychol.* 113, 45–65. doi: 10.1016/S0001-6918(02)00150-6
- Ward, T. B., Patterson, M. J., and Sifonis, C. M. (2004). The role of specificity and abstraction in creative idea generation. *Creat. Res. J.* 16, 1–9. doi: 10.1207/s15326934crj1601_1
- Wiley, J., and Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Curr. Dir. Psychol. Sci.* 21, 258–262. doi: 10.1080/17470210802479329
- Zavagnin, M., De Beni, R., Borella, E., and Carretti, B. (2016). Episodic future thinking the role of working memory and inhibition on age-related differences. *Aging Clin. Exp. Res.* 28, 109–119. doi: 10.1007/s40520-015-0368-6

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 Truelove-Hill, Erickson, Anderson, Kossoyan and Kounios. 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.



What Enables Novel Thoughts? The Temporal Structure of Associations and Its Relationship to Divergent Thinking

Peng Wang, Maarten L. Wijnants and Simone M. Ritter*

Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, Netherlands

OPEN ACCESS

Edited by:

Amory H. Danek,
Universität Heidelberg, Germany

Reviewed by:

Boris Forthmann,
Universität Münster, Germany
Alexander Christensen,
*University of North Carolina at
Greensboro, United States*

*Correspondence:

Simone M. Ritter
s.ritter@psych.ru.nl

Specialty section:

*This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology*

Received: 30 March 2018

Accepted: 03 September 2018

Published: 25 September 2018

Citation:

Wang P, Wijnants ML and Ritter SM
(2018) What Enables Novel
Thoughts? The Temporal Structure
of Associations and Its Relationship
to Divergent Thinking.
Front. Psychol. 9:1771.
doi: 10.3389/fpsyg.2018.01771

The aim of the current study is to enhance our understanding of cognitive creativity, specifically divergent thinking, by employing an interdisciplinary methodological approach. By integrating methodology from computational linguistics and complex systems into creativity research, the current study aims to shed light on the relationship between divergent thinking and the temporal structure of semantic associations. In complex systems, temporal structures can be described on a continuum from random to flexible-stable and to persistent. Random structures are highly unpredictable, persistent structures are highly predictable, and flexible-stable structures are in-between, they are partly predictable from previous observations. Temporal structures of associations that are random (e.g., dog-graveyard-north pole) or persistent (e.g., dog-cat-rat) are hypothesized to be detrimental to divergent thinking. However, a flexible-stable structure (e.g., dog-police-drugs) is hypothesized to be related to enhanced divergent thinking (inverted-U). This notion was tested ($N = 59$) in an association chain task, combined with a frequently used measure of divergent thinking (i.e., Alternative Uses Test). Latent Semantic Analysis from computational linguistics was used to quantify the associations, and methods from complex systems in form of Power Spectral Density analysis and detrended fluctuation analysis were used to estimate the temporal structure of those associations. Although the current study does not confirm that a flexible-stable (vs. random/persistent) temporal structure of associations is related to enhanced divergent thinking skills, it hopefully challenges fellow researchers to refine the recent methodological developments for assessing the (temporal) structure of associations. Moreover, the current cross-fertilization of methodological approaches may inspire creativity researchers to take advantage of other fields' ideas and methods. To derive a theoretically sound cognitive theory of creativity, it is important to integrate research ideas and empirical methods from a variety of disciplines.

Keywords: creativity, divergent thinking, associations, LSA, semantic distance, complex systems, temporal structure, interdisciplinary

INTRODUCTION

Creativity has often been defined as the generation of novel and useful insights or solutions to a problem (e.g., Stein, 1953; Runco and Jaeger, 2012). However, the question of what creativity really should be, is rather complex. Some scholars have offered conceptual frameworks that capture a wide scope of many research directions that entail creativity. One example is the Four C Model of creativity by Kaufman and Beghetto (2009, 2013) where a distinction between mini-c, little-c, pro-c, and big-c is made. This separation of four levels of creativity is mainly driven by the indirect assumption of different gradients of experiences. Therefore, mini-c focuses on developmental and transformative experiences in children and little-c more on everyday life accomplishments. For example, a child that learns to tie their shoes in a different way solved a problem in a new manner. This accomplishment wouldn't be regarded as 'creative' for an adult in their daily life routine. Pro-c, on the other hand, distinguishes accomplishments in professional settings that are transformative for certain arts or crafts (e.g., inventing a new statistical method) but is lacking the eminent accomplishment that revolutionizes the world (e.g., formulating probability theory). Those eminent accomplishments could be understood as big-c following Kaufman and Beghetto. Consequently, the Four C Model helps to embed different creative outputs in settings that are hardly comparable to another. Research assessing little-c creativity has gained considerable knowledge to this date. For example, creativity is found to be linked to intelligence, in that creative potential benefits from intelligence (or vice versa). There is evidence that creativity might benefit from intelligence up to a certain level but not above that level (e.g., Jauk et al., 2013; Karwowski and Gralewski, 2013). Others argue that intelligence might be necessary but not sufficient for creative potential and that this relationship does not stringently follow a curvilinear shape (e.g., Karwowski et al., 2016). Further, there is evidence that attentional flexibility is linked to creative potential (e.g., Zabelina et al., 2015a,b), that creativity can be trained (e.g., Scott et al., 2004; Ritter and Mostert, 2017), and that positive and negative mood moderate creative thought differently (e.g., De Dreu et al., 2008). What has remained relatively unexplored are the cognitive underpinnings and foundations of creativity. It has been proposed that a core ability in the process of generating creative solutions involves divergent thinking, which refers to the process of producing multiple answers to a problem (Guilford, 1967; Plucker et al., 2011). Divergent thinking, in turn, is believed to rely on the ability to generate remote semantic associations (Mednick, 1962; Levin, 1978; Acar and Runco, 2014). A semantic association—in this study—is the written lexical response (e.g., door) to another lexical stimuli (e.g., house) (Osipow and Grooms, 1966). Thus far, little research has been conducted to uncover the possible mechanism that allows people to express divergent thinking. Semantic associations are proposed to contribute to divergent thinking (e.g., Acar and Runco, 2014).

In a recent special issue of the *Journal of Creative Behavior*, celebrating its 50th anniversary, it has been argued that creativity research would benefit from more interdisciplinary work, encompassing different perspectives (Ambrose, 2017). In the

current study, we combined *creativity research* on divergent thinking with methodology from *computational linguistics* and *complex systems*. Computational linguistics approaches phenomena in language from a computational perspective, utilizing statistical models (Manning and Schütze, 1999). Complex systems, briefly, is the study of how parts in a system interact to create behavior that cannot be explained by examining the parts alone (i.e., an holistic approach; Bassingthwaite et al., 1994). The outline of the introduction is as follows: First, an introduction to divergent thinking and association formation with respect to creativity is presented. Second, ideas from complex systems are discussed in light of association formation in creativity. Third, techniques from computational linguistics, such as Latent Semantic Analysis (LSA), and in particular semantic distance (SmD) are introduced as a measurement technique for studying creativity. LSA has been increasingly applied in creativity research recently (e.g., Green, 2016; Hass, 2017b; Forthmann et al., 2018).

The Role of Associations in Divergent Thinking

Since Guilford (1950) introduced the idea of divergent and convergent thinking, these concepts have been very prominent in creativity research. While convergent thinking is defined as the process of finding one single, correct solution to a problem, the notion of divergent thinking is regarded as the opposite process. Divergent thinking relies on the generation of various solutions to a problem (Cropley, 2006). A widely used method to measure divergent thinking is the Alternative Uses Test (AUT; Guilford, 1950; Kaufman et al., 2008). In this test, participants are asked to generate as many ideas as possible about the usage of a commonplace object (i.e., "What can you do with a brick?"). A typical scoring scheme incorporates creativity (i.e., perceived creativity of the ideas generated), novelty (i.e., originality of the ideas), usefulness (i.e., applicability of ideas), and fluency (i.e., the number of ideas generated).

An influential theoretical model advanced by Mednick (1962) proposes that remote associations play a vital role in the formation of creative ideas. For example, Benedek et al. (2012) assessed dissociative ability and associative combination as different associative abilities with regard to divergent thinking. Dissociative ability reflects the ability to form unrelated concepts from previous concepts (e.g., summer: computer, bridge,...). Associative combination refers to the ability to form reasonable associations to seemingly unrelated concepts (e.g., summer-high: airplane, temperature,...). The authors found that dissociative ability and associative combination predicted divergent thinking, which was a composite of fluency and originality. Another interesting study from Forthmann et al. (2016) used eight different AUTs and the authors manipulated the instructions and the lemma frequencies of the objects in the tasks. That is, there was one condition that received a to "be creative" instruction while the other condition should focus on fluency. Lemma frequency or word stem frequency was varied across the objects in the AUTs so that some objects would have more natural occurrences in language and others less. Results indicated that

objects with a high lemma frequency also led to more generated ideas. The interaction between instruction and lemma frequency on the fluency of ideas revealed that objects with high frequency evoked less ideas in the “be creative” condition compared to the other one. For low frequency objects the fluency of ideas was similar in both conditions. Those findings showcase that associations play a different role depending on the task-relevant goals in the instructions.

According to Mednick (1962), individual differences in creative abilities are due to differences in their hierarchy of associations. More creative individuals should possess a “flatter” associative hierarchy, in which the strength of the associations is more similar to each other, whereas less creative individuals would show a “steeper” hierarchy. To illustrate this with an example, a highly creative individual would respond to the word “dog” with more unusual associations (e.g., work, police, and drugs) than less creative individuals (e.g., cat, pet, and bird). For a highly creative individual, distinct concepts are more closely related, in that the concept “dog–drugs” has the same associative strength as “dog–cat.” On the other hand, a less creative individual has a much stronger association between “dog” and “cat,” and a rather weak association between “dog” and “drugs.” Hence, highly creative individuals form a “flatter” associative hierarchy (dog–police = dog–cat = dog–drugs . . .), whereas it is suggested that less creative individuals have a “steeper” associative hierarchy (dog–cat > dog–police > dog–drugs . . .). Consequently, highly creative individuals should be able to access remote associates with more ease, to ultimately form a creative solution. In a similar vein, Rossmann and Fink (2010) tested students with higher creativity-related demands (i.e., enrolled in an art college) and students with lower creativity-related demands (i.e., enrolled in psychology and geosciences) in a word-pair task. The participants were instructed to judge the associative distance between indirectly related word pairs (e.g., cat–cheese) and unrelated word pairs (e.g., subject–marriage) on a 6-point scale. Results of this study indicated that students with higher creativity-related demands, compared to students with lower creativity-related demands, estimated the associative distance between unrelated word-pairs as lower and hence, more proximate to each other.

After reviewing inconclusive studies, Benedek and Neubauer (2013) conducted an experiment to test Mednick’s assumption. The authors used a continuous free association task in which associations had to be created for six predefined words within 60 s per word. Subsequently, participants were categorized as high or low creatives based on their performance on two other divergent thinking tasks (i.e., unrelated to the free association task). The results indicated that individuals scoring high on creativity, as measured by the two divergent thinking tasks, also formed more associations and more uncommon responses. The authors could not find evidence that corroborates differences in associative hierarchy and concluded that hierarchy does not contribute to divergent thinking. However, there is ample evidence suggesting that, for example, the semantic networks of high versus low creative people inherit different properties. Kenett et al. (2014, 2016) used methodologies from network science to test Mednick’s hypothesis and found that highly creative people have a denser,

less modular (less sub-parts) and more connected semantic network. Hence, high creatives are supposed to access remote associations more efficiently, with more interconnections and shorter routes between two or more concepts.

We discussed that associations may play a vital role in divergent thinking. Further, it has been theorized that the structure of those associations is crucial for divergent thinking (Mednick, 1962) and there are indications that properties of semantic networks might be related to divergent thinking (e.g., Kenett et al., 2014). To further explore the theorized associative structure underlying divergent thinking, this study investigates the temporal structure of associations. Temporal structure, in this case, refers to the change of associations over time (e.g., dog–cat–milk–supermarket. . .). In other words, how are differently organized “chains of thoughts,” which unfold over time, related to divergent thinking? It is hypothesized that different temporal structures of associations underlie the ability to perform well in divergent thinking. This idea will be elaborated in more detail after complex systems and SmD have been introduced. In the following part, ideas and methodologies from complex systems will be discussed.

Complex Systems

Complex systems are mainly embraced in mathematics and natural sciences to describe processes that originate in nature. What characterizes these systems are their non-linear, dynamical and mutual influential, and interdependent properties (Friedenberg, 2009). The core assumption is that knowledge about only one part of a system will not lead to knowledge about the overarching behavior of the system. Consider for example the case of population swings in a predator–prey model as an illustration (Berryman, 1992). Let there be a population of foxes (predator) and a population of rabbits (prey) interacting in an environment. If we observe both population changes over time, the number of rabbits and foxes will vary in a non-linear way. It will be the case that the number of rabbits decreases if the number of foxes increases. However, if there are very few rabbits left, the population of foxes will decline dramatically as there is not enough prey. This is accompanied by a minor increase of rabbits at first, which is then followed by an exponential grow of rabbits. By gaining knowledge about only the rabbits (physiology and ethology) or only the foxes, one cannot explain the change in population over time. In order to understand the change in population (either rabbits, foxes, or both), the whole interacting system has to be observed. This example illustrates that as little as two variables or parts of a system can produce behavior that is neither linear, nor independent, as these two parts of the system mutually interact.

Another crucial aspect is the temporal order of observations in a complex system. The temporal order is important as parts of a complex system are mutually influencing one another (which changes over time). An analogy will exemplify the reasoning behind this idea. Assume we test three players in a skill game where they throw a tennis ball in one of five bowls, representing 1–5 points. Each player has a different skill level/technique and throws for 100 trials. We call these 100 trials a *time-series*. Player A randomly hits the bowls (e.g., instruct a computer to generate

a throwing scheme). Player A ends up with an average of 3 points across all trials. Player C would display a very persistent throwing order. He would decide to hit the bowls in an ascending order (e.g., hit bowl 1, then 2, then 3, . . .), repeating the pattern until he reaches 100 trials. Thus, he would be very persistent in his throws, but his average would also be 3 points. Player B is flexibly varying his throws in that he sometimes hits one bowl more often, and then switches. However, he is neither random nor very persistent in his actions, so that the pattern of throwing is not obvious at first glance. He also ends up with an average of 3 points. Further, all three players would deviate from the mean by approximately 1.4 points. With classical frequentist statistics, we would not be able to discern the underlying difference between the three players. Data would be pooled and the temporal structure of the time-series was lost. It is apparent that the temporal order is substantially different between the three players (see **Figure 1**, right panel). The order or strategy of player A (random) would substantially differ from player C (very persistent) or B (flexible-stable). Therefore, the temporal structure or dynamics of the time-series should be taken into account to capture the whole picture. Models from complex systems adhere to the temporal structure in a time-series and preserve them in the analysis.

Mathematically, the temporal structure of time-series can be distinguished into at least three classes of patterns. Those patterns are typically called either random, persistent or flexible-stable (Wijnants, 2014). Notice that this is a continuum, where randomness is the one extreme and persistence the other, with flexible-stability residing in the middle balancing random and persistent processes out. Processes in a time-series constitute change over time, they reflect what happened at time 1, time 2, time 3, and so on. A time-series can be regarded as random when the next observation (e.g., time 4) can hardly be predicted from the previous data (e.g., time 1–3). Thus, it does not build on previous observations. A time-series can be regarded as persistent when the next observation (e.g., time 4) can be determined with high certainty from the previous data. Thus, it strongly builds on previous observations. A time-series can be regarded as flexible-stable when the next observation (e.g., time 4) can be determined with moderate certainty. That is, higher than random and lower than persistent. Thus, it does partly build on previous observations but can vary to some extent. Linking to the previously mentioned ball game, each player's time-series could be described in the continuum from random (i.e., player A) to flexible-stable (i.e., player B) to persistent (i.e., player C).

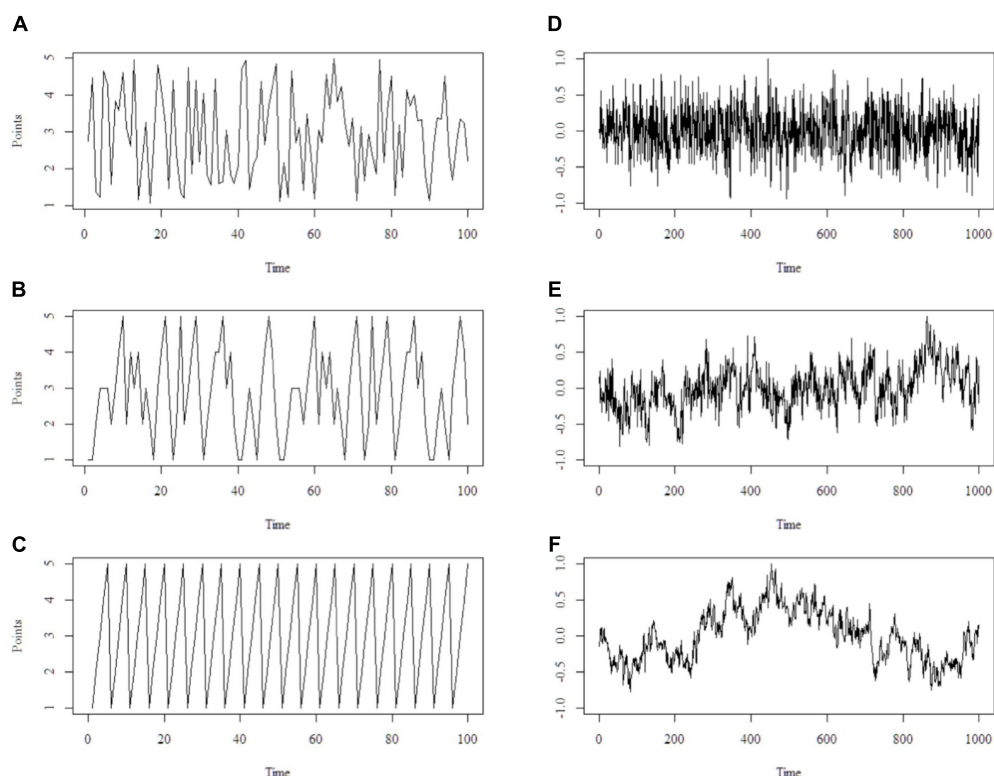


FIGURE 1 | Schematic illustration of the outcomes of the ball-throw game on the left, points are displayed on the y-axis and time on the x-axis. Throw pattern of player A is random (**A**), throw pattern of player B is flexible-stable (**B**) and throw pattern of player C is persistent (**C**). On the right, three different time-series that correspond to random pattern (**D**), flexible pattern (**E**) and persistent pattern (**F**) are shown. Those time-series reflect self-similarity across different scales. That is, the pattern of (**E**) observed over 1000 trials (bigger scale) is already reflected in, for example, the first 200 trials (smaller scale) or trials 400 to 800 (intermediate scale). It is self-similar as e.g., the first 200 trials of (**E**) are resembling the whole time-series of (**E**). Put differently, the first 200 trials of (**E**) “look” similar to trial 0–500 or 0–1000 of (**E**). Notice that not only the geometrical shape but also statistical properties are similar across scales.

Complex Systems in Physiological and Psychological Research

Research indicates that the functioning of humans also obeys temporal regularities. Physiological examples show that the time-series of performances on a motor task is associated with flexible-stable patterns (Wijnants et al., 2009) and likewise does a healthy heartbeat fluctuation reflect a flexible-stable pattern (Van Orden et al., 2011) whereas abnormal fluctuations will not (Peng et al., 1995; Goldberger, 1997). Studies on cognitive processes show that the reaction-time of mental rotation (Gilden and Hancock, 2007), word naming and simple reaction tasks (Van Orden et al., 2003) are associated with flexible-stable patterns. Accordingly, it has been proposed that those processes which lie in between persistent and random can be regarded as optimal (Corona et al., 2013).

Interestingly, it has been argued that complex systems might be a promising approach to study creativity. For example, Piccardo (2017) examined the potential role of plurilingualism and the associated dynamical engagement with one's language and environment as a beneficial factor in creative thought. Further, Poutanen (2013) suggests that different levels of inquiries (i.e., individual level, group level and organizational level) about creative phenomena can be embedded within a complex systems framework. The present study investigates a time-series reflecting a chain of associations. Put differently, it examines how associations unfold and change over time. By implementing methodology from a complex systems perspective, the temporal structure of this chain of associations can be inferred. Distinctions can be attributed to a chain or structure of associations to discern random, flexible-stable or persistent patterns. In order to utilize this approach, a time-series is needed. In the following part of the introduction, we will address the possibility to quantify a time-series from a chain of written associations.

Computational Linguistics, Latent Semantic Analysis, and Semantic Distance

One approach in computational linguistics involves the quantification of word similarities using large amounts of texts. This line of reasoning relies on the distributional hypothesis, stating that words with similar meaning tend to occur together in a similar context (Harris, 1954; Sahlgren, 2008). A prominent method for modeling word similarities is LSA (Deerwester et al., 1990; Landauer and Dumais, 1997). LSA is a computational method that allows the user to compress substantial quantity of texts and retrieve their word meanings or semantics. It does so by creating a highly dimensional spatial space (*semantic space*) where semantic concepts are represented as points (vectors) in this space. Subsequently, it is possible to infer the relative position between, for example, two words in this semantic space by calculating the cosine (number between 0 and 1) of the angle of those two points. For instance, there are studies showing that the similarity between words, expressed by the cosine in LSA, significantly predicts reaction time in lexical priming experiments (e.g.,

Jones et al., 2006; Hutchison et al., 2008; Günther et al., 2016).

Recently, researchers in the field of creativity have started to measure creative performance by using LSA to estimate the similarity between words as SmD (e.g., Green et al., 2006; Beaty et al., 2014). SmD is now defined as the inverted cosine ($1 - \text{cosine}$) of two semantic concepts (two points or words in the semantic space), where higher decimals represent more dissimilarity and lower decimals represent more similarity between two concepts. Hence, the higher the SmD of two concepts, the less common they are, and vice versa. For example, the concept of “university” and the concept of “cook” do not share much common ground (SmD of 0.87, greater distance and lower similarity). If you now compare “university” to “study,” the SmD drops as those concepts are more related to each other and appear together more often (SmD of 0.44, smaller distance and higher similarity).

Evidence and Validity of Semantic Distance as a Measure for Creativity

In an extensive study by Hass (2017a), responses from two AUTs were quantified using LSA. The author found, for example, that the SmD from the responses, directed to the target concept (i.e., brick and bottle), was non-linearly increasing as the task progresses. As a result, this means that participants generated associations more closely related to the target concept at the beginning but that this relatedness slowly decays as the process continues. Similarly, per-trial response time was positively correlated with SmD between adjacent responses and creativity scores provided by raters. In another paper by the same author, a large data set with divergent thinking tasks was reanalyzed with LSA. As in the previous study, responses from different AUTs were first quantified using LSA and then analyzed in a regression model. The results indicated that subjective creativity ratings were positively predicted by SmD. That is, more dissimilar concepts or ideas of the AUTs were associated with more creative ratings (Hass, 2017b). Another line of evidence comes from neuropsychological studies. For example, Green et al. (2010) manipulated SmD of word pairings and measured (left) frontopolar cortex activity. The (left) frontopolar cortex is believed to be involved in the process of analogical reasoning, which plays a vital role in innovative outcomes. The authors showed that a higher SmD was associated with greater frontopolar cortex activity. In another study by the same group of researchers (Green et al., 2015), participants were now tested in an analogy generation task. Here, a verb had to be generated to a noun shown to the participants. Crucially, half of the trials were cued, signaling that a creative verb should be formed. Explicit instruction to think creatively has previously been shown to be effective in enhancing creative performance (e.g., Runco and Okuda, 1991; Chen et al., 2005; Hong et al., 2016). Results indicated that SmD was significantly higher for cued trials than for uncued trials. Further, for cued trials, the (left) frontopolar cortex exhibited increased activity. As a direct extension to those results, stimulating the frontopolar cortex through transcranial Direct Current Stimulation has been observed to enhance performance in the same cued analogy task

(Green et al., 2016). SmD has also been used directly as an outcome variable for creativity in a study testing a large online sample (Weinberger et al., 2016). Here, an analogy finding task was used and the authors were able to find that SmD was higher for blocks that were paired with the instruction to think creatively than without the instruction.

Finally, construct validity for SmD as a measure for creativity was provided by Prabhakaran et al. (2014). The authors showed that SmD positively correlates with creativity measurements, specifically with a divergent thinking task, story writing task, the Torrance figural test and the Creative Achievement Questionnaire. Based on these findings, SmD has been suggested as a novel measurement tool to reliably measure creativity. Other work on the reliability of LSA in creativity research has been done by Forthmann et al. (2018) who revisited several studies which applied LSA to quantify responses from divergent thinking tasks. They argued that estimations of SmDs are potentially biased due to response length (i.e., multiple words) and conducted a simulation study. When responses were removed from stop words and corrected for biases, the authors found the correlation between SmD and creativity ratings in divergent thinking tasks to be highest.

In the present study LSA and SmD is utilized to quantify a time-series of a chain of associations. Hereby, the change over time in the similarity of a chain of associations can be inferred by implementing methodologies from complex systems. That is, a persistent pattern in a chain of associations would reflect similar concepts that build closely on previous concepts (e.g., dog–cat–milk–cow, etc., hence, all low SmD). A random pattern would reflect extremely dissimilar concepts that are loosely related to previous concepts (e.g., dog–graveyard–north pole–whisky–etc., hence, all high SmD). Lastly, a flexible-stable pattern would reflect dissimilar concepts that build on previous concepts (e.g., dog–police–helicopter–fan–etc., hence, intermediate SmD).

The Present Study

It was discussed that the ability to generate associations plays a vital role in divergent thinking. That is, associative abilities (e.g., combining distinct associations) contribute to divergent thinking (Benedek et al., 2012). Further, Mednick (1962) suggested that the associative hierarchy is crucial, which received support from studies involving semantic networks (Kenett et al., 2014; but for an exception, see Benedek and Neubauer, 2013). This study proposes that the temporal structure of associations is related to divergent thinking. Temporal structure here refers to the change of generated (written) associations over time. By applying methodologies from complex systems, a time-series can be characterized on a continuum from random to persistent. LSA and SmD will be used to quantify a testable time-series, where the association between two concepts serves as observations that unfold over time. More precisely, it is hypothesized that a flexible-stable (in between random and persistent) temporal structure of associations is linked to the highest performance on the creativity and originality dimension of divergent thinking. It is assumed that associations that are too random would make less sense and are therefore regarded as less creative (e.g., from “dog” to arbitrary concepts like “graveyard”). On the other hand, too

persistent associations do not create novel and creative insights as they too strongly rely on previous associations (e.g., from “dog” to “cat”). In the middle, where flexible-stable associations are found, it is expected that concepts are unrelated enough to create novel and thus creative insights (e.g., “dog” and “drugs”).

Hence, novelty and creativity in divergent thinking are hypothesized to be predicted by a quadratic term (or inverted U) of the temporal structure of associations. In a novel paradigm, where a chain of associations had to be formed, SmD was calculated using LSA. Subsequently, the temporal structure was inferred through Power Spectral Density (PSD) analysis and detrended fluctuation analysis (DFA, described in the Methods section). The Brick Version of the AUT served as a measure for divergent thinking. Additionally, it is expected to observe a positive relationship between mean SmD, originality and creativity in the divergent thinking task. This would add to previous findings on the validity of SmD as a creativity measure (e.g., Green, 2016) To complement the analysis, a convergent thinking task and a real-life creative achievement questionnaire are added for explorative analysis.

MATERIALS AND METHODS

Participants

Participants were 59 students (41 female) from the Radboud University Nijmegen with a mean age of 21.95 years ($SD = 2.58$). All participants were of German nationality and spoke German on a native level. Participation was voluntary and rewarded with either €7.50 or credit points, which were to be obtained as part of the participants' curriculum. One participant was excluded from the descriptive and correlation analysis ($N = 58$), however, the main analysis ($N = 59$) included all participants. Exclusion was based on unreasonable scores on the CAQ (87, highest score for all categories, fulfilled several times). Descriptive scores are shown in Table 1.

Divergent Thinking, the Alternative Uses Test (AUT)

A widely used test for divergent thinking is the AUT (Guilford, 1950; Runco and Acar, 2012). The brick version of the AUT

TABLE 1 | Descriptive statistics of creativity measures and PSD, DFA.

Variable	<i>M</i>	<i>SD</i>	Range
AUT			
Creativity	2.29	0.44	1.38–3.79
Novelty	2.43	0.48	1.38–3.79
Usefulness	3.51	0.57	1.61–4.50
Fluency	7.47	3.19	2–16
CRAT	4.27	2.59	0–11
CAQ	5.10	2.60	0–11
PSD	–1.24	0.48	–2.49 – –0.26
DFA	0.54	0.08	–0.39–0.79

N = 58 for CAQ statistics as one participant was excluded due to an unreasonable CAQ score of 87. N = 59 for all other variables.

was used in the present study and was introduced as an idea generation task to the question: “What can you do with a Brick?–List your ideas below”¹. The task was fully computerized and participants were instructed to insert their answers in an empty text box. In total, the task lasted for 3 min and was automatically terminated afterward. Responses of the AUT were rated on the creativity, novelty and usefulness dimension by two judges. Judges were instructed to first get an overview of the responses and be in a neutral mood when rating. While rating, scores should be consistent (e.g., same ideas should receive the same score) and their focus should lie only on the respective dimension to be rated. As for creativity, judges should follow their first impression on how creative they perceive the idea to be. For novelty, judges should evaluate the ideas on how novel and unique they are. For usefulness, judges should evaluate the ideas on how well they think the idea will work and can be implemented. Each item in the respective dimension was to be rated on a Likert-scale from 1 to 5, where 1 was the lowest and 5 the highest score (i.e., 1 = not at all creative and 5 = very creative). ICC was calculated with the ICC function of the psych package (Revelle, 2016) in R (R Core Team, 2016). For all rating dimensions, the intraclass correlation (ICC2k) was excellent ($AUT_{creativity} ICC = 0.92$; $AUT_{novelty} ICC = 0.87$; and $AUT_{usefulness} ICC = 0.94$). Accordingly, the average score of the two raters were used for further analysis. *Fluency* was determined by counting the number of answers that were provided.

Convergent Thinking, the Compound Remote Associate Task (CRAT)

To measure convergent thinking, the German version of the CRAT, validated by Landmann et al. (2014), was used. In the CRAT, participants are asked to find a matching word that relates to three other words previously mentioned. For example, the solution for the triplet “cottage–Swiss–cake” is “cheese.” The CRAT in this study consisted of 20 randomly chosen (and matched by difficulty) triplets using the sample function of the core package in R. All 20 triples were simultaneously shown. Participants had 5 min to solve as many CRAT items as possible. The CRAT was scored according to the solution scheme provided by Landmann et al. (2014). Every response was manually checked for correctness (by the researcher) and the final score was the sum of all correct answers.

Real-Life Creativity, the Creative Achievement Questionnaire (CAQ)

The CAQ measures real-life creative achievement, and has been validated by Carson et al. (2005). The original English version of the CAQ was administered without a time constraint and included ten categories (with each eight response options) covering different real-life domains. Those domains were: visual arts, music, dance, architectural design, creative writing, humor, inventions, scientific discoveries, theater and film, and culinary arts. All 10 domains were shown at the same time, and response

options ranged from 0 to 7, where 7 was the highest score. The CAQ score is the sum of all questions, where in each domain the highest score (if applicable) is multiplied with the amount of times it was fulfilled. For example, in the theater and film domain the most extreme answer is: “*My theatrical work has been recognized in a national publication.*” (7 points). If someone had fulfilled this condition several times (e.g., 3) than the score gets multiplied (i.e., $3 \times 7 = 21$). The CAQ was scored in accordance to Carson et al. (2005).

Association Chain Task (ACT)

The ACT is a task designed to capture semantic associations in order to first calculate the mean SmD, and to subsequently derive scaling estimates from its times-series. We used an adapted version from Benedek et al. (2012). Therefore, a customized computer script in python (version 2.7.12), using mainly the PsychoPy package 1.82.1 (Peirce, 2007), was created. The script displayed the instructions and enabled participants to insert written responses on a computer. In the ACT, participants were asked to generate a chain of associations. A definition of associations was provided as follows: “Think of associations as ideas or thoughts between two (or more) concepts.” Accordingly, the task was to repeatedly form an association to the previous word they had generated and continue to do so until the program indicated to stop. As an illustration, let us assume that the starting word was “airplane.” Now the participant had to form an association, for example “vacation,” to the previous word “airplane.” The next association, for example “beach,” had now to be based on the previous word, which is “vacation.” This procedure was now to be repeated. One example has then been displayed to the participant (i.e., apple–tree–leaf–bird), followed by the instruction to only use words which could be found in a dictionary (e.g., no names of friends or actors, book titles, movies, etc.). This was due to the restrictions of the LSA. Then, a practice block of four trials was provided. Hereafter, it was stated that the target block will start. The first word for all participants was “house.” Participants inserted their responses using a keyboard and continued each trial by pressing the “Enter” button. In each trial, only the previously inserted word was shown, together with the instruction to form an association to it. Importantly, it was also mentioned that participants should “try to be creative” in their responses, which was displayed during every trial in the target block. All written responses were recorded within a time window of 35 min. This duration was chosen based on a pilot to maximize the number of trials while considering the tiresome characteristic of the experiment. There were no time restrictions between the trials. That is, participants were free to “think” as long as needed. The task was automatically terminated hereafter.

LSA and SmD

LSA was used to derive the SmD from the associations formed in the ACT. Conceptually, in LSA a semantic space is created by counting word frequencies of a large body of documents. Accordingly, the first step of LSA is to count the occurrence and co-occurrence of words in documents. That is, each word reflects a row and each document reflects a column in a large matrix. The cells of this matrix are then populated with the

¹Notice that some researchers do include an explicit instruction to “be creative” in AUTs (e.g., Forthmann et al., 2016; Harrington, 1975) which has effects on the results of divergent thinking tasks.

number of occurrence of the respective word in that respective document. This matrix is then transformed so that less frequent words have an increased impact, since less frequent words normally convey more detailed and specific meaning, and more frequent words decrease in their impact. Then, a so called, singular value decomposition is applied to the matrix. A method that shares conceptual resemblance with principal component analysis, where factors are formed. Lastly, the dimensions or factors are reduced (mostly to around 300 dimensions) to remove noise and redundant dimensions from the matrix. The result is a (still) highly dimensional semantic space, where words are represented as vectors in this space. The SmD is now defined as the inverted cosine of the angle between two words or vectors in the space. A value of one is interpreted as unrelated words, whereas zero indicates identical words. Thus, words with similar meaning tend to have low values and vice versa. It is possible to receive negative values. However, these values cannot be interpreted and are usually set to one (Deerwester et al., 1990; Landauer et al., 1998; Günther et al., 2015). Subsequently, a mean SmD per participant was calculated to analyze its correlation with the other creativity measures.

Temporal Structure and Complexity Estimates

The techniques used to infer the temporal structure and to calculate the complexity estimates were all applied in MATLAB. One method to estimate the temporal structure is found in PSD (Gilden et al., 1995). The purpose of PSD is to calculate an estimate of the fractal dimension which informs us about the temporal structure of a time-series (i.e., from random to flexible-stable to persistent). Fractal dimension refers to the presence of self-similar patterns across multilayered scales (see **Figure 1**, left panel). PSD functions most reliable with large time-series consisting of any number that is the power of 2^n (e.g., ... 1024, 512, 256, 128 ...). Conceptually, in PSD a time-series is transformed into a linear combination of sinus waves, called Fourier transformation. The result is a summation of all frequencies and amplitudes of the time-series. All frequencies and amplitudes are log transformed and plotted with (log) frequency on the x-axis and (log) amplitude on the y-axis. The best-fitted line (linear regression) represents the fractal dimension where a slope of -1.0 reflects a (perfect) flexible-stable structure, 0 reflects a (perfect) random structure and -2 reflects a (perfect) persistent structure. The PSD analysis was conducted using the PSD function, available in the Signal Processing Toolbox in MATLAB (The Math Works, 2017).

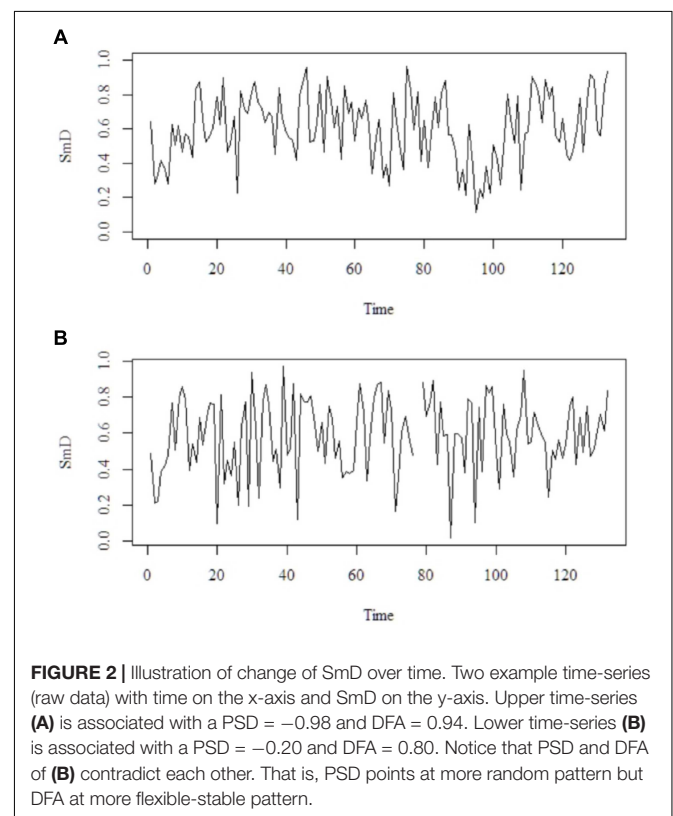
Detrended fluctuation analysis is another method to inquire the same question in the time domain (Peng et al., 1994). It takes a time-series and computes the cumulative sum. In the present study, the cumulative sum of the SmD time-series is taken. Then, the new time-series of the cumulative sum is divided into several windows with different lengths. For example, a time-series with 100 data points could be divided into 4 (windows) \times 25 (length), 5 \times 20, 10 \times 10, and so on. For each window, a slope (linear regression) is fitted which

represents the “local trend.” The “global trend,” which is the regression of the whole time-series, is then subtracted from each “local trend” and hence, detrended. Now, the standard deviation in each window is calculated where the mean is taken and log transformed. Ultimately, the log mean is plotted against the log window sizes and the best fitted line or slope (as in PSD) represents the fractal dimension. A slope of 1 reflects a (perfect) flexible-stable structure, 0.5 reflects a (perfect) random structure and 1.5 reflects a (perfect) persistent structure. The DFA function for MATLAB can be found on <https://github.com/FredHasselmann/toolboxML/blob/master/Ddfa.m> (Schmidt, 2001).

These methods are complementary in that the strengths of one compensates for the weaknesses of the other. For instance, PSD, while robust in many respects, requires preprocessing of the signal because extreme observations can contaminate the outcome of the analysis (see Holden, 2005). DFA can be applied to non-stationary signals and is not susceptible to most statistical artifacts or long-term trends, but it can falsely classify certain types of signals as fractal (Rangarajan and Ding, 2001). Finally, each participant received a PSD and DFA estimate in addition to the mean SmD to their time-series derived from the ACT (see **Figure 2** for an example of the time-series).

Procedure

Because of the demanding characteristic of the association task (ACT), it was chosen to split the study in two parts. In part



one, an online experiment administered in *Qualtrics*, participants had to complete the AUT, the CRAT, and the CAQ. Moreover, demographic information was assessed. Upon completion of part one, part two, which took place in a lab setting, was scheduled. In the lab study, the ACT was administered and participants had to form a chain of associations. The first word “house” was provided by the program, participants then started to generate associations for 35 min. After the task was finished, participants were thanked, rewarded and debriefed if wished so.

Data Aggregation

LSA and SmD

Before conducting the main analysis, responses of the ACT had to be preprocessed. Therefore, an R script was written in which all responses (10722 words generated by 59 participants) were first cleaned from unwanted characters (i.e., whitespace, special characters, and upper case characters). Hereafter, the SmD was calculated using the Cosine function of the LSAfun package (Günther et al., 2015). The semantic space (“dewak100k_lsa,” a semantic space of the German language) was retrieved from <http://www.lingexp.uni-tuebingen.de/z2/LSAspaces/>, which was created by the package maintainer. In the first iteration, there were 2627 instances where a SmD could not be calculated. This was mostly due to typos but also due to words (very rare words or compound words) not present in the semantic space. Typos were identified (1624 misspelled words) using the `hunspell_check` and `hunspell_suggest` function of the `hunspell` package (Ooms, 2017). Those words were manually corrected (for a complete list see the **Supplementary Table 1**). In a second iteration, SmDs were newly calculated with the updated data file, which decreased the number of missing values to 857 (8% of all words).

Temporal Structure and Complexity Estimates

Before calculating the complexity estimates, the time-series of the SmD for each participant was ensured to be an integer power of 2^n (e.g., 64, 128, 256, 512, ...). Some algorithms require the length of a time-series to be an integer power of 2. Further the algorithm works faster if this requirement is met. As many time-series did not obey this rule, we used zero-padding to guarantee the length of any time-series to be an integer power of 2. This was done by adding zeros at the end of the time-series. For example, if a time-series had 115 data points, 13 zeros were added at the end to obtain 128 observations (or e.g., $230 + 26$ zeros). This, so called zero-padding, is assumed to have no distorting effect on the complexity estimates (Holden, 2005).

RESULTS

Main Analysis

To test the main hypothesis that novelty and creativity in the AUT can be modeled as a quadratic function of PSD and DFA, two multiple regressions, one with $AUT_{novelty}$ and the other with $AUT_{creativity}$ as dependent variable and PSD and DFA

(and their quadratic transformations) as independent variables were conducted. The regression models for $AUT_{creativity}$ and for $AUT_{novelty}$ were non-significant, $F(4,52) = 2.01$, $p = 0.11$, $R^2 = 0.07$ and $F(4,52) = 1.52$, $p = 0.21$, $R^2 = 0.04$, respectively. No linear or quadratic effect were found, and the main hypothesis was not confirmed. Notice that PSD and DFA was not found to be correlated in this study ($r = -0.03$, $p = .82$), which is unusual. For example, in more repetitive tasks the observed correlation were rather high and around $r = 0.8$ (Wijnants et al., 2012). We further discuss this issue in the exploratory analysis section.

Correlation Between Creativity Measures and Semantic Distance

To assess the bivariate correlation between the AUT, CRAT, CAQ, and SmD, Pearson correlations were calculated (see **Table 2**). The creativity and novelty dimension of the AUT were significantly positively related with each other, usefulness was negatively related to them and to fluency. This is in line with previous research (e.g., Diedrich et al., 2015), in that more useful ideas are usually rated less novel and less creative. On the other side, the more novel an idea is, the more creative it will also be rated. Surprisingly, SmD was not related to the creativity measures, which contradicts previous research (e.g., Prabhakaran et al., 2014; Green, 2016).

Exploratory Analysis

Previous research successfully made use of SmD as a creativity measure (e.g., Green et al., 2010; Beaty et al., 2014; Prabhakaran et al., 2014). However, the design in these studies was different to the design used in the current study. For example, the experimental duration was approximately 9 min in Prabhakaran et al. (2014) over 72 trials, but 35 min in the present study. To test whether the mean SmD of different task durations (9, 5, and 2 min) would be associated with the creativity measures in this study, bivariate correlations were calculated. All three SmD task durations remained non-significant in relation to $AUT_{usefulness}$, $AUT_{novelty}$, $AUT_{creativity}$, $AUT_{fluency}$, CRAT, and

TABLE 2 | Bivariate correlation of the AUT, CRAT, CAQ, and SmD.

	AUT						
	Creativity	Novelty	Usefulness	Fluency	CRAT	CAQ	SmD
AUT							
Creativity							
Novelty	0.93***						
Usefulness	-0.28*	-0.40**					
Fluency	0.19	0.28*	-0.28*				
CRAT	0.27*	0.20	-0.03	0.14			
CAQ	0.04	0.04	-0.07	-0.02	0.19+		
SmD	0.00	0.02	0.02	0.02	-0.07	0.02	

N = 58 as one CAQ scores was unreasonable, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and + $p < 0.10$.

CAQ (all $p > 0.05$). That is, neither was the SmD of the first 9 min related to any creativity measure, nor the SmD of the first 5 or 2 min. These results would strongly argue against SmD as a measure for creativity. However, SmD was significantly predicted by the mean response time in a linear regression $F(1,57) = 14.38$, $p < 0.001$, $R^2 = 0.20$, and $\beta = 0.04$, in that longer response time was associated with higher SmD. This is also often described as the serial order effect stating that more creative outcomes tend to appear after increased amount of time (e.g., Christensen et al., 1957; Beaty and Silvia, 2012). This suggest that SmD is likely to reflect the similarity of semantic concepts, as the longer someone thinks, the more uncommon, distinct and hence less similar the response should be.

For the complexity measures, the data showed no significant correlation between PSD and DFA, $r(57) = -0.03$, $p = 0.83$. This is unusual as PSD and DFA are estimating the same relationship and should therefore corroborate each other. A reason could be the variability in the length of the time-series, which varied from as low as 42 to more than 370 data-points. PSD and DFA tend to be more reliable with more observations, while PSD performs best with 2nd observations (Delignieres et al., 2006). To introduce a more robust measure, also for smaller time-series, the *sample entropy* was calculated (Richman and Moorman, 2000). The implementation in MATLAB can be found on <https://www.physionet.org/physiotools/sampen/matlab/1.1-1/sampenc.m> (Lake et al., 2008). Sample entropy is another method to infer the fractal dimension of a time-series, where the higher the sample entropy the more random a system is. However, it is a relative method in that no absolute statements (i.e., time-series A reflects a flexible-stable pattern) can be made. That is, different time-series can be described in their structure to each other in that particular sample (time-series A is more random than B). A bivariate correlation between sample entropy and all creativity measures (AUT, CRAT, and CAQ), mean response time and mean SmD was conducted. Sample entropy did not significantly correlate with AUT_{creativity}, $r(57) = 0.20$, $p = 0.13$, or AUT_{novelty}, $r(57) = 0.19$, $p = 0.15^2$. Which further argues against a relationship between the temporal structure of associations and divergent thinking. Interestingly, the correlation between sample entropy and mean SmD was found to be significant, $r(57) = 0.50$, $p < 0.001$. This would indicate that the more random the temporal structure was, the higher the SmD. Although a significant correlation could be detected, caution is advised. There was no predefined hypothesis, and multiple testing inflates type I errors. Moreover, sample entropy was not correlated with PSD, $r(57) = 0.03$, $p = 1$ or DFA, $r(57) = 0.01$, $p = 1$.

DISCUSSION

This study investigated the temporal structure (i.e., random/flexible-stable/persistent structure) of associations

and its relationship to a core component of creativity, divergent thinking. It was hypothesized that novelty and creativity in divergent thinking, as measured by the AUT, would be a quadratic function of the temporal structure of the associations. That is, random and persistent structures of associations were assumed to be related to less novelty and creativity ratings, whereas flexible-stable structures of associations would predict high novelty and creativity ratings on the AUT. The current findings provide no evidence for the hypothesis that the structure of associations is related to an individual's potential for divergent thinking. There was neither a linear nor quadratic trend found. Initially, this would imply that different structures of associations do not contribute to the ability to generate novel and creative responses. The temporal order of how each association leads to another was irrelevant. Hence, any sort of temporal structure would equally enable people to utter creative behaviors, and a more random structure of associations would be found to display the same relationship with divergent thinking as a persistent structure. If this holds true, it would not matter which structure of associations someone possess, distinct streams of thought would play no role in creativity. This would be in line with Benedek and Neubauer (2013) who found that associative hierarchy is not predictive for divergent thinking (cf. Mednick, 1962).

On the other hand, previous research on semantic properties and creativity were found to corroborate the idea that there might be a relationship (e.g., Forthmann et al., 2016; Hass, 2017a). Furthermore, studies on semantic networks suggest the same, network properties do influence divergent thinking abilities (Kenett et al., 2016, 2017). In a recent study by Kenett and Austerweil (2016), it was tested whether a simulated “search” over modeled semantic networks of more and less creative individuals would lead to different results. Results indicated that, indeed, a simulated “search” in the semantic network of more creative individuals yielded more unique words. Hence, there is growing evidence in the literature that divergent thinking benefits from distinct characteristics in semantic structures. The present study examined the role of the temporal structure of association, which is not fully comparable to associative hierarchy but more relatable to semantic networks. Considering the characteristics of different temporal structures in complex systems (i.e., random, flexible-stable and persistent) one would reason that, e.g., a persistent structure of associations will not facilitate divergent thinking. That is, if every next association heavily builds on the earlier association (e.g., dog-cat-mouse-cheese-etc.) creative thoughts are rare or slowly to appear. Too random structures will presumably not connect concepts meaningful enough. Flexible-stable structures, in turn, would enable new associations to arise which still form enough coherence to previously generate thoughts (see, e.g., Kenett et al., 2018). If the results are indeed trustworthy, those ideas could be questioned. However, there are reasons to believe that the methodology in the current study did not truly capture the temporal structure, which is laid out in the next paragraph.

²However, there is the possibility that our sample had not enough power to detect the potentially meaningful correlations as they reside between small and medium effect sizes.

Limitations

Temporal Structure of Associations

There are indications that the current results might be less reliable due to noticeable deviations in the data. Firstly, PSD and DFA were not correlated (also not with sample entropy), which could mean that those measures were not capable of estimating the temporal structure of the time-series. One reason might lie in the fluctuation in the number of observations (42 to more than 370) within the time-series and the missing data (8% on average) (Delignieres et al., 2006). Another reason could be the operationalization of associations. It was hypothesized that SmD would capture change in a cognitive process (association formation). PSD and DFA are suggested to reveal natural processes which have been successfully implemented in biologically sound concepts such as heart rate variation (Van Orden et al., 2011) or reaction time (Van Orden et al., 2003). Those processes are outcomes of a natural system. On the other hand, the current research made use of SmD as a proxy for a natural process which was the change of association forming. Because SmD is based on a computational method (LSA), it is likely that it is not an inherently natural and ontologically concise cognitive outcome. When applying techniques to infer the fractal dimension, it is not guaranteed that the result will reflect a true temporal structure of a natural system.

SmD and Creativity

SmD was not related to any creativity measure, which strongly contradicts the literature on SmD and creativity (e.g., Beaty et al., 2014; Prabhakaran et al., 2014; Weinberger et al., 2016). Even after considering the greater length of this experiment (previous studies measured SmD in shorter designs), that is, assessing the SmD of the first 9, 5 and 2 min, no relationship with any creativity measure was found. Thus, it is unlikely that the length of the experiment confounded the correlation between SmD and creativity measures. As other studies confirmed the effective application of SmD (e.g., Green et al., 2010; Beaty et al., 2014; Prabhakaran et al., 2014), it is highly likely that another feature of the design in this experiment confounded the effect. For example, the ACT challenged the participants to form associations based on the previous concept. In earlier studies using SmD, associations were to be formed toward one single concept. That is, to form a verb to a noun (Prabhakaran et al., 2014), synonyms to a word (Beaty et al., 2014) or analogies between two words (Weinberger et al., 2016). Notice, however, that reaction time in the ACT significantly predicted SmD in a positive direction (longer reaction time equals greater SmD). Thus, it seems that SmD is related to uncommonness, where the longer someone thinks, the more unusual the response should be (support for internal validity). This is also in line with previous findings, which found that instances of more unusual responses increase over time (Benedek and Neubauer, 2013) and that category switching in divergent thinking tasks was indicated by a higher latency (Acar and Runco, 2017).

Another difference lies in the language of the experiment. All participants were of German nationality and spoke German as

their first language. Accordingly, the semantic space of the LSA was based on the German language, whereas earlier studies were conducted in English. However, it is unreasonable, although not impossible, to assume that SmD or similarity between concepts are differently perceived by different nationalities and differently reflected in the LSA.

To conclude, several constraints can be attested to the design of this study. SmD might not capture the cognitive process of association formation. Consequently, analysis methods trying to estimate the temporal structure might fail due to the inappropriateness of the data.

Future Directions

Future research should, therefore, pursue to refine the methodology for assessing the (temporal) structure of associations. Network science could bring benefits to the researcher seeking to investigate semantic structure and how this relates to divergent thinking and creativity. Additionally, the connection between SmD and creativity should be further explored to concisely pinpoint its relationship. Some authors successfully used LSA to also study sentence-like responses in divergent thinking tasks compared to single-word responses as in this study (e.g., Forthmann et al., 2018). Although this might be possible for LSA and SmD, techniques to infer the temporal structure, such as DFA and PSD, would only yield meaningful results with many more observations (favorably 256 and more) than usually available in common divergent thinking tasks. Another example could be to also study the phonological similarity between words or to apply different computational methods, as LSA is only one of several methods (see, e.g., HAL: Lund and Burgess, 1996) to infer the similarity of semantic concepts (Günther et al., 2015). SmD has the potential to complement established creativity measures (which are mainly subjective) as an objective instrument for assessing creative potential. Therefore, we encourage fellow researchers to venture new and potentially fruitful paths by taking inspiration from other fields.

CONCLUSION

As stated in prominent journals (e.g., current *Frontiers Research Topic description*, special issue *Journal of Creative Behavior*), the creativity research field could benefit from more interdisciplinary work and a broader range of methodological approaches. Existing creativity research often applies a relatively small number of empirical methodologies. In the current study we integrated methodology from computational linguistics and complex systems into creativity research to further enhance our understanding of cognitive creativity. Although the current study does not corroborate the idea that a flexible-stable (vs. random/persistent) temporal structure of associations is related to enhanced performance in divergent thinking, it hopefully challenges fellow researchers to refine the recent methodological developments for assessing the

(temporal) structure of associations. Moreover, we hope that the current cross-fertilization of methodological approaches inspires researchers to take advantage of other fields' ideas and methods. To derive at a theoretically sound cognitive theory of creativity, it is important to integrate research ideas and empirical methods from a variety of disciplines.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Code of Ethics for Research in the Social and Behavioral Sciences involving Human Participants by the Ethiek Commissie Sociale Wetenschappen of the Radboud University Nijmegen. The protocol was approved by the Ethiek Commissie Sociale Wetenschappen. All subjects gave

written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

PW conceived the idea, carried out the experiments, analyzed data, and wrote the manuscript. SR and MW supervised, refined the design, and edited the manuscript. MW also analyzed data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01771/full#supplementary-material>

REFERENCES

- Acar, S., and Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creat. Res. J.* 26, 229–238. doi: 10.1080/10400419.2014.901095
- Acar, S., and Runco, M. A. (2017). Latency predicts category switch in divergent thinking. *Psychol. Aesthet. Creat. Arts* 11, 43–51. doi: 10.1037/aca0000091
- Ambrose, D. (2017). Interdisciplinary invigoration of creativity studies. *J. Creat. Behav.* 51, 348–351. doi: 10.1002/jocb.205
- Bassingthwaite, J. B., Liebovitch, L. S., and West, B. J. (1994). *Fractal Physiology*. New York, NY: Springer, doi: 10.1007/978-1-4614-7572-9
- Beatty, R. E., and Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychol. Aesthet. Creat. Arts* 6, 309–319. doi: 10.1037/a0029171
- Beatty, R. E., Silvia, P. J., Nusbaum, E. C., Jaak, E., and Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Mem. Cogn.* 42, 1186–1197. doi: 10.3758/s13421-014-0428-8
- Benedek, M., Könen, T., and Neubauer, A. C. (2012). Associative abilities underlying creativity. *Psychol. Aesthet. Creat. Arts* 6, 273–281. doi: 10.1037/a0027059
- Benedek, M., and Neubauer, A. C. (2013). Revisiting mednick's model on creativity-related differences in associative hierarchies. Evidence for a common path to uncommon thought. *J. Creat. Behav.* 47, 273–289. doi: 10.1002/jocb.35
- Berryman, A. A. (1992). The origins and evolution of predator-prey theory. *Ecology* 73, 1530–1535. doi: 10.2307/1940005
- Carson, S. H., Peterson, J. B., and Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creat. Res. J.* 17, 37–50. doi: 10.1207/s15326934crj1701_4
- Chen, C., Kasof, J., Himsel, A., Dmitrieva, J., Dong, Q., and Xue, G. (2005). Effects of explicit instruction to “be creative” across domains and cultures. *J. Creat. Behav.* 39, 89–110. doi: 10.1002/j.2162-6057.2005.tb01252.x
- Christensen, P. R., Guilford, J. P., and Wilson, R. C. (1957). Relations of creative responses to working time and instructions. *J. Exp. Psychol.* 53, 82–88. doi: 10.1037/h0045461
- Corona, O. L., Padilla, P., Escolero, O., Frank, A., and Fossion, R. (2013). Lévy flights, 1/f noise and self organized criticality in a traveling agent model. *J. Mod. Phys.* 04, 337–343. doi: 10.4236/jmp.2013.43046
- Cropley, A. (2006). In praise of convergent thinking. *Creat. Res. J.* 18, 391–404. doi: 10.1207/s15326934crj1803_13
- De Dreu, C. K. W., Baas, M., and Nijstad, B. A. (2008). Hedonic tone and activation level in the mood-creativity link: toward a dual pathway to creativity model. *J. Pers. Soc. Psychol.* 94, 739–756. doi: 10.1037/0022-3514.94.5.739
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.* 41, 391–407. doi: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9
- Delignieres, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M., and Ninot, G. (2006). Fractal analyses for ‘short’ time series: a re-assessment of classical methods. *J. Mathematical Psychology* 50, 525–544. doi: 10.1016/j.jmp.2006.07.004
- Diedrich, J., Benedek, M., Jaak, E., and Neubauer, A. C. (2015). Are creative ideas novel and useful? *Psychol. Aesthet. Creat. Arts* 9, 35–40. doi: 10.1037/a0038688
- Forthmann, B., Gerwig, A., Holling, H., Çelik, P., Storme, M., and Lubart, T. (2016). The be-creative effect in divergent thinking: the interplay of instruction and object frequency. *Intelligence* 57, 25–32. doi: 10.1016/j.intell.2016.03.005
- Forthmann, B., Oyeade, O., Ojo, A., Günther, F., and Holling, H. (2018). Application of latent semantic analysis to divergent thinking is biased by elaboration. *J. Creat. Behav.* doi: 10.1002/jocb.240
- Friedenberg, J. (2009). *Dynamical Psychology: Complexity, Self-Organization and Mind*. Lichtfield Park, AZ: ISCE Publishing Dynamical Psychology.
- Gilden, D. L., and Hancock, H. (2007). Response variability in attention-deficit disorders. *Psychol. Sci.* 18, 796–802. doi: 10.1111/j.1467-9280.2007.01982.x
- Gilden, D. L., Thornton, T., and Mallon, M. (1995). 1/f noise in human cognition. *Science* 267, 1837–1839. doi: 10.1126/science.7892611
- Goldberger, A. L. (1997). Fractal variability versus pathologic periodicity: complexity loss and stereotypy in disease. *Perspect. Biol. Med.* 40, 543–561. doi: 10.1353/pbm.1997.0063
- Green, A. E. (2016). Creativity, within reason: semantic distance and dynamic state creativity in relational thinking and reasoning. *Curr. Dir. Psychol. Sci.* 25, 28–35. doi: 10.1177/0963721415618485
- Green, A. E., Cohen, M. S., Raab, H. A., Yedibalian, C. G., and Gray, J. R. (2015). Frontopolar activity and connectivity support dynamic conscious augmentation of creative state. *Hum. Brain Mapp.* 36, 923–934. doi: 10.1002/hbm.22676
- Green, A. E., Fugelsang, J. A., and Dunbar, K. N. (2006). Automatic activation of categorical and abstract analogical relations in analogical reasoning. *Mem. Cogn.* 34, 1414–1421. doi: 10.3758/BF03195906
- Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., and Dunbar, K. N. (2010). Connecting long distance: semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cereb. Cortex* 20, 70–76. doi: 10.1093/cercor/bhp081
- Green, A. E., Spiegel, K. A., Giangrande, E. J., Weinberger, A. B., Gallagher, N. M., and Turkeltaub, P. E. (2016). Thinking cap plus thinking zap: tdc of frontopolar cortex improves creative analogical reasoning and facilitates conscious augmentation of state creativity in verb generation. *Cereb. Cortex* 27:bhw080. doi: 10.1093/cercor/bhw080
- Guilford, J. P. (1950). Creativity. *Am. Psychol.* 5, 444–454. doi: 10.1037/h0063487
- Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York, NY: McGraw-Hill.
- Günther, F., Dudschig, C., and Kaup, B. (2015). LSAfun - An R package for computations based on latent semantic analysis. *Behav. Res. Methods* 47, 930–944. doi: 10.3758/s13428-014-0529-0
- Günther, F., Dudschig, C., and Kaup, B. (2016). Latent semantic analysis cosines as a cognitive similarity measure: evidence from priming studies. *Q. J. Exp. Psychol.* 69, 626–653. doi: 10.1080/17470218.2015.1038280

- Harrington, D. M. (1975). Effects of explicit instructions to “be creative” on the psychological meaning of divergent thinking test scores. *J. Pers.* 43, 434–454. doi: 10.1111/j.1467-6494.1975.tb00715.x
- Harris, Z. S. (1954). Distributional structure. *Word* 10, 146–162. doi: 10.1080/00437956.1954.11659520
- Hass, R. W. (2017a). Semantic search during divergent thinking. *Cognition* 166, 344–357. doi: 10.1016/j.cognition.2017.05.039
- Hass, R. W. (2017b). Tracking the dynamics of divergent thinking via semantic distance: analytic methods and theoretical implications. *Mem. Cogn.* 45, 233–244. doi: 10.3758/s13421-016-0659-y
- Holden, J. G. (2005). “Gauging the fractal dimension of response times from cognitive tasks,” in *Contemporary Nonlinear Methods for Behavioral Scientists: A Webbook Tutorial*, eds M. A. Riley and G. C. Van Orden (Arlington, VA: National Science Foundation), 267–318.
- Hong, E., O’Neil, H. F., and Peng, Y. (2016). Effects of explicit instructions, metacognition, and motivation on creative performance. *Creat. Res. J.* 28, 33–45. doi: 10.1080/10400419.2016.1125252
- Hutchison, K. A., Balota, D. A., Cortese, M. J., and Watson, J. M. (2008). Predicting semantic priming at the item level. *Q. J. Exp. Psychol.* 61, 1036–1066. doi: 10.1080/17470210701438111
- Jauk, E., Benedek, M., Dunst, B., and Neubauer, A. C. (2013). The relationship between intelligence and creativity: new support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence* 41, 212–221. doi: 10.1016/j.intell.2013.03.003
- Jones, M. N., Kintsch, W., and Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *J. Mem. Lang.* 55, 534–552. doi: 10.1016/j.jml.2006.07.003
- Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D. M., Gajda, A., et al. (2016). Is creativity without intelligence possible? A necessary condition analysis. *Intelligence* 57, 105–117. doi: 10.1016/j.intell.2016.04.006
- Karwowski, M., and Gralewski, J. (2013). Threshold hypothesis: fact or artifact? *Think. Skills Creat.* 8, 25–33. doi: 10.1016/j.tsc.2012.05.003
- Kaufman, J. C., and Beghetto, R. A. (2009). Beyond big and little: the four c model of creativity. *Rev. Gen. Psychol.* 13, 1–12. doi: 10.1037/a0013688
- Kaufman, J. C., and Beghetto, R. A. (2013). Do people recognize the four Cs? Examining layperson conceptions of creativity. *Psychol. Aesthet. Creat. Arts* 7, 229–236. doi: 10.1037/a0033295
- Kaufman, J. C., Plucker, J. A., and Baer, J. (2008). *Essentials of Creativity Assessment*, Vol. 53. Hoboken, NJ: John Wiley & Sons.
- Kenett, Y. N., Anaki, D., and Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Front. Hum. Neurosci.* 8:407. doi: 10.3389/fnhum.2014.00407
- Kenett, Y. N., and Austerweil, J. L. (2016). “Examining search processes in low and high creative individuals with random walks,” in *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (Austin, TX: Cognitive Science Society), 313–318.
- Kenett, Y. N., Beaty, R. E., Silvia, P. J., Anaki, D., and Faust, M. (2016). Structure and flexibility: investigating the relation between the structure of the mental lexicon, fluid intelligence, and creative achievement. *Psychol. Aesthet. Creat. Arts* 10, 377–388. doi: 10.1037/aca0000056
- Kenett, Y. N., Levi, E., Anaki, D., and Faust, M. (2017). The semantic distance task: quantifying semantic distance with semantic network path length. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 1470–1489. doi: 10.1037/xlm0000391
- Kenett, Y. N., Levy, O., Kenett, D. Y., Stanley, H. E., Faust, M., and Havlin, S. (2018). Flexibility of thought in high creative individuals represented by percolation analysis. *Proc. Natl. Acad. Sci. U.S.A.* 115, 867–872. doi: 10.1073/pnas.1717362115
- Lake, D. K., Moorman, J. R., and Hanqing, C. (2008). *Sample Entropy Estimation Using Sampen*. Available at: <https://www.physionet.org/physiotools/sampen/matlab/1.1-1/sampenc.m>
- Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato’s problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240. doi: 10.1037/0033-295X.104.2.211
- Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Process.* 25, 259–284. doi: 10.1080/01638539809545028
- Landmann, N., Kuhn, M., Piosczyk, H., Feige, B., Riemann, D., and Nissen, C. (2014). Entwicklung von 130 deutschsprachigen compound remote associate (CRA)-worträtseln zur untersuchung kreativer prozesse im deutschen sprachraum. *Psychol. Rundsch.* 65, 200–211. doi: 10.1026/0033-3042/a000223
- Levin, I. (1978). Creativity and two modes of associative fluency: chains and stars. *J. Pers.* 46, 426–437. doi: 10.1111/j.1467-6494.1978.tb01010.x
- Lund, K., and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* 28, 203–208. doi: 10.3758/BF03204766
- Manning, C. D., and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*, 1st Edn. Cambridge, MA: MIT Press.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* 69, 220–232. doi: 10.1037/h0048850
- Ooms, J. (2017). *Hunspell: High-Performance Stemmer, Tokenizer, and Spell Checker for R. R Package Version 2.3*. Available at: <https://CRAN.R-project.org/package=hunspell>
- Osipow, S. H., and Grooms, R. R. (1966). The temporal stability of chains of word associations. *J. Gen. Psychol.* 75, 157–175. doi: 10.1080/00221309.1966.9710361
- Peirce, J. W. (2007). PsychoPy—psychophysics software in python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017
- Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Phys. Rev. E.* 49, 1685–1689. doi: 10.1103/PhysRevE.49.1685
- Peng, C.-K., Havlin, S., Hausdorff, J.-J. M., Mietus, J. E., Stanley, H. E., and Goldberger, A. L. (1995). Fractal mechanisms and heart rate dynamics long-range correlations and their breakdown with disease. *J. Electrocardiol.* 28, 59–65. doi: 10.1016/S0022-0736(95)80017-4
- Piccardo, E. (2017). Plurilingualism as a catalyst for creativity in superdiverse societies: a systemic analysis. *Front. Psychol.* 8:2169. doi: 10.3389/fpsyg.2017.02169
- Plucker, J. A., Qian, M., and Wang, S. (2011). Is originality in the eye of the beholder? comparison of scoring techniques in the assessment of divergent thinking. *J. Creat. Behav.* 45, 1–22. doi: 10.1002/j.2162-6057.2011.tb01081.x
- Poutanen, P. (2013). Creativity as seen through the complex systems perspective. *Interdiscip. Stud. J.* 2, 207–221.
- Prabhakaran, R., Green, A. E., and Gray, J. R. (2014). Thin slices of creativity: using single-word utterances to assess creative cognition. *Behav. Res. Methods* 46, 641–659. doi: 10.3758/s13428-013-0401-7
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.
- Rangarajan, G., and Ding, M. (2001). An integrated approach to the assessment of long range correlation in time series data. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* 61, 4991–5001. doi: 10.1103/PhysRevE.61.4991
- Revelle, W. (2016). *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Available at: <https://CRAN.R-project.org/package=psych>
- Richman, J. S., and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049. doi: 10.1152/ajpheart.2000.278.6.H2039
- Ritter, S. M., and Mostert, N. (2017). Enhancement of creative thinking skills using a cognitive-based creativity training. *J. Cogn. Enhanc.* 1, 243–253. doi: 10.1007/s41465-016-0002-3
- Rossmann, E., and Fink, A. (2010). Do creative people use shorter associative pathways? *Personal. Individ. Dif.* 49, 891–895. doi: 10.1016/j.paid.2010.07.025
- Runco, M. A., and Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creat. Res. J.* 24, 66–75. doi: 10.1080/10400419.2012.652929
- Runco, M. A., and Jaeger, G. J. (2012). The standard definition of creativity. *Creat. Res. J.* 24, 92–96. doi: 10.1080/10400419.2012.650092
- Runco, M. A., and Okuda, S. M. (1991). The instructional enhancement of the flexibility and originality scores of divergent thinking tests. *Appl. Cogn. Psychol.* 5, 435–441. doi: 10.1002/acp.2350050505
- Sahlgren, M. (2008). The distributional hypothesis. *Ital. J. Linguist.* 20, 33–54.
- Schmidt, R. C. (2001). *Detrended Fluctuation Analysis*. GitHub Repository. Available at: <https://github.com/FredHasselmann/toolboxML/blob/master/Ddfa.m>
- Scott, G., Leritz, L. E., and Mumford, M. D. (2004). The effectiveness of creativity training: a quantitative review. *Creat. Res. J.* 16, 361–388. doi: 10.1080/10400410409534549
- Stein, M. I. (1953). Creativity and culture. *J. Psychol.* 36, 311–322. doi: 10.1080/00223980.1953.9712897

- The Math Works (2017). *MATLAB and Signal Processing Toolbox*. Natick, MA: The MathWorks.
- Van Orden, G. C., Holden, J. G., and Turvey, M. T. (2003). Self-organization of cognitive performance. *J. Exp. Psychol. Gen.* 132, 331–350. doi: 10.1037/0096-3445.132.3.331
- Van Orden, G. C., Kloos, H., and Wallot, S. (2011). “Living in the pink: intentionality, wellbeing, and complexity,” in *Philosophy of Complex Systems*, Vol. 10, ed. C. Hooker (New York, NY: Elsevier), 629–672. doi: 10.1016/B978-0-444-52076-0.50022-5
- Weinberger, A. B., Iyer, H., and Green, A. E. (2016). Conscious augmentation of creative state enhances “real” creativity in open-ended analogical reasoning. *PLoS One* 11:e0150773. doi: 10.1371/journal.pone.0150773
- Wijnants, M. L. (2014). A review of theoretical perspectives in cognitive science on the presence of scaling in coordinated physiological and cognitive processes. *J. Nonlinear Dyn.* 2014, 1–17. doi: 10.1155/2014/962043
- Wijnants, M. L., Bosman, A. M. T., Hasselman, F., Cox, R. F. A., and Van Orden, G. C. (2009). $1/f$ scaling in movement time changes with practice in precision aiming. *Nonlinear Dyn. Psychol. Life Sci.* 13, 75–94.
- Wijnants, M. L., Cox, R. F. A., Hasselman, F., Bosman, A. M. T., and Van Orden, G. C. (2012). A trade-off study revealing nested timescales of constraint. *Front. Physiol.* 3:116. doi: 10.3389/fphys.2012.00116
- Zabelina, D. L., O’Leary, D., Pornpattananangkul, N., Nusslock, R., and Beeman, M. (2015a). Creativity and sensory gating indexed by the P50: selective versus leaky sensory gating in divergent thinkers and creative achievers. *Neuropsychologia* 69, 77–84. doi: 10.1016/j.neuropsychologia.2015.01.034
- Zabelina, D. L., Saporta, A., and Beeman, M. (2015b). Flexible or leaky attention in creative people? Distinct patterns of attention for different types of creative thinking. *Mem. Cogn.* 44, 488–498. doi: 10.3758/s13421-015-0569-4

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 Wang, Wijnants and Ritter. 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.



Engaging in Creativity Broadens Attentional Scope

Marta K. Wronska^{1,2*}, Alina Kolańczyk¹ and Bernard A. Nijstad²

¹ Faculty in Sopot, SWPS University of Social Sciences and Humanities, Sopot, Poland, ² Department of Human Resource Management and Organizational Behavior, University of Groningen, Groningen, Netherlands

OPEN ACCESS

Edited by:

Amory H. Danek,
Universität Heidelberg, Germany

Reviewed by:

Jasmin M. Kizilirmak,
University of Hildesheim, Germany
Valerio Santangelo,
University of Perugia, Italy
Gillian Hill,
University of Buckingham,
United Kingdom

*Correspondence:

Marta K. Wronska
m.k.wronska@rug.nl

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 28 March 2018

Accepted: 03 September 2018

Published: 21 September 2018

Citation:

Wronska MK, Kolańczyk A and
Nijstad BA (2018) Engaging
in Creativity Broadens Attentional
Scope. *Front. Psychol.* 9:1772.
doi: 10.3389/fpsyg.2018.01772

Previous studies have shown that creativity is enhanced by a broad attentional scope, defined as an ability to utilize peripheral stimuli and process information globally. We propose that the reverse relationship also holds, and that breadth of attention also is a consequence of engaging in a creative activity. In Study 1, participants showed increased breadth of attention in a visual scanning task after performing a divergent thinking task as opposed to an analytic thinking task. In Study 2, participants recognized peripheral stimuli displayed during the task better after performing a divergent thinking task as compared to an analytic task, whereas recognition performance of participants performing a task that involves a mix of divergent and analytic thinking (the Remote Associates Test) fell in between. Additionally, in Study 2 (but not in Study 1), breadth of attention was positively correlated with performance in a divergent thinking task, but not with performance in an analytic thinking task. Our findings suggest that the adjustment of the cognitive system to task demands manifests at a very basic, perceptual level, through changes in the breadth of visual attention. This paper contributes a new, motivational perspective on attentional breadth and discusses it as a result of adjusting cognitive processing to the task requirements, which contributes to effective self-regulation.

Keywords: creativity, idea generation, divergent thinking, breadth of attention, self-regulation, analytic thinking, Remote Associates Test, convergent thinking

INTRODUCTION

What is the temperature in the place you are currently in and what background sounds can you hear? Unless the environmental conditions are extreme, you probably did not register these peripheral, seemingly unimportant stimuli. Indeed, doing so would only be distracting and may interfere with other activities. When generating creative ideas (ideas that are both novel and useful; Amabile, 1983), however, having a broad attentional scope and noticing peripheral stimuli can be beneficial. For example, Mendelsohn and Griswold (1964) found that people who score high on creativity tests, as compared to less creative problem solvers, are better able to take advantage of peripheral cues (prompts) to solve the task at hand, and similar results were obtained in later experiments (Mendelsohn and Griswold, 1966; Mendelsohn and Lindholm, 1972; Ansburg and Hill, 2003). More recent studies also found strong support for the beneficial effect of broad attention on creative idea generation: Creativity is enhanced by meditation techniques that broaden attention (Colzato et al., 2012, 2017; see also Lebeda et al., 2016), as well as by experimental manipulations that increase attentional breadth (Friedman et al., 2003; Förster et al., 2004; Jia et al., 2009; Liu, 2016; Moraru et al., 2016). It has even been found that alcohol intake can facilitate creative problem

solving, which is expected to be driven by reduced attentional control and higher sensitivity to peripheral information (Jarosz et al., 2012).

Although it is clear that breadth of attention influences creative performance, here we explore the intriguing possibility of the reverse causal relation: that engaging in creative activity can influence attentional breadth. Just as certain types of meditation or experimental manipulations enhance breadth of attention, engaging in a creative task may broaden the attentional field. This possibility is intriguing because it would suggest that the cognitive system is able to adapt to task demands at a fundamental (perceptual) level. Indeed, Vartanian (2009) suggests that successful problem solving requires the cognitive system to flexibly adjust to task requirements. Because creative tasks are ill-defined and demand exploration of problem space (e.g., Arreola and Reiter-Palmon, 2016), a broader attentional scope is beneficial and may be triggered by the particular activities for which it is needed. For example, it is possible that engaging in brainstorming activates a completely different mindset (which manifests in the attentional breadth) than engaging in planning an agenda, because planning does not require a broad search for solutions whereas brainstorming does.

The current paper reports two experiments in which we manipulated engagement in certain activities (creative idea generation or other) and measured attentional breadth in different ways. Specifically, we measured attentional breadth as a consequence of engagement in a divergent thinking task, an analytic thinking task, or (only in Study 2) in the Remote Associates Test (RAT; Mednick, 1962), a test that involves a mix of divergent and analytic thinking. Together, these studies suggest that the cognitive system adaptively responds to task demands at a very basic level of information processing (breadth of attention) and that breadth of attention is also a consequence of engaging in a creative activity. Based on these and prior findings, we propose that the attentional breadth–creativity relation is, in fact, bi-directional.

Broad Attention Stimulates Creativity

According to Mednick's (1962) associative theory, creativity requires finding elements that are remotely associated and combining them in a meaningful way. This theory explains why a broad scope of attention should increase creativity: It gives access to a larger pool of elements, and therefore, facilitates original combinations of these elements (Mendelsohn, 1976). Mednick proposed that people differ in the strength of their associations to certain concepts (e.g., “table”), with some people having a steep association hierarchy and others a flatter one. If one association dominates (e.g., “chair”), then the remaining, potentially creative associations are less likely to be activated, and the association hierarchy is steep; however, if various associations are similarly strong, the association hierarchy is relatively flat, which may lead to more creative outcomes. Because a broad scope of attention implies that diverse elements in the perceptual field are similarly important, broad attention should facilitate a flat association hierarchy. This idea has been proposed by Martindale (1989), who suggested that the mind can be represented as a set of interconnected nodes, similar to

neural networks, which may be activated in different degrees (see also Spreading-Activation Theory of Semantic Processing; Collins and Loftus, 1975). When attention is narrow, strong activation of a single node prevents activation from spreading to other nodes in the network – in this case, a single concept (like “chair” in response to “table”) is activated strongly and adjacent nodes (like “tablecloth”) are inhibited. However, when more nodes are activated simultaneously and attention is broad, then the activation of each node is weaker, and there is no inhibiting effect on other nodes. Such situations lead to the generation of more remote, and potentially creative, associations. A similar idea has been proposed by the extensive–intensive attention theory (Kolańczyk, 1989, 1991, 2011, 2012): extensive attention relates to more sensitivity toward peripheral stimuli (rather than strong focus on central stimuli) and consequently, weak activation of a large pool of nodes in the semantic network.

Consistent with this idea, Mendelsohn (1976) found that those who are able to connect remote ideas are also those who can take advantage of seemingly irrelevant, peripheral stimuli to solve the task at hand. Furthermore, Ansburg and Hill (2003) confirmed this idea by showing that scores on the RAT (Mednick, 1962), a test which measures the ability to make remote associations, positively predict the number of word puzzles (anagrams) solved with peripheral cues (answers to the word puzzles played on the tape recorder in the background). Other evidence is also consistent with this reasoning. For example, Kasof (1997) found a positive relation between creativity of poems and sensitivity to peripheral stimuli in the environment. Experimental studies confirm that it is indeed a broad conceptual scope that increases creative performance (Isen and Daubman, 1984; Isen et al., 1987; Jarosz et al., 2012; Deuja et al., 2014; Chiu, 2015; Liu, 2016). Finally, studies on meditation suggest that attending to the surroundings in a broad and defocused manner boosts creativity: Open monitoring, compared with focused attention meditation, has been found to increase performance in creative idea generation (Colzato et al., 2012, 2017; Baas et al., 2014). Together, these findings provide converging evidence that broad attention facilitates creative performance by expanding the scope of concepts that may be combined into a potentially creative outcome.

Does Engaging in Creativity Lead to Broader Attention?

Although it is well established that broad attention increases creativity, the idea that engaging in creative activity could alter the breadth of the attentional field has not yet been investigated. If generating creative ideas requires broadening the conceptual scope to transcend from obvious solutions to more original ones, it is also possible that attempting to produce creative output in itself will broaden the attentional field. This can be true especially when we compare it with engaging in an activity that does not require such expansion of horizons, or even asks for the opposite – focusing only on the task-relevant information to arrive at a single correct solution (cf. Ansburg and Hill, 2003; Liu, 2016).

Indirect support for this idea comes from the studies that contrasted divergent thinking tasks with the RAT (Akbari Chermahini and Hommel, 2012; Fischer and Hommel, 2012). In divergent thinking tasks, participants were asked to generate multiple creative uses of an everyday object (e.g., a brick), whereas in the RAT participants had to provide a single word that is a common associate for three words that were provided; here, only one solution was correct. Hommel (2012) argued that engaging in these tasks induces a certain control state, which either favors flexible switching between options with little “top-down” guidance (divergent thinking) or releases a strong top-down bias, which guides a person toward one specific option (solving the RAT). The first case is associated with achieving creativity through flexible and relatively effortless processing (i.e., low cognitive control and low self-control; Kolańczyk, 2012), whereas the second refers to creativity achieved through persistent and effortful processing (i.e., high cognitive control; Nijstad et al., 2010). Results have shown that engaging in divergent thinking, compared with solving the RAT, led to higher multitasking performance (Fischer and Hommel, 2012), and to a more positive mood (Akbari Chermahini and Hommel, 2012), which is associated with broad attention and global processing (Isen and Daubman, 1984; Fredrickson and Branigan, 2005; Bramesfeld and Gasper, 2008; Kuhbandner et al., 2011; Schmid et al., 2011).

Overview of the Present Studies

Overall, these results suggest that the weak top-down control state induced by divergent thinking should be connected with defocused and broader attention (see also Martindale, 1989; Kolańczyk, 2012; Zhou et al., 2017). However, to the best of our knowledge, there is no direct evidence showing this effect. Providing a direct test of this idea is the aim of the present contribution. If indeed engaging in a divergent thinking task broadens the attentional field (as compared with engaging in analytic thinking task) this would indicate that, at a very basic perceptual level, the cognitive system can adapt to task demands.

To test the idea that engaging in creative activity leads to a broader attentional field, we performed two studies, in which we compared a divergent thinking task with an analytic thinking task (Study 1) and a divergent thinking task with both an analytic thinking task and the RAT (Study 2). We expected that performing a divergent thinking task would lead to a broader attentional field than performing an analytic thinking task, because top-down cognitive control is lower for a task that requires flexible and explorative processing (i.e., divergent thinking) than for a task that requires careful evaluation of task-related information to arrive at a single correct solution (i.e., analytic thinking). In turn, these differences in mindset and cognitive control state will translate to differences in breadth of attention.

Both studies used a between-subjects design and employed different measures of breadth of attention. Study 1 measured attentional breadth with a task specifically designed to measure extensive–intensive attention states (Roczniewska et al., 2011), with a state of extensive attention defined as broader and more sensitive to peripheral stimuli than a state of intensive attention. In the second study, we drew from the peripheral cues

paradigm (Mendelsohn and Griswold, 1964) to measure breadth of attention through recognition of peripheral stimuli. We also assessed performance on each task and examined whether performance in each of the tasks correlates with our measure of breadth of attention. As discussed above, previous research suggests that breadth of attention should correlate positively with creative performance but not with analytic performance.

STUDY 1

Method

Participants and Design

Ninety undergraduate students participated in an experiment on the “properties of cognitive processes” in exchange for credit points. However, 14 participants were excluded from analysis due to: disrupted procedure during attention measurement (e.g., talking to the experimenter, the door being opened, noise), using a touchpad instead of a mouse, failing to understand the attentional breadth measure instruction (e.g., selecting very few stimuli, see the description of the Ellipses Test in Measures), and a computer malfunction. Data from 76 participants was analyzed (59 females and 17 males), whose age ranged from 18 to 53 years ($M = 21.59$, $SD = 4.16$). Average age did not differ between conditions, $t(74) = 0.26$, $p = 0.798$.

Participants were randomly assigned to two conditions of a between-subjects design. In the *divergent thinking* task condition ($n = 40$; 30 female, $M_{\text{age}} = 21.48$), participants performed the Unusual Uses Task with instructions developed by Silvia et al. (2008). Participants were asked to write down all original and creative uses of a brick they could think of. Participants in the *analytic thinking* task condition ($n = 36$; 29 female, $M_{\text{age}} = 21.72$) were asked to solve a task from the analytic reasoning section of the Law School Admission Test (Princeton Review, 2015; also see Kray et al., 2006). This test measures the ability to derive conclusions from a set of assumptions and asks participants to apply logic to multifaceted problems, understand how rules affect outcomes and decisions, and identify connections between concepts. The task that we employed required the participants to follow five rules (e.g., “the student must clean the kitchen first before shopping for groceries”) to determine the correct order of household chores (e.g., “grocery shopping”) performed by a student.

Measures

Breadth of attention

To measure breadth of attention, we used the Ellipses Test (Roczniewska et al., 2011), which consisted of 363 letters (*a*, *d*, *e*, *k*, *s*, and *w*) arranged in the shape of ellipses on a computer screen. Ellipses made of letters varied in size, with smaller ellipses located inside bigger ellipses (see **Figure 1**). Letters were displayed in a black font on a white background. Participants had to select letters *d* with mouse clicks. After a letter had been clicked, its color turned to green to mark its selection. Some *ds* were spread out ($n = 17$) and others ($n = 43$) appeared in small clusters, which made them easier to spot with a broader attentional field. We used the distance between selections (percentages of the screen size)

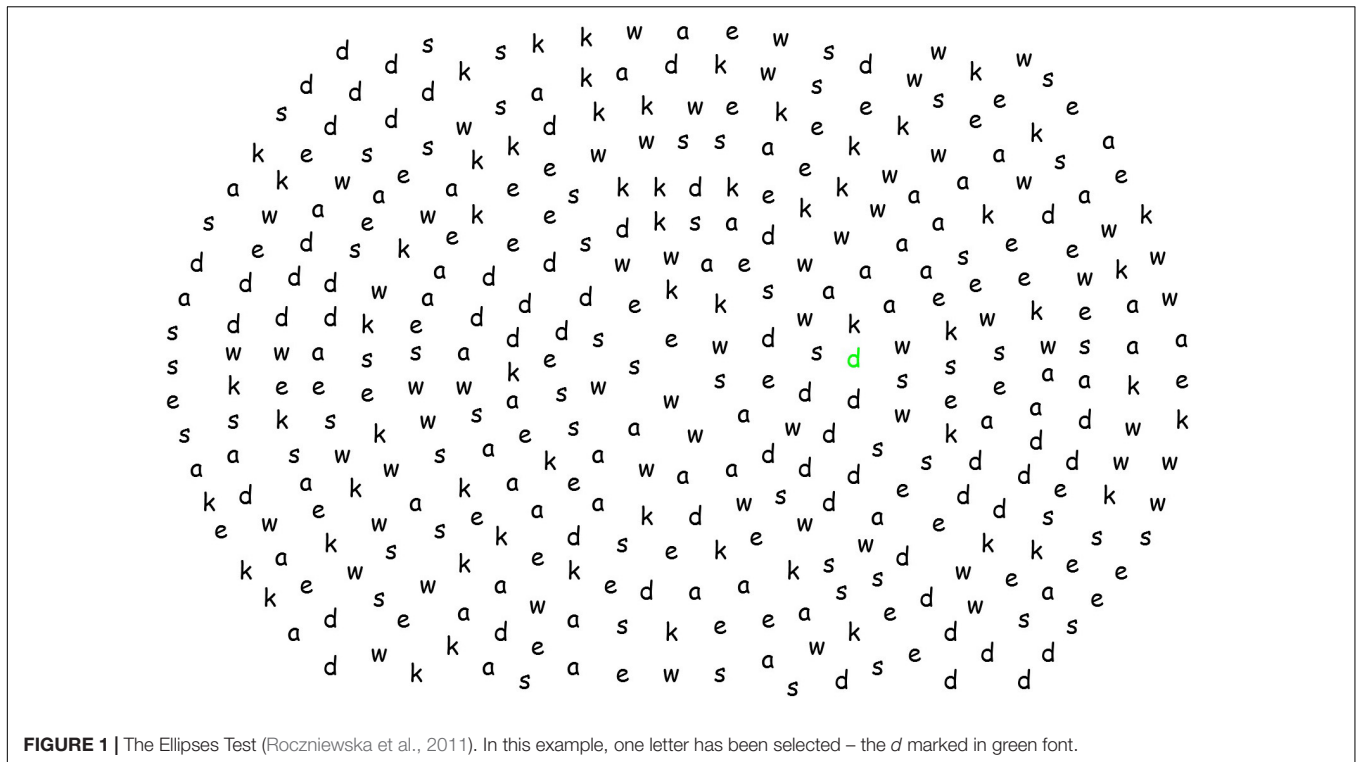


FIGURE 1 | The Ellipses Test (Roczniewska et al., 2011). In this example, one letter has been selected – the *d* marked in green font.

as indicator of attentional breadth and computed two indicators: *total distance* covered (“travelled”) by the solver while searching for *ds* and the *standard deviation* (SD) of distance between clicks.

Total distance was computed as the sum of distances between all clicks. High total distance indicates that the solver searched for *ds* globally, within a broad perceptual field; low total distance indicates that the solver searched for *ds* locally, within a narrow perceptual field. SD of distance was computed to examine the amount of variation in distances. Because most *ds* appeared in clusters and broad (but not narrow) attention should facilitate spotting such clusters, this should result in small distances within each cluster and big distances between the clusters, thus creating a high standard deviation. Participants may differ in how many letters they selected in total, so we controlled for the total number of clicked letters, as this could bias the attentional breadth indicators.

Control measures

It is possible that engaging in different tasks influenced participants’ mood state. Because moods affect creative performance (Baas et al., 2008), we employed two versions of a 4-item questionnaire to measure pretest and posttest mood (Wojciszke and Baryła, 2004). Participants rated statements (e.g., “I’m in a bad mood”) on a 5-point scale (1 = *disagree*, 5 = *agree*). Scale reliabilities were good (Cronbach’s $\alpha = 0.91$ for version A and Cronbach’s $\alpha = 0.89$ for version B). We also controlled the subjective difficulty of the task (Bujacz et al., 2014), because task difficulty may affect attentional processes (e.g., Santangelo et al., 2011; see also Santangelo and Spence, 2008). Participants indicated to what extent they found the previous task: “easy,”

“undemanding,” “unproblematic” (all reverse scored), “difficult,” “complicated,” and “challenging” (Cronbach’s $\alpha = 0.85$). We employed a 7-point scale ranging from 0 (*not at all*) to 6 (*very much*). Moreover, participants were asked to rate their task enjoyment (Friedman and Förster, 2002; “How much did you enjoy the task?”), using the same 7-point scale.

Procedure

Upon arrival, participants gave written informed consent and were randomly assigned to a divergent thinking or analytic thinking task condition. The experiment was run in Inquisit Lab4. First, participants answered four pretest mood items (Wojciszke and Baryła, 2004). Subsequently, they engaged in a divergent thinking (Silvia et al., 2008) or analytic thinking task (Princeton Review, 2015; also see Kray et al., 2006) for 1.5 min. Participants could take notes on a sheet of paper, and after 1.5 min, an audio sound signaled that they had to look at the screen again. They were asked to stop the task and were informed that they would be able to finish it later. Next, the Ellipses Test was administered to measure attentional breadth (Roczniewska et al., 2011). Participants were instructed that a number of letters would appear on the screen. Their task was to select as many letters *d* as possible with mouse clicks.

After 2 min, the test ended and the participants were instructed to finish the divergent thinking or analytic thinking task. In the divergent thinking task condition, participants continued writing down possible uses of a brick for another 1.5 min. In total (before and after the Ellipses Test), they thus performed the divergent thinking task for 3 min (see also Silvia et al., 2008). When the time was up, they had to choose their

two most creative ideas and underline them. Participants in the analytic thinking task condition had 5 min to finish their task. The longer time was chosen to ensure that it was sufficient and proportional to task difficulty. However, participants were allowed to finish earlier, on condition that they had completed the task (finishing early was not allowed in the divergent thinking task condition). Participants were not informed about the time limit to avoid the confounding effect of time pressure (e.g., Hsu and Fan, 2010).

In the final part, participants rated their posttest mood with items differing from those used at the beginning (Wojciszke and Baryła, 2004). Subsequently, they evaluated the subjective difficulty and their enjoyment of the task; they indicated their gender, age, and were thanked for participation.

Coding Performance

Divergent thinking task

For the divergent thinking task, we closely followed the subjective scoring procedures developed by Silvia et al. (2008). Responses to the divergent thinking task were typed into a spreadsheet and sorted alphabetically. We engaged three coders (including the first author), all of whom were the alumni or students of an advanced university course on the psychology of creativity (including creativity diagnosis). They were trained by the first author and asked to read each response. Each coder independently scored the responses on a scale from 1 (*not at all creative*) to 5 (*highly creative*). Scoring instructions were translated from Silvia et al. (2008) by the first author and then back-translated by a professional English teacher (Polish native speaker). We obtained two indicators of creative performance: average creativity of all responses of each participant (average creativity) and an average from the two responses that the participant marked as the most creative (top 2 creativity). The interrater reliability was satisfactory: intraclass correlation coefficient (ICC; two-way random model, absolute agreement) was 0.811 ($p < 0.001$) for the average creativity and 0.680 ($p < 0.001$) for top 2 creativity, which indicates good and moderate reliability, respectively (Koo and Li, 2016).

Analytic thinking task

The aim of the analytic thinking task was to order household chores according to rules given (Princeton Review, 2015; also see Kray et al., 2006). Two possible orders could be correctly derived from the rules. In the 0–1 indicator, participants scored one point when the entire sequence of chores was correct; otherwise, the score was 0 points. In the 0–5 indicator, one point was given for each condition that was met (e.g., if all conditions were met, the participant scored five points).

Results

Control Variables

Control variables (task enjoyment and subjective difficulty, pretest and posttest mood¹) did not differ between experimental

conditions (all t s < 1.14 ; p s > 0.257). Mean accuracy (the number of clicked d s divided by the number of all clicked letters) in the Ellipses Test did not differ between conditions either, $t(74) = 0.26$, $p = 0.794$. Similarly, there were no differences between conditions in the number of all clicked letters ($t[74] = 0.28$, $p = 0.783$), number of clicked d s ($t[74] = 0.35$, $p = 0.732$; $M_{\text{divergent}} = 59.58$, $M_{\text{analytic}} = 59.28$), and in the number of other clicked letters ($t[74] = 0.20$, $p = 0.845$, $M_{\text{divergent}} = 0.88$, $M_{\text{analytic}} = 0.92$). This is in line with the assumptions of the method, which diagnoses attentional breadth not through the effectiveness of finding the d s but through the strategy of searching the perceptual field.

Effect of the Task on Breadth of Attention

We performed a multivariate analysis of covariance (MANCOVA) with task type (divergent vs. analytic) as independent variable, total distance and SD of distance as dependent variables, and total number of clicked letters as a covariate. We found a significant multivariate effect: participants who solved the divergent thinking task had broader attention ($M_{\text{total}} = 962.43$, $M_{\text{SD}} = 17.56$) than participants who solved the analytic thinking task ($M_{\text{total}} = 909.81$, $M_{\text{SD}} = 16.07$), $F(2, 72) = 3.19$, $p = 0.047$ (see **Figure 2**). Total number of clicked letters was a significant covariate, $F(2, 72) = 35.57$, $p < 0.001$. In a follow-up univariate analyses, the effect of task type (divergent vs. analytic) on total distance did not reach significance level when corrected for multiple comparisons, $F(1, 73) = 3.99$, $p = 0.050$ ($p = 0.100$ with Bonferroni correction), but the univariate effect of task type (divergent vs. analytic) on SD of distance was significant, $F(1, 73) = 6.24$, $p = 0.015$ ($p = 0.030$ with Bonferroni correction). Confidence intervals for both effects did not include zero, 95% CI (0.09, 102.93) for total distance and 95% CI (0.31, 2.76) for SD of distance, which suggests a significant difference for both indicators. Total number of clicked letters was a significant covariate for SD of distance, $F(1, 73) = 7.13$, $p = 0.009$ ($p = 0.018$ with Bonferroni correction), but not for total distance, $F(1, 73) = 1.78$, $p = 0.186$ ($p = 0.372$ with Bonferroni correction). The effect size was small to moderate (Cohen's $d = 0.47$ for total distance; Cohen's $d = 0.54$ for SD of distance; Cohen, 1977).

In order to verify whether mood, subjective task difficulty and enjoyment can account for the influence of the task type (divergent vs. analytic) on breadth of attention, we performed another MANCOVA. In this analysis, we additionally entered the following covariates: pretest mood, subjective task difficulty, and task enjoyment. We found that the additional covariates had no multivariate effect on breadth of attention (all p s > 0.13) and that the multivariate effect of the task type (divergent vs. analytic) on breadth of attention remained significant, $F(2, 29) = 3.84$, $p = 0.026$. Both effects in a univariate follow-up analyses remained at the same significance level as in the analysis without additional covariates, $F(1, 70) = 4.02$, $p = 0.049$ ($p = 0.098$ with Bonferroni

variable) and mood as a within-subjects factor (dependent variable). We found a significant interaction between task type and mood $F(1, 74) = 11.77$, $p = 0.001$. A follow-up simple effects analysis revealed that solving a divergent thinking task induced a more positive mood ($M_{\text{pretest}} = 3.79$, $M_{\text{posttest}} = 3.97$, $p = 0.044$, Cohen's $d = 0.39$), whereas solving an analytic thinking task elicited a more negative one ($M_{\text{pretest}} = 4.04$, $M_{\text{posttest}} = 3.79$, $p = 0.007$, Cohen's $d = 0.66$). This is in line with findings of Akbari Chermahini and Hommel (2012).

¹Additionally, we tested whether the divergent thinking task triggered a more positive mood and the analytic task triggered a more negative mood, as reported by Akbari Chermahini and Hommel (2012). We performed a two-way repeated measures ANOVA with task type as a between-subjects factor (independent

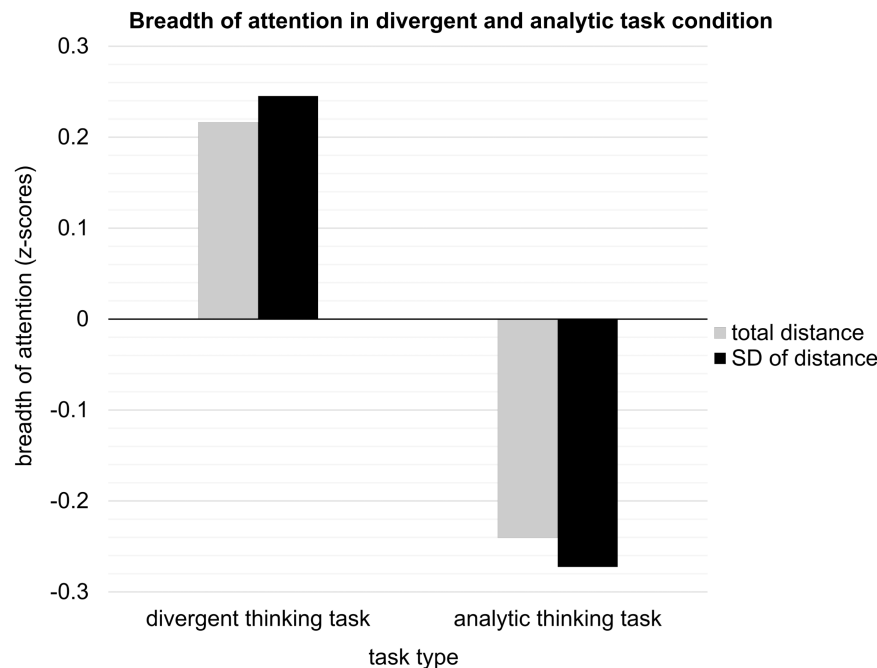


FIGURE 2 | Standardized breadth of attention (total distance and SD of distance) in divergent thinking task condition and analytic thinking task condition in Study 1. Total distance (in z-scores) ranged from -2.41 to 2.26 , and SD of distance ranged from -2.25 to 2.34 .

correction) for total distance and $F(1, 70) = 7.13$, $p = 0.009$ ($p = 0.018$ with Bonferroni correction) for SD of distance. Thus, pretest mood, subjective task difficulty, and enjoyment cannot explain the influence of task type (divergent vs. analytic) on attentional breadth.

Performance

We performed a correlation analysis separately for the divergent and analytical thinking task condition to examine whether breadth of attention correlated with performance. We found that performance in the divergent thinking task, as well as in the analytic thinking task, was unrelated to total distance and SD of distance ($-0.16 < r < 0.05$; all $ps > 0.351$).

TABLE 1 | Education and main activity of participants in Study 2.

	Divergent thinking task (<i>n</i> = 47) (%)	Analytic thinking task (<i>n</i> = 45) (%)	RAT (<i>n</i> = 46) (%)	Total sample (<i>N</i> = 138) (%)
Education				
Primary school	0	0	2.2	0.7
High school	53.2	37.8	37.0	42.8
University level	42.5	62.2	60.8	55.1
Missing	4.3	–	–	1.4
Main activity				
Education	61.7	26.7	65.2	65.9
Paid work	27.7	71.1	23.9	26.1
Other	6.4	2.2	10.9	6.5
Missing	4.3	–	–	1.4

Discussion of Study 1

Study 1 provided initial evidence that engaging in a divergent thinking task, compared with engaging in an analytic thinking task, broadens the scope of attention. We found a significant multivariate effect on attention indicators (total distance and SD of distance), both when we did and did not control for pretest mood, subjective task difficulty, and enjoyment. This suggests that these control variables cannot explain the effect of task type (divergent vs. analytic) on attentional breadth. We found a significant univariate effect on SD of distance, but the univariate effect on total distance did not reach significance. This suggests that broad attention triggered by the divergent thinking task was not so strongly visible in global search for the target letters within a broad perceptual field; instead, it was more reliably reflected in higher variation of distances obtained when the solver noticed and clicked on the *ds* that appeared in clusters. Furthermore, and somewhat surprisingly, in this study, attentional breadth was unrelated to creative performance. A possible explanation is that breadth of attention was measured in the middle of task performance. Switching attention between idea generation and the Ellipses Test potentially disrupted the flow of ideas while participants were generating creative solutions, which may have weakened the correlation between attentional breadth and creative performance.

STUDY 2

To replicate the findings of Study 1 and generalize the results to other divergent and analytic thinking tasks, we performed

Study 2. In this study, we wanted to avoid interrupting participants by the Ellipses Test, and therefore measured attentional breadth via the recognition of peripheral stimuli which were displayed during task performance. This method builds on the paradigm of incidental (peripheral) stimuli in creative problem solving (Mendelsohn and Griswold, 1964, 1966; Mendelsohn and Lindholm, 1972; Mendelsohn, 1976). In the studies of Mendelsohn and colleagues, participants were exposed to words played on a tape recorder while memorizing a list of other words. Next, they were asked to solve multiple anagrams. Some of the answers to the anagrams were earlier played on the tape recorder (answers to “peripheral anagrams”) and some were present on the list (answers to “central anagrams”). Those participants who achieved high scores on the creativity test also solved more peripheral anagrams (cf. Ansburg and Hill, 2003). Our study, however, used recognition of visual peripheral stimuli as a dependent variable, with the assumption that incidental recognition of peripheral cues would be better when attention is broad (vs. narrow) during task performance.

We also added a condition in which participants performed the RAT (Mednick, 1962). Interestingly, the RAT requires both divergent thinking (coming up with multiple candidates for the solution) and analytic thinking (evaluating the correctness of possible answers; Mendelsohn, 1976). Although previous research argued that solving the RAT requires more cognitive control than solving a divergent thinking task (Fischer and Hommel, 2012; Hommel, 2012), it has been found that the RAT can be solved both through an insight strategy (spontaneous activation of diverse associations) and through an analytic strategy (effortful and sequential search for close associations; see e.g., Bowden et al., 2005; Harkins, 2006). Furthermore, Topolinski and Strack (2008) found that just reading a RAT trial (three remotely associated words) triggers spreading activation in the semantic network: Participants who only read the RAT trials recognized solutions to those trials faster than unrelated, random words. However, the authors also found that intentional search for the solution *blocks* spreading activation in the semantic network: participants who intentionally searched for solutions recognized the solutions to those trials as quickly as unrelated, random words. This suggests that on the one hand, just reading RAT trials primes broad activation of the semantic network. Since such conceptual breadth translates into perceptual breadth of attention (Förster and Dannenberg, 2010), reading the RAT trial should broaden the perceptual field of attention. On the other hand, converging on a single solution should block spreading activation in the semantic network, and thus narrow the attentional field. In other words, performing the RAT may have mixed effects on breadth of attention, and therefore, we decided to explore its effects.

Method

Participants and Design

One hundred thirty-eight participants were recruited through the university participant recruitment system and social networks to participate in an experiment on “solving different tasks” (107 females and 31 males). Their age ranged from 19 to 53 years

($M = 26.79$, $SD = 8.31$). Average age did not differ between conditions, $F(2, 133) = 0.812$, $p = 0.446$. To diversify our sample, in this study, apart from student participants ($n = 97$), we also recruited people who were not enrolled at university and who pursued a creative career or had a creative hobby ($n = 38$; background of three participants was not saved due to an internet connection error). Education and main activity of participants are summarized in **Table 1**. Student participants earned credit points for participation. Student and non-student participants could obtain one of seven shopping vouchers worth 50 PLN (around 12 €). Participants were seated at computers separated by screening walls in the laboratory, and were run individually or in small groups (maximally four participants).

Participants were randomly assigned to three conditions of a between-subjects design: divergent thinking ($n = 47$, 38 female, $M_{\text{age}} = 25.78$), analytic thinking ($n = 45$, 33 female, $M_{\text{age}} = 26.67$), and the RAT ($n = 46$, 36 female, $M_{\text{age}} = 28.00$). As a divergent thinking task, we employed the Unusual Uses Task (Silvia et al., 2008). Participants in the analytic thinking task condition were asked to solve a task inspired by a task from the mathematical competition for pupils (Towarzystwo Upowszechniania Wiedzy i Nauk Matematycznych, 2016) and by a task used by Ansburg and Hill (2003). It required the participants to determine the order of men, from the tallest to the smallest. Participants were informed that the men have different height and different eye colors. Three premises were given, which enabled participants to derive a correct solution (e.g., “Adam is not the tallest, and Lucas does not have green eyes”). As a third condition, we used a Polish adaptation of the RAT (Sobków et al., 2016). Eight trials were included.

Measures

Breadth of attention

Participants’ task in each of the conditions was displayed in the middle of the screen on a white background, which was surrounded by a gray frame. Twenty-five peripheral stimuli, geometric shapes and symbols, were displayed on the gray frame, always in the same locations and for the duration of the whole task (see **Figure 3**; the same number of peripheral stimuli – 25 – was used in previous research, e.g., Mendelsohn and Griswold, 1964). Participants were given no information or explanation about why the symbols were there. Breadth of attention was measured by recognition of peripheral stimuli that were displayed on the screen during task performance. The recognition test started after the main task and a mood check had been completed, but participants were not informed earlier that they would perform the recognition test. It included 25 peripheral and 20 filler symbols (i.e., symbols that were not present on the screen during the task solution). Participants indicated whether a symbol was present on the screen during the task solution by pressing a number on the keyboard (1 = *definitely no*, 2 = *rather no*, 3 = *rather yes*, and 4 = *definitely yes*). We recoded these scores into 0 (*no*; score 1 or 2) and 1 (*yes*; score 3 or 4). A recognition index was computed by taking the difference between the percentage of hits (i.e., the proportion of peripheral symbols that were correctly classified as present) and percentage of false alarms (i.e., the proportion of filler symbols that were

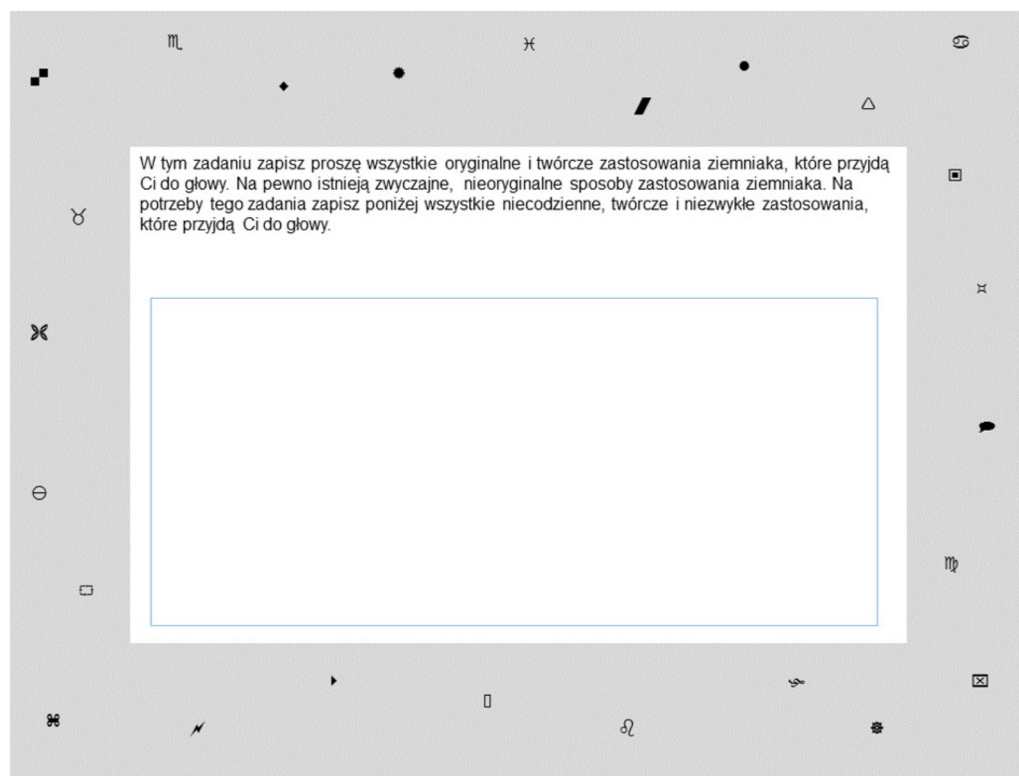


FIGURE 3 | The task with instructions in the heading, response window in the middle of the screen, and symbols in the peripheries of the screen (divergent thinking task condition) in Study 2.

falsely classified as present). Higher recognition index scores indicate more accurate recognition, which means that attention was broader during the task solution. Possible values for this indicator vary from -100 (e.g., 0% of hits and 100% of false alarms) to 100 (e.g., 100% of hits and 0% of false alarms).

Control measures

We employed an Affect Grid to measure mood after the task (Russell et al., 1989). Participants were presented with a square grid divided into 9×9 square fields. The vertical dimension represented arousal, from sleepiness in the lower part to high arousal in the upper part, and the horizontal dimension represented valence, from unpleasant feelings on the left to pleasant feelings on the right. Participants were instructed to click on the field that reflected their feelings most accurately. In this way, we obtained two mood indicators from each participant: valence (ranging from 1 = *unpleasant* to 9 = *pleasant*) and arousal (ranging from 1 = *sleepiness* to 9 = *high arousal*). We also controlled for the subjective difficulty with an adjective scale (Cronbach's $\alpha = 0.87$ for all six items) and enjoyment of the task, using the same measures as in Study 1. Controlling for subjective difficulty is particularly important, because this may affect attention to peripheral stimuli (Santangelo et al., 2011).

Procedure

The study consisted of two parts: online and laboratory. In the online part, participants gave an informed consent and filled

in questionnaires, results of which are not reported in this paper.² Upon arrival in the lab, participants were randomly assigned to the divergent thinking, analytic thinking, or the RAT condition. The experiment was run in the Inquisit Lab5. In the first part, not reported in this paper, participants took part in another experiment, in which they also solved a divergent thinking task, analytic thinking task or the RAT. After the first part, the divergent thinking, analytic thinking task or the RAT was displayed in the middle of the screen and 25 stimuli were displayed in the peripheries. Participants remained in the same condition that they were assigned to in the first part (i.e., they solved a different task of the same type for the second time).

In the divergent thinking task condition, people generated creative uses of a potato (Silvia et al., 2008) and entered their ideas in the field located in the middle of the screen. The time limit was not mentioned, and the task automatically terminated after 180 s. In the analytic thinking task condition, participants ordered four men from the smallest to the tallest, based on the premises given (Ansburg and Hill, 2003; Towarzystwo Upowszechniania Wiedzy i Nauk Matematycznych, 2016). Their task was to write the names of men in the correct order in the field located in the middle of the screen. The time limit set for this task – 360 s – was not mentioned. The longer time was chosen to ensure that it was sufficient and proportional to the difficulty. However,

²Details on these measures may be obtained from the first author.

participants were allowed to finish earlier, on condition that they completed the task (which was not allowed in the divergent thinking task condition). In the RAT condition, participants solved eight RAT trials (Sobków et al., 2016). One trial consisted of three words and a response field displayed on a single screen. The time limit was not mentioned, and each trial automatically terminated after 30 s (Sobków et al., 2016). However, participants were allowed to finish earlier, on condition that they completed the trial. We intended to provide participants in each of the conditions with a similar amount of time to solve the task. Solving all eight RAT trials could take a maximum of 240 s, which was similar to the solution time in other conditions.

Next, all participants performed a mood check (Affect Grid; Russell et al., 1989), and proceeded with the recognition test (“Was this stimulus present on the screen?”). One stimulus at a time was displayed on the screen, and participants responded to all 25 peripheral and 20 filler stimuli in random order. In the end, participants indicated their gender and age, and were thanked for participation.

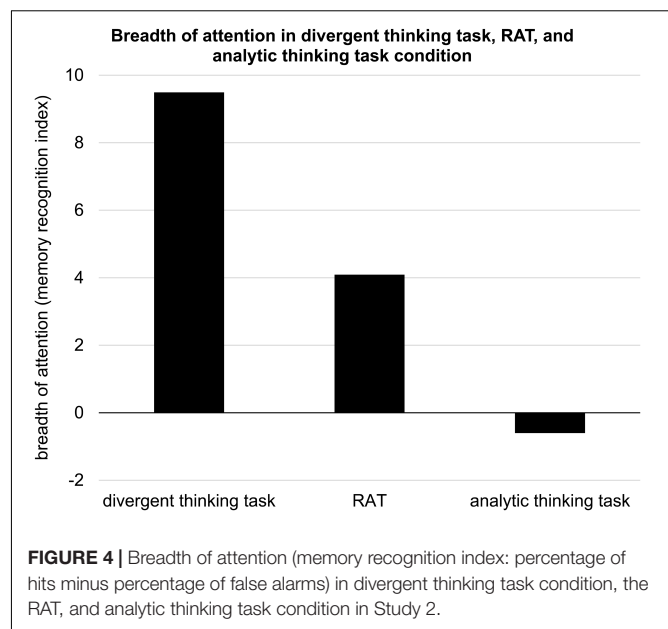
Coding Performance

Divergent thinking task

We trained three independent coders to score three classic indicators of creativity: fluency, flexibility, and originality (Guilford, 1950, 1967). Participants’ responses were typed into a spreadsheet and sorted alphabetically. To obtain fluency measure, the coders counted all generated ideas. To score flexibility, the coders classified each idea into one of 15 categories predefined by the first author and verified with other coders before scoring (e.g., “using potato as a container: making some kind of a container from a potato, where other objects can be stored”). Flexibility of a participant was the number of non-redundant categories in which we could classify the responses. Originality of an idea was rated on a scale from 1 (*not original at all*) to 5 (*very original*), with an original idea defined as “an idea that is infrequent, novel, and original.” Therefore, coders were asked to bear in mind both the objective frequency of a specific idea in a sample, as well as subjective novelty and originality. Originality of a participant was the average originality of all participant’s ideas. A similar coding procedure was employed by De Dreu et al. (2008). The interrater reliability was high: ICC (two-way random model, absolute agreement) for fluency = 0.999, $p < 0.001$, for flexibility ICC = 0.940, $p < 0.001$, and for originality ICC = 0.845, $p < 0.001$ (Koo and Li, 2016). We used the average scores across raters as indicators of divergent thinking performance.

Analytic thinking task

The aim of the analytic task was to order men from the lowest to the tallest (Ansburg and Hill, 2003; Towarzystwo Upowszechniania Wiedzy i Nauk Matematycznych, 2016). Four men were listed in the task. Establishing the correct order required deriving three correct pairings (e.g., Rafael→Adam, Adam→Michael, Michael→Lucas). One point was given for each correct pairing. Therefore, participants could score between 0 and 3 points for the analytic thinking task.



The RAT

Participants scored 1 point for each correctly solved trial (Sobków et al., 2016). Therefore, participants could score between 0 and 8 points in the RAT condition.

Results

Control Variables and Solution Time

We performed separate analyses of variance (ANOVAs) to examine the effects of our manipulation on task enjoyment, subjective difficulty, valence, and arousal, and found no effects (all $ps > 0.128$). On average, participants spent 144 s on solving analytic thinking task ($SD = 57$ s) and 144 s on solving the RAT ($SD = 76$ s). The solution time of divergent thinking task was fixed and was 180 s.

Effect of the Task on Breadth of Attention

We performed a one-way ANOVA with task type (divergent thinking vs. analytic thinking vs. RAT) as independent variable and the recognition index as dependent variable. We found a significant difference in the memory recognition index among the three conditions, $F(2,135) = 4.25$, $p = 0.016$ (see Figure 4). A follow-up simple effects analysis revealed that recognition in the divergent thinking condition ($M = 9.49$, $SD = 17.03$) was significantly higher than in the analytic thinking condition ($M = -0.60$, $SD = 11.95$, $p = 0.013$ with Bonferroni correction). Confidence interval for this comparison did not include zero, 95% CI (1.69, 18.49), and the effect size was moderate (Cohen’s $d = 0.69$). Recognition in the RAT condition ($M = 4.09$, $SD = 19.79$) did not differ significantly from the other conditions ($ps > 0.357$). Confidence intervals for comparisons between RAT and other conditions included zero, 95% CI (−2.95, 13.75) with divergent thinking task and 95% CI (−3.75, 13.13) with analytic thinking task.

In order to verify whether time on task, mood, subjective task difficulty, and enjoyment can account for the influence of

the condition (divergent thinking vs. analytic thinking vs. RAT) on breadth of attention, we performed analysis of covariance with condition as independent variable, recognition index as a dependent variable and solution time, valence, arousal, subjective task difficulty, and enjoyment as covariates. All covariates apart from subjective task difficulty were not significant ($ps > 0.101$). However, subjective task difficulty was a significant covariate, $F(1,129) = 5.44$, $p = 0.021$, but the effect of manipulation (divergent thinking vs. analytic thinking vs. RAT) on recognition index remained significant after controlling for covariates, $F(2, 129) = 5.48$, $p = 0.005$.

Performance

We performed a correlation analysis separately for each condition to test whether recognition (the index of attentional breadth) is related to performance in each of the conditions. Performance was related to recognition in the divergent thinking task condition ($r_{\text{flexibility}} = 0.34$, $p = 0.019$; $r_{\text{originality}} = 0.29$, $p = .047$, $n = 47$) and the RAT condition ($r = 0.44$, $p = 0.002$, $n = 46$), but not in the analytic thinking task condition ($r = 0.14$, $p = 0.352$, $n = 45$).

Discussion of Study 2

Using a different measure of attentional breadth, Study 2 conceptually replicated the findings of Study 1 and strengthened the evidence that engaging in divergent thinking tasks, compared with engaging in analytic thinking tasks, broadens the scope of attention. Interestingly, attentional breadth triggered by the RAT did not differ significantly from attentional breadth triggered by other tasks. A reason for this may be that the RAT can be solved with different strategies that employ more divergent or analytic thinking (Jung-Beeman et al., 2004; Bowden et al., 2005); therefore, the effects of engagement in RAT may vary depending on the method of solution. Furthermore, while reading the RAT triads should broaden attentional field, looking for a single solution is more likely to narrow the attentional breadth (Topolinski and Strack, 2008), and this makes the effects of RAT more similar to the effects of divergent thinking and analytic thinking task, respectively. This also implies that the differences in top-down control state caused by divergent thinking and the RAT may show up in contexts that favor strong but not weak top-down control. For example, these effects may be more pronounced when attention is measured with tasks that favor narrow attentional breadth (see Fischer and Hommel, 2012).

In this study, we found the expected positive relationship between breadth of attention and performance in divergent thinking task. In contrast to Study 1, task performance was not interrupted in this study, and this seems a plausible reason why the correlation was stronger than in Study 1. Similarly to previous findings, analytic thinking performance did not correlate with attentional breadth (Ansburg and Hill, 2003; Liu, 2016). However, we also found a positive relationship between attentional breadth and the RAT performance, which is consistent with the idea that a weaker top-down control state facilitates finding remote

associates (cf. Martindale, 1989; Kolańczyk, 2011; Kenett et al., 2014).

GENERAL DISCUSSION

The ability to process peripheral stimuli together with the ability to broaden the attentional field has been suggested to characterize creative problem solvers (Mendelsohn and Griswold, 1964, 1966; Mendelsohn and Lindholm, 1972; Kasof, 1997; Ansburg and Hill, 2003; Zmigrod et al., 2015). What is more, evidence from experimental research has shown that attentional breadth has a causal effect on creativity (Friedman et al., 2003; Förster et al., 2004; Jia et al., 2009; Colzato et al., 2012, 2017; Liu, 2016; Moraru et al., 2016). However, attentional breadth has not been examined as a *result* of engaging in a creative activity. The present research shows that engaging in creative idea generation indeed broadens the scope of attention compared with engaging in analytic thinking and that this broadened attention relates to higher creative performance. These results suggest that the adjustment of the cognitive system to task demands manifests at a fundamental, perceptual level, through changes in breadth of visual attention. Below, we interpret these results in terms of self-regulation, discuss the limitations of our studies, and suggest questions for further research.

Attentional Breadth as a Self-Regulation Mechanism

Showing the reversed causal relationship between creativity and attentional breadth provides a new perspective, in which attentional breadth has a motivational basis. In this view, attentional breadth is a result of adjusting cognitive processing to task requirements, which ensures effective self-regulation (cf. Bargh et al., 2001). During task engagement, people represent task requirements as their goals, and these goals regulate cognitive processing (e.g., Locke and Latham, 2002; Ferguson et al., 2008). We extend this line of research by showing that attentional breadth results from specific task requirements and may play a self-regulatory role.

This perspective is consistent with several theoretical approaches. For example, the extensive-intensive attention theory (Kolańczyk, 2011, 2012) suggests that ambiguous and ill-defined goals (as in creative idea generation tasks) trigger broad attention, while specific goals (as in analytic thinking tasks) narrow the field of attention. When a goal is ill-defined, broad attention enables exploration and flexibility, which in turn facilitates goal attainment (cf. Johnson et al., 2006). The broaden-and-built theory of positive emotions also points to a similar function of broad attention (Fredrickson and Branigan, 2005). It postulates that attentional breadth results from emotions, with positive emotions broadening the scope of attention and providing room for exploration and novel behaviors. This self-regulatory role of attentional breadth has also been found in the present research: Creative idea generation led to broader attention than analytic thinking, and broad

attention was related to increased creative performance. Therefore, attentional breadth seems to align with task requirements, which may support effective self-regulation and goal attainment.

This line of reasoning is also compatible with construal level theory (Trope and Liberman, 2010), which posits that people represent objects at lower, concrete levels or at higher, abstract levels. The level of representation depends on psychological (temporal or physical) distance between the self and represented objects: the greater the distance, the more abstract and broad object representations. Therefore, construal level adjusts to psychological distance, similar to how attention adjusts to task demands. Indeed, studies have shown that greater temporal, physical, or social distance facilitates global (vs. local) processing, and thus broadens the attentional field (Liberman and Förster, 2009). Similar to broad attention increasing creative performance, temporal (Förster et al., 2004) and physical distance (Jia et al., 2009) also increase creative performance. Therefore, our results are consistent with construal level theory findings. Engaging in creative idea generation, compared with engaging in analytical thinking, is likely to elicit simultaneously higher level construals and a broader attentional field; however, the interdependence of these effects is yet to be examined.

Limitations

Results of our experiments have to be interpreted in the light of some limitations. A first limitation is that the two studies differed on various aspects, including different divergent thinking and analytic thinking tasks (and the inclusion of the RAT only in Study 2), and different measures of attentional breadth. In Study 1, task performance was interrupted and breadth of attention was measured with a separate task, but in Study 2, participants encoded the symbols during the task performance and later reported their recognition of symbols. Therefore, attention was measured without interrupting task performance in Study 2 and participants were not even aware that the recognition of symbols would later be measured. Although findings were consistent in that the divergent thinking task in both studies led to broader attention than the analytical task, one finding was clearly different: breadth of attention correlated with divergent thinking performance in Study 2 but not in Study 1. It is likely that this correlation was not obtained in Study 1 because the measurement of attentional breadth interfered with performance on the divergent thinking task.

Second, we established our effects of type of task on breadth of attention using between-participants designs. We cannot exclude that *a priori* differences between conditions existed in breadth of attention, although such differences should be eliminated by random assignment of participants to conditions. Nonetheless, within-participants designs would offer the opportunity to observe changes in breadth of attention as a consequence of performing a certain task, which would offer strong evidence for the effects of task performance on breadth of attention. One difficulty, however, with such a design is that breadth of attention had to be measured

twice, and preferably with similar tasks, which may be problematic because of learning effects (e.g., peripheral stimuli might be intentionally memorized if the recognition test was anticipated).

Third, and related, although we found that performing a divergent thinking task led to higher breadth of attention than performing an analytical task, we cannot conclude whether the divergent task increased breadth of attention or the analytical task lowered it. Again, a within-participants design may solve the issue. Alternatively, some control condition could be used, although it is not clear *a priori* which tasks would have no effect on breadth of attention and could function as a neutral control condition.

Future Directions

Besides addressing these limitations, we also see other opportunities for future research. An interesting issue relates to different manifestations of cognitive adjustment to task demands. One line of research has linked creativity to the tendency toward global vs. local processing (i.e., whether people perceive an object as a whole or whether they attend to the details of the object; Navon, 1977; Förster and Dannenberg, 2010). Our findings indicate that engaging in creativity can increase breadth of attention, which may relate to more global processing (cf. Hommel, Akbari Chermahini, van den Wildenberg, and Colzato, unpublished manuscript). This was visible, for example, in the identification of clusters of target letters in Study 1, and “jumping” among these clusters, rather than engaging in local and sequential search for single target letters. However, other research has linked creativity to breadth of attention through the functioning of an “attentional filter” (Mendelsohn and Griswold, 1964; Mendelsohn and Lindholm, 1972; Zabelina et al., 2015, 2016). This work proposes that creativity benefits from a “leaky” attentional filter, which allows peripheral stimuli to enter the field of attention. Thus, engaging in a creative task may also lead to increased sensitivity to peripheral cues and a more “leaky” attentional filter, which is consistent with our findings regarding recognition of peripheral stimuli in Study 2. Perhaps *the same* adaptation of attentional breadth to the ongoing situational demands can manifest in *different* ways, and therefore, can be captured with different methods. This is in line with self-regulatory role of attention, and future research can examine this issue more closely.

Future research could also clarify the role of the orienting mechanism of attention – how attention aligns with an internal (e.g., memory structure) or an external sensory (e.g., object from the surroundings) stimulus (Posner, 1980). This mechanism consists of overt and covert orienting. Overt orienting can be observed through head and eye movements, whereas covert orienting occurs when the object of attention changes without eye or head movements. Broad attention can be achieved both through overt (exploring the environment with multiple fixations while thinking about the task solution) as well as covert orienting (enhanced peripheral vision through more global processing while still fixating on the task). Both mechanisms may be responsible for our results, and future work could examine this.

Importantly, Santangelo et al., 2011 (see Santangelo and Spence, 2008 for a review) found that covert orienting toward peripheral cues depends on how (objectively) difficult the main task is in terms of perceptual load (amount of information to be attended to). Even though we controlled for subjective task difficulty, it is likely that perceptual load in each of the tasks in Study 2 was different. For example, the divergent thinking task consisted of a general instruction (“write down all original and creative uses of a brick”) but had no further restrictions (low perceptual load), whereas the analytic thinking task consisted of a task instruction and a set of restrictions which had to be respected in order to reach the solution (high perceptual load). It is therefore possible that perceptual load was responsible for the effect in Study 2. However, in Study 1, attentional breadth was measured independently from the main task – perceptual load during the Ellipses Test was identical across conditions – which is inconsistent with this alternative explanation. Nevertheless, future studies could provide a more nuanced perspective on attentional breadth triggered by divergent and analytical thinking tasks, through manipulations of perceptual load.

Finally, it would be interesting to further investigate the relationship between attentional breadth and solving the RAT (Mednick, 1962). Although we did not detect a difference in the attentional breadth triggered by the RAT vs. other tasks, we did find a positive relationship between breadth of attention and the RAT performance. A possible explanation is that the RAT may involve characteristics of both divergent thinking (e.g., employing various strategies in the search for the solution) and analytic thinking (e.g., arriving at a single correct solution through careful examination of existing options). In contrast to analytic thinking tasks, the pathway to the solution in the RAT is not straightforward and the most obvious associations are often not correct, which requires the solver to look for solutions in multiple directions. Therefore, the RAT may benefit from broad attention, especially when the correct solution is remotely related to all of the three provided words; at the same time, it may trigger narrow attention, because the task instruction emphasizes the goal of finding the single correct solution (cf. Topolinski and Strack, 2008). This explanation is partly supported by Harkins (2006), who showed that inducing greater effort facilitates performance in easy RAT items and inhibits performance in difficult RAT items. The author found that the activation of close associates – narrow attention – was responsible for worse performance on difficult items under increased effort. Further work could examine whether the positive relationship between attentional breadth and the RAT performance holds only for the difficult RAT items or whether it depends on the solution strategy (insightful vs. analytic; Bowden et al., 2005). Additionally, it would be interesting to examine how only reading RAT trials vs. reading and searching for the solutions affects attentional breadth.

Conclusion

The present research showed that engaging in creative idea generation, as compared with engaging in analytic thinking,

broadens the scope of attention. Interestingly, we found that broadened attention also relates to higher performance in creative tasks. Our findings converge with the control-state approach to creativity (Hommel, 2012), in which engaging in creativity triggers stronger or weaker “top-down” guidance, and spills over into how subsequent tasks are performed, depending on whether the goal is to produce multiple different ideas or to arrive at a single correct answer. The present findings shed light on attentional breadth as a self-regulation mechanism: we show that activating a goal embedded in a task leads not only to adjustment of attentional breadth, but that this adjustment may also support task performance. As such, this work indicates that the cognitive system is highly adaptable to task demands and that such adaptation can be observed at the basic, perceptual level, through changes in breadth of visual attention.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

These studies were carried out in accordance with the recommendations of the ethical standards, Ethical Research Committee of SWPS University of Social Sciences and Humanities, Faculty in Sopot, Poland. The protocol was approved by the Ethical Research Committee of SWPS University of Social Sciences and Humanities, Faculty in Sopot, Poland, project number WKE/S 14/V/6. All subjects gave informed consent in a written (Study 1) or online (Study 2) form in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

MW and AK developed the theoretical conception and the study design. MW performed the data collection and analyses. MW, AK, and BN interpreted the results. MW drafted the manuscript. BN and AK revised the manuscript and contributed to the final version.

FUNDING

This research was financially supported by Ministry of Science and Higher Education in Poland from the budget for science in 2014–2018 as a research project within the “Diamond Grant” (grant no. DI2013 010843) awarded to MW and by grant 453-15-002 of the Netherlands Organization for Scientific Research (NWO) awarded to BN.

REFERENCES

- Akbari Chermahini, S., and Hommel, B. (2012). Creative mood swings: divergent and convergent thinking affect mood in opposite ways. *Psychol. Res.* 76, 634–640. doi: 10.1007/s00426-011-0358-z
- Amabile, T. M. (1983). *The Social Psychology of Creativity*. New York, NY: Springer. doi: 10.1007/978-1-4612-5533-8
- Ansburg, P. L., and Hill, K. (2003). Creative and analytic thinkers differ in their use of attentional resources. *Pers. Individ. Dif.* 34, 1141–1152. doi: 10.1016/S0191-8869(02)00104-6
- Arreola, N. J., and Reiter-Palmon, R. (2016). The effect of problem construction creativity on solution creativity across multiple everyday problems. *Psychol. Aesthetics Creat. Arts* 10, 287–295. doi: 10.1037/a0040389
- Baas, M., De Dreu, C. K. W., and Nijstad, B. A. (2008). A meta-analysis of 25 years of mood-creativity research: hedonic tone, activation, or regulatory focus? *Psychol. Bull.* 134, 779–806. doi: 10.1037/a0012815
- Baas, M., Nevelick, B., and Ten Velden, F. S. (2014). Specific mindfulness skills differentially predict creative performance. *Personal. Soc. Psychol. Bull.* 40, 1092–1106. doi: 10.1177/0146167214535813
- Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., and Trötschel, R. (2001). The automated will: nonconscious activation and pursuit of behavioral goals. *J. Pers. Soc. Psychol.* 81, 1014–1027. doi: 10.1037/0022-3514.81.6.1014
- Bowden, E. M., Jung-Beeman, M. J., 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
- Bramesfeld, K. D., and Gasper, K. (2008). Happily putting the pieces together: a test of two explanations for the effects of mood on group-level information processing. *Br. J. Soc. Psychol.* 47, 285–309. doi: 10.1348/000712607X218295
- Bujacz, A., Dunne, S., Fink, D., Gatej, A. R., Karlsson, E., Ruberti, V., et al. (2014). Does creativity make you happy? The influence of creative activity on hedonic and eudaimonic well-being. *J. Eur. Psychol. Stud.* 5, 19–23. doi: 10.5334/jeps.by
- Chiu, F.-C. (2015). Improving your creative potential without awareness: overinclusive thinking training. *Think. Ski. Creat.* 15, 1–12. doi: 10.1016/j.tsc.2014.11.001
- Cohen, J. (1977). *Statistical Power Analysis for Behavioral Sciences*. New York, NY: Academic Press.
- Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 10.1037/0033-295X.82.6.407
- Colzato, L. S., Ozturk, A., and Hommel, B. (2012). Meditate to create: the impact of focused-attention and open-monitoring training on convergent and divergent thinking. *Front. Psychol.* 3:116. doi: 10.3389/fpsyg.2012.00116
- Colzato, L. S., Szapora, A., Lippelt, D., and Hommel, B. (2017). Prior meditation practice modulates performance and strategy use in convergent- and divergent-thinking problems. *Mindfulness* 8, 10–16. doi: 10.1007/s12671-014-0352-9
- De Dreu, C. K. W., Baas, M., and Nijstad, B. A. (2008). Hedonic tone and activation level in the mood-creativity link: toward a dual pathway to creativity model. *J. Pers. Soc. Psychol.* 94, 739–756. doi: 10.1037/0022-3514.94.5.739
- Deuja, A., Kohn, N. W., Paulus, P. B., and Korde, R. M. (2014). Taking a broad perspective before brainstorming. *Gr. Dyn. Theory, Res. Pract.* 18, 222–236. doi: 10.1037/gdn0000008
- Ferguson, M. J., Hassin, R., and Bargh, J. A. (2008). “Implicit motivation: Past, present, and future,” in *Handbook of Motivation Science*, eds J. Y. Shah and W. L. Gardner (New York, NY: Guilford Press), 150–166. doi: 10.1007/s13398-014-0173-7
- Fischer, R., and Hommel, B. (2012). Deep thinking increases task-set shielding and reduces shifting flexibility in dual-task performance. *Cognition* 123, 303–307. doi: 10.1016/j.cognition.2011.11.015
- Förster, J., and Dannenberg, L. (2010). GLOMOsys: a systems account of global versus local processing. *Psychol. Inq.* 21, 175–197. doi: 10.1080/1047840X.2010.487849
- Förster, J., Friedman, R. S., and Liberman, N. (2004). Temporal construal effects on abstract and concrete thinking: consequences for insight and creative cognition. *J. Pers. Soc. Psychol.* 87, 177–189. doi: 10.1037/0022-3514.87.2.177
- Fredrickson, B. L., and Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cogn. Emot.* 19, 313–332. doi: 10.1080/02699930441000238
- Friedman, R. S., Fishbach, A., Förster, J., and Werth, L. (2003). Attentional priming effects on creativity. *Creat. Res. J.* 15, 277–286. doi: 10.1080/10400419.2003.9651420
- Friedman, R. S., and Förster, J. (2002). The influence of approach and avoidance motor actions on creative cognition. *J. Exp. Soc. Psychol.* 38, 41–55. doi: 10.1006/jesp.2001.1488
- Guilford, J. P. (1950). Creativity. *Am. Psychol.* 9, 444–454. doi: 10.1037/h0063487
- Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York, NY: McGraw-Hill.
- Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *J. Pers. Soc. Psychol.* 91, 436–455. doi: 10.1037/0022-3514.91.3.436
- Hommel, B. (2012). “Convergent and divergent operations in cognitive search,” in *Cognitive Search: Evolution, Algorithms, and the Brain*, ed. J. Lupp (Cambridge, MA: The MIT Press), 221–236.
- Hsu, M. L. A., and Fan, H. L. (2010). Organizational innovation climate and creative outcomes: exploring the moderating effect of time pressure. *Creat. Res. J.* 22, 378–386. doi: 10.1080/10400419.2010.523400
- Isen, A. M., and Daubman, K. A. (1984). The influence of affect on categorization. *J. Pers. Soc. Psychol.* 47, 1206–1217. doi: 10.1037//0022-3514.47.6.1206
- Isen, A. M., Daubman, K. A., and Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *J. Pers. Soc. Psychol.* 52, 1122–1131. doi: 10.1037/0022-3514.52.6.1122
- Jarosz, A. F., Colflesh, G. J. H., and Wiley, J. (2012). Uncorking the muse: alcohol intoxication facilitates creative problem solving. *Conscious. Cogn.* 21, 487–493. doi: 10.1016/j.concog.2012.01.002
- Jia, L., Hirt, E. R., and Karpen, S. C. (2009). Lessons from a Faraway land: the effect of spatial distance on creative cognition. *J. Exp. Soc. Psychol.* 45, 1127–1131. doi: 10.1016/j.jesp.2009.05.015
- Johnson, R. E., Chang, C.-H., and Lord, R. G. (2006). Moving from cognition to behavior: what the research says. *Psychol. Bull.* 132, 381–415. doi: 10.1037/0033-2909.132.3.381
- Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biol.* 2:e97. doi: 10.1371/journal.pbio.0020097
- Kasof, J. (1997). Creativity and breadth of attention. *Creat. Res. J.* 10, 303–315. doi: 10.1207/s15326934crj1004_2
- Kenett, Y. N., Anaki, D., and Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Front. Hum. Neurosci.* 8:407. doi: 10.3389/fnhum.2014.00407
- Kolańczyk, A. (1989). How to study creative intuition? *Polish Psychol. Bull.* 20, 149–155.
- Kolańczyk, A. (1991). *Intuicyjność Procesów Przetwarzania Informacji [Intuitiveness of information processing]*. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.
- Kolańczyk, A. (2011). Uwaga ekstensywna. Model ekstensywności vs. intensywności uwagi [Extensive attention. Model of extensiveness vs. intensiveness of attention]. *Stud. Psychol.* 49, 7–27.
- Kolańczyk, A. (2012). “Building bridges in psychology as exemplified by creative intuition,” in *Warsaw Lectures on Personality and Social Psychology. Personality, Cognition and Emotion*, Vol. 2, eds M. W. Eysenck, M. Fajkowska, and T. Maruszewski (New York, NY: Eliot Werner Publications, Inc), 67–88.
- Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012
- Kray, L. J., Galinsky, A. D., and Wong, E. M. (2006). Thinking within the box: the relational processing style elicited by counterfactual mind-sets. *J. Pers. Soc. Psychol.* 91, 33–48. doi: 10.1037/0022-3514.91.1.33
- Kubbandner, C., Lichtenfeld, S., and Pekrun, R. (2011). Always look on the broad side of life: happiness increases the breadth of sensory memory. *Emotion* 11, 958–964. doi: 10.1037/a0024075
- Lebuda, I., Zabelina, D. L., and Karwowski, M. (2016). Mind full of ideas: a meta-analysis of the mindfulness-creativity link. *Pers. Individ. Dif.* 93, 22–26. doi: 10.1016/j.paid.2015.09.040
- Liberman, N., and Förster, J. (2009). The effect of psychological distance on perceptual level of construal. *Cogn. Sci.* 33, 1330–1341. doi: 10.1111/j.1551-6709.2009.01061.x

- Liu, S. (2016). Broaden the mind before ideation: the effect of conceptual attention scope on creativity. *Think. Ski. Creat.* 22, 190–200. doi: 10.1016/j.tsc.2016.10.004
- Locke, E. A., and Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. *Am. Psychol.* 57, 705–717. doi: 10.1037//0003-066X.57.9.705
- Martindale, C. (1989). "Personality, situation, and creativity," in *Handbook of Creativity: Perspectives on Individual Differences*, ed. R. T. Brown (New York, NY: Plenum Press), 211–232.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* 69, 220–232. doi: 10.1037/h0048850
- Mendelsohn, G. A. (1976). Associative and attentional processes in creative performance. *J. Pers.* 44, 341–369. doi: 10.1111/j.1467-6494.1976.tb00127.x
- Mendelsohn, G. A., and Griswold, B. B. (1964). Differential use of incidental stimuli in problem solving as a function of creativity. *J. Abnorm. Soc. Psychol.* 68, 431–436. doi: 10.1037/h0040166
- Mendelsohn, G. A., and Griswold, B. B. (1966). Assessed creative potential, vocabulary level, and sex as predictors of the use of incidental cues in verbal problem solving. *J. Pers. Soc. Psychol.* 4, 423–431. doi: 10.1037/h0023783
- Mendelsohn, G. A., and Lindholm, E. P. (1972). Individual differences and the role of attention in the use of cues in verbal problem solving. *J. Pers.* 40, 226–241. doi: 10.1111/j.1467-6494.1972.tb01000.x
- Moraru, A., Memmert, D., and van der Kamp, J. (2016). Motor creativity: the roles of attention breadth and working memory in a divergent doing task. *J. Cogn. Psychol.* 5911, 1–12. doi: 10.1080/20445911.2016.1201084
- Navon, D. (1977). Forest before trees: the precedence of global features in visual perception. *Cogn. Psychol.* 9, 353–383. doi: 10.1016/0010-0285(77)90012-3
- Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., and Baas, M. (2010). The dual pathway to creativity model: creative ideation as a function of flexibility and persistence. *Eur. Rev. Soc. Psychol.* 21, 34–77. doi: 10.1080/10463281003765323
- Posner, M. (1980). Orienting of Attention. *Q. J. Exp. Psychol.* 32, 3–25. doi: 10.1080/0033558008248231
- Princeton Review (2015). *LAST sections*. Available at: <http://www.princetonreview.com/law/lsat-sections> [Accessed June 24, 2015].
- Roczniewska, M., Sterczyński, R., Popławska, A., Szamotulska, B., and Kolańczyk, A. (2011). Ellipses Test - a new research tool to measure extensive vs intensive attention. *Stud. Psychol.* 49, 115–128. doi: 10.2478/v10167-010-0031-y
- Russell, J. A., Weiss, A., and Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *J. Pers. Soc. Psychol.* 57, 493–502. doi: 10.1037/0022-3514.57.3.493
- Santangelo, V., Botta, F., Lupiáez, J., and Spence, C. (2011). The time course of attentional capture under dual-task conditions. *Attent. Percept Psychophys.* 73, 15–23. doi: 10.3758/s13414-010-0017-2
- Santangelo, V., and Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. *Conscious. Cogn.* 17, 989–1015. doi: 10.1016/j.concog.2008.02.006
- Schmid, P. C., Mast, M. S., Bombardieri, D., Mast, F. W., and Lobmaier, J. S. (2011). How mood states affect information processing during facial emotion recognition: an eye tracking study. *Swiss J. Psychol.* 70, 223–231. doi: 10.1024/1421-0185/a000060
- Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., et al. (2008). Assessing creativity with divergent thinking tasks: exploring the reliability and validity of new subjective scoring methods. *Psychol. Aesthetics Creat. Arts* 2, 68–85. doi: 10.1037/1931-3896.2.2.68
- Sobków, A., Poleć, A., and Nosal, C. (2016). RAT-PL - Konstrukcja i walidacja polskiej wersji Testu Odległych Skojarzeń. *Stud. Psychol.* 54, 1–13.
- Topolinski, S., and Strack, F. (2008). Where there's a will—there's no intuition. *The unintentional basis of semantic coherence judgments. J. Mem. Lang.* 58, 1032–1048. doi: 10.1016/j.jml.2008.01.002
- Towarzystwo Upowszechniania Wiedzy i Nauk Matematycznych (2016). *Kangur Matematyczny - Przykładowe Zadania - rączek*. Available at: http://www.kangur-mat.pl/zad_przy_zaczek.php [Accessed December 13, 2016].
- Trope, Y., and Liberman, N. (2010). Construal-level theory of psychological distance. *Psychol. Rev.* 117, 440–463. doi: 10.1037/a0018963
- Vartanian, O. (2009). Variable attention facilitates creative problem solving. *Psychol. Aesthetics Creat. Arts* 3, 57–59. doi: 10.1037/a0014781
- Wojciszke, B., and Baryła, W. (2004). Skale do pomiaru nastroju i szczęścia emocji. *Czas. Psychol.* 10, 1–17.
- Zabelina, D. L., O'Leary, D., Pornpattananangkul, N., Nusslock, R., and Beeman, M. (2015). Creativity and sensory gating indexed by the P50: selective versus leaky sensory gating in divergent thinkers and creative achievers. *Neuropsychologia* 69, 77–84. doi: 10.1016/j.neuropsychologia.2015.01.034
- Zabelina, D. L., Saporta, A., and Beeman, M. (2016). Flexible or leaky attention in creative people? Distinct patterns of attention for different types of creative thinking. *Mem. Cogn.* 44, 488–498. doi: 10.3758/s13421-015-0569-4
- Zhou, Y., Zhang, Y., Hommel, B., and Zhang, H. (2017). The impact of bodily states on divergent thinking: evidence for a control-depletion account. *Front. Psychol.* 8:1546. doi: 10.3389/fpsyg.2017.01546
- Zmigrod, S., Zmigrod, L., and Hommel, B. (2015). Zooming into creativity: individual differences in attentional global-local biases are linked to creative thinking. *Front. Psychol.* 6:1647. doi: 10.3389/fpsyg.2015.01647

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 Wronska, Kolańczyk and Nijstad. 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.



Investigating the Role of the Primary Motor Cortex in Musical Creativity: A Transcranial Direct Current Stimulation Study

Aydin Anic^{1,2*}, Kirk N. Olsen^{1,2} and William Forde Thompson^{1,2,3}

¹ Department of Psychology, Macquarie University, Sydney, NSW, Australia, ² Centre for Elite Performance, Expertise and Training, Macquarie University, Sydney, NSW, Australia, ³ Australian Research Council Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, NSW, Australia

OPEN ACCESS

Edited by:

Michele Biasutti,
Università degli Studi di Padova, Italy

Reviewed by:

Maria Herrojo Ruiz,
Goldsmiths, University of London,
United Kingdom
Barbara Colombo,
Champlain College, United States

*Correspondence:

Aydin Anic
aydin.anic@mq.edu.au

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 15 May 2018

Accepted: 30 August 2018

Published: 01 October 2018

Citation:

Anic A, Olsen KN and Thompson WF
(2018) Investigating the Role of the
Primary Motor Cortex in Musical
Creativity: A Transcranial Direct
Current Stimulation Study.
Front. Psychol. 9:1758.
doi: 10.3389/fpsyg.2018.01758

Neuroscientific research has revealed interconnected brain networks implicated in musical creativity, such as the executive control network, the default mode network, and premotor cortices. The present study employed brain stimulation to evaluate the role of the primary motor cortex (M1) in creative and technically fluent jazz piano improvisations. We implemented transcranial direct current stimulation (tDCS) to alter the neural activation patterns of the left hemispheric M1 whilst pianists performed improvisations with their right hand. Two groups of expert jazz pianists ($n = 8$ per group) performed five improvisations in each of two blocks. In Block 1, they improvised in the absence of brain stimulation. In Block 2, one group received inhibitory tDCS and the second group received excitatory tDCS while performing five new improvisations. Three independent expert-musicians judged the 160 performances on creativity and technical fluency using a 10-point Likert scale. As the M1 is involved in the acquisition and consolidation of motor skills and the control of hand orientation and velocity, we predicted that excitatory tDCS would increase the quality of improvisations relative to inhibitory tDCS. Indeed, improvisations under conditions of excitatory tDCS were rated as significantly more creative than those under conditions of inhibitory tDCS. A music analysis indicated that excitatory tDCS elicited improvisations with greater pitch range and number/variety of notes. Ratings of technical fluency did not differ significantly between tDCS groups. We discuss plausible mechanisms by which the M1 region contributes to musical creativity.

Keywords: creativity, expertise, musical improvisation, primary motor cortex, transcranial direct current stimulation

Abbreviations: ACC, anterior cingulate cortex; AIC, anterior insula cortex; DLPFC, dorsolateral prefrontal cortex; DMN, default mode network; dPMC, dorsal premotor cortex; ECN, executive control network; EEG, electroencephalography; fMRI, functional magnetic resonance imaging; IHIC, inter-hemispheric inhibition connection; M1, primary motor cortex; MEP, motor-evoked potential; PCC, posterior cingulate cortex; PFC, prefrontal cortex; pre-SMA, pre-supplementary motor area; rTMS, repetitive transcranial magnetic stimulation; TMS, transcranial magnetic stimulation; tDCS, transcranial direct current stimulation; vMPFC, ventromedial prefrontal cortex; vPMC, ventral premotor cortex.

INTRODUCTION

The ability of humans to generate novel ideas has fascinated scientists and philosophers for centuries. Such ideas are defined as *creative* when they involve both novelty and congruency (Benedek et al., 2014; Schwab et al., 2014). Novelty pertains to the originality of a specific idea; congruency is said to have occurred if an idea is contextually appropriate (Dietrich, 2004; Jauk et al., 2015). Other theorists include a third defining feature, arguing that acts can only be considered creative if they are also *non-obvious* (Boden, 2004).

Creative thought and behavior have significant implications for human life, and a large body of research has focused on understanding psychological mechanisms that underpin the creative process (e.g., Batey and Furnham, 2006; Simonton, 2010; Jauk et al., 2014; Tan et al., 2016). Over the past 10 years, researchers have begun to reveal the neural underpinnings of creative thought and action, employing methods such as fMRI (e.g., Limb and Braun, 2008) and EEG (e.g., Fink and Benedek, 2014). The present investigation used a novel method of online bihemispheric tDCS to investigate the neuroscience of creativity in the context of artistic enactment (Lucchiari et al., 2018). Specifically, tDCS was used to investigate the role of the M1 in creative piano improvisations performed by expert jazz pianists.

Musical improvisation represents an ecologically valid domain in which to explore the process of creativity because it requires novelty and continuous production of non-obvious but contextually appropriate passages of music (Bengtsson et al., 2007). Musical improvisation is a form of creative expression that can be defined as the composition or invention of music in real-time (Biasutti, 2015, 2017). Its implementation in real time means that no corrections can be made to creative output. Instead, improvisation is a temporally dynamic behavior that unfolds over time (Biasutti, 2015; Adhikari et al., 2016).

Improvisation plays a role in many genres of music but is most prominent in jazz, where musicians routinely generate novel melodies while observing complex rhythmic and harmonic templates that can be modulated to generate creative output (Biasutti and Frezza, 2009). In the context of neuroscientific research, musical improvisation is commonly used in studies designed to highlight brain networks involved in movement-based creativity (e.g., Bengtsson et al., 2007; Limb and Braun, 2008; Pinho et al., 2014). However, this research has focused primarily on regions of the brain involved in higher-order cognitive processing, without consideration of the M1. The M1 is usually known for low-level functions such as motor learning and consolidation of motor skills (Karok and Witney, 2013; Sosnik et al., 2014), yet its role in creativity is unknown.

Previous fMRI studies investigating the neural mechanisms that underpin musical creativity often report activation of the ECN (Bengtsson et al., 2007). The ECN is located in the frontal lobe and comprises the DLPFC, ACC, and AIC (Kuhn et al., 2013). The ECN mediates three distinct cognitive mechanisms associated with creativity: inhibition, working memory, and cognitive flexibility (Diamond, 2013; Sowden et al., 2015; Bendetowicz et al., 2017; Kenett et al., 2018). The DLPFC is particularly important in mediating attention, working memory,

and goal-orientation (Boccia et al., 2015). The DMN is another neural network that underpins creative cognition in a musical context, yet operates in direct contrast to the ECN (Limb and Braun, 2008). The DMN is a combination of brain areas that include the vMPFC, the PCC, and the medial and lateral temporal lobes (Kuhn et al., 2013; Zhu et al., 2017). The vMPFC is of particular importance since it mediates mind wandering, future imagination, and is activated during tasks requiring musical creativity (Limb and Braun, 2008; Bashwiler et al., 2016; Kenett et al., 2018).

The PFC and specifically the DMN and ECN are of paramount importance to processes involved in creative cognition and behavior. This is true irrespective of the domain (e.g., artistic creativity vs. insightful problem solving; Gonen-Yaacovi et al., 2013). Moreover, the output of information processed by the DLPFC that forms part of the ECN branches to the motor cortices (Dietrich, 2004). To date, it is known that premotor cortices such as the pre-SMA and the ventral and dorsal counterparts of the premotor cortex (vPMC and dPMC, respectively) are involved in high-level motor planning and execution (Berkowitz and Ansari, 2008; de Manzano and Ullén, 2012; Sosnik et al., 2014). The pre-SMA is important in the temporal components of motor performance, whereas the vPMC and dPMC are both involved in selection and performance of novel motor outputs – features that are vitally important for creative improvisation in music performance (Chouinard and Paus, 2006; Hoshi and Tanji, 2007; Berkowitz and Ansari, 2008; de Manzano and Ullén, 2012).

It is clear from this brief review that some of the brain networks that underpin creative musical improvisations are associated with higher-order cognitive processing and motor planning. It is not yet clear, however, whether brain regions involved in low-level processes such as the M1 also play a significant role in creative musical performance. The M1 is important for motor acquisition, consolidation, and importantly for pianists, the orientation, velocity, and direction of movement in the arms and hands (Karok and Witney, 2013; Sosnik et al., 2014). Stimulation of the M1 also results in greater muscular synergies in the hand that enhance the ability to “generate novel patterns of muscular activity” (Waters-Metenier et al., 2014, p. 1037). Indeed, creativity in performances that require rapid changes in the muscular activity in the hand may be modulated by the M1 in two important ways. First, precise temporal and spatial hand movements are required for technically fluent piano performances. It is likely that with high levels of technical fluency comes the increased probability of realizing creative cognition through performed improvisation. Second, the M1 may function directly to control the implementation of motor plans arising from higher-order processes, acting as a neural gateway that impacts upon creative artistic enactment (Lucchiari et al., 2018). The present study was specifically designed to address these overarching hypotheses by investigating the role of the M1 in creative and technically fluent piano improvisations. The improvisations were performed by expert jazz pianists and creativity and technical fluency were adjudicated by expert musician adjudicators (see Anic et al., 2017 for pilot data).

The M1 is located in both hemispheres of the brain. The left hemispheric M1 tends to exert superior control of the right hand,

whereas the right hemispheric M1 tends to exert superior control of the left hand (Brinkman and Kuypers, 1973; Vines et al., 2008b). The two hemispheres of the M1 are linked by an IHIC. When the left hemispheric M1 is activated during movement in the right hand, the right hemispheric M1 is naturally inhibited through the IHIC to facilitate right-handed movement (see also van den Berg et al., 2011).

In the present study, we investigated whether excitatory tDCS over the left hemispheric M1 enhances creativity and technical fluency of right-handed piano improvisations, when compared with inhibitory tDCS. If creativity is modulated by the M1, then creativity and technical fluency in right-handed piano improvisations should vary as a function of the type of tDCS administered to the left hemispheric M1. Specifically, we hypothesized that excitatory tDCS over the left M1 will result in an increase in creativity and technical fluency compared to inhibitory tDCS. A subsidiary aim was to examine the correlation between ratings of creativity and technical fluency by expert musician adjudicators.

MATERIALS AND METHODS

Participants

Sixteen proficient jazz pianists ($M = 24.1$ years, $SD = 7.2$, 7 females) and three independent expert musical adjudicators were recruited for the study. Each musician produced 10 improvisations which were judged on two separate scales by all three adjudicators, resulting in a total of 960 ratings that were then subjected for analysis. Three of the 16 proficient jazz pianists reported to be left-handed; one reported to be mixed-handed. All pianists had undergone considerable formal musical training on piano ($M = 9.6$ years, $SD = 4.4$). A TMS safety screener with a series of health-related questions (e.g., do you, or anyone in your family, have epilepsy?) was administered to participants prior to tDCS stimulation to ensure the safe application of brain stimulation. All participants satisfied the requirements of the safety screener and no participant subsequently experienced adverse effects from the procedure. The pianists were reimbursed \$50 or course credit for an undergraduate psychology unit for their participation. Three expert musicians were recruited as judges to rate the improvisations. All three judges had completed doctoral level education in music-related fields, had received an average of 12.67 years of formal music training (8, 10, and 20 years), and were experienced as adjudicators of music performances. The judges were independent in that they did not know each other and did not adjudicate the performances together. They were reimbursed up to \$150 for the approximate time of 3 h to adjudicate the performances. All participants and judges gave informed consent and the study was approved by the Macquarie University Human Research Ethics Committee (HREC Reference number: 5201600392).

Stimuli

Ten short pieces of music were custom-written by the first author (AA) for this study using *Notion* (Version 2.0.183) music

software. These pieces were written to conform to a quintessential contemporary jazz style and provided participants with a musical context from which to perform their improvisations. All pieces incorporated an electronic drum kit, electric piano, grand piano from the *GarageBand* (Version 10.2.0) music software, and a live electric bass was played and recorded by the first author. Each musical piece contained 10 bars and lasted 30 s in total. An example score is shown in **Figure 1**. In each score, the first bar provided a four-beat count-in with an electronic high-hat cymbal on the drum kit to prepare participants for the beginning of the performance. Bars 2–5, labeled by the rehearsal marker “A” in **Figure 1**, contained a custom-written novel melody with the electronic drum kit, electronic piano and live electric bass acting as accompaniment for the harmonic and rhythmic qualities. In this “sight-reading” section, participants were instructed to reproduce the melody as accurately as possible, only on the treble clef and only with their right hand. Bars 6–10, labeled with the rehearsal marker “B” in **Figure 1**, comprised the improvisation section of the piece. In the section B – the “improvisation” section – the custom-written melody in section “A” was removed but the instrumental accompaniment remained to ensure rhythmic and harmonic quality and consistency. The participants were instructed to only use their right hand for both the sight-reading and improvisation sections. Seven of the 10 pieces were written in major key signatures (A, B, C \times 2, D \times 2, E^b); the remaining three pieces were written in minor key signatures (B, D, G). All pieces were written in a 4:4 time signature with a swing feel at 90 beats per minute. See **Supplementary Material** for the scores of all 10 pieces.

Equipment

The tDCS montage used in the present study comprised two saline-soaked electrodes diametric in charge: anode (positive) and cathode (negative) (Nitsche et al., 2003; Colombo et al., 2015). The anode charge heightens neural activity, whereas the cathode charge inhibits neural activity (Nitsche et al., 2003). An online bihemispheric tDCS configuration was implemented where both electrodes were placed on the scalp to stimulate the left and right M1 while each participant was engaged in the experimental task (see the “Experimental Design” subsection below for more detail). A study was conducted by Karok and Witney (2013) to determine the optimal tDCS configuration (placement of electrodes) and mode of tDCS (offline vs. online), and found that online bihemispheric tDCS is the optimal method for experiments designed to elicit significant changes in neural activity and subsequent behavior (see also Vines et al., 2008a; Waters-Metenier et al., 2014).

The online bihemispheric tDCS montage was set at 1.4 mA using two 25 cm² electrodes to ensure a current density of 0.056/cm², as recommended in Bikson et al. (2009). The saline-soaked electrodes (the anode and cathode) were attached onto an electroencephalogram (EEG) cap and worn by the participants with the tDCS device attached to the back of the cap. In accordance to the 10–20 EEG system, the electrodes were placed



FIGURE 1 | This figure shows an example of a musical score presented to participants in a trial. This example is written in the key of D major at a tempo of 90 beats per minute. The score shown at Section A marks the beginning of the sight-reading stage, where participants were instructed to play the displayed melody as accurately as possible. The score shown at Section B marks the beginning of the improvisation stage, where participants were instructed to play an improvisation based on the melody in the previous section and the harmonic structure of the music. Participants were instructed to play with their right-hand only.

on the C3 and C4 electrode sites with the Cz electrode site situated on top of the scalp (Karak and Witney, 2013).

A 27-inch iMac was used to present each score to participants during each trial. The iMac was connected via a thunderbolt cable to a MacBook Air that played each piece of music and recorded each performance. All performances were conducted on a MIDI keyboard that was connected via USB to the MacBook Air. An additional MacBook Pro was used to run the Neuro-electrics Instrument Controller (NIC) (Version 1.4.10) that controlled the configuration and stimulation for the tDCS device. Once configured on the NIC software, the tDCS device was connected remotely to the MacBook Pro via Bluetooth and was attached to the cap on the back of the participants' head. Each piece of music was played to participants through two external computer speakers.

Experimental Design

Two tDCS stimulation conditions were developed for the experiment: Anodal-Left M1/Cathodal-Right M1 (excitatory tDCS group, $n = 8$) and Cathodal-Left M1/Anodal-Right M1 (inhibitory tDCS group, $n = 8$). These tDCS conditions were developed to target the right hand of participants. The 16 participants were pseudo-randomized into the two conditions to ensure an equal distribution of participants in the conditions. The 10 musical pieces were used to create 10 experiment trials (one piece per trial) that were further subdivided into two blocks. Block 1 contained five pieces (five trials or "takes") to perform *without* tDCS stimulation. This served to evaluate a baseline rating of creativity and technical fluency under normal (no brain stimulation) performance conditions. Block 2 contained the remaining five pieces to perform during tDCS stimulation. The set of 10 pieces were initially randomly placed into the two blocks and to mitigate order effects, were further randomized within each

block for each participant. In Block 2 – the stimulation block – either excitatory or inhibitory tDCS was applied to the participant's left hemispheric M1, depending on the tDCS group they were placed in prior to the commencement of the experiment. Participants were blind to the type of tDCS stimulation they received and were tested individually in separate sessions. The duration of the experiment lasted approximately 90 min.

Procedure

First, the TMS screener was administered to ensure that tDCS brain stimulation was safe to administer. Participants then gave informed consent to participate in the experiment and completed a demographic questionnaire. After this, participants completed five trials in the no-stimulation Block 1. Each trial consisted of two stages: familiarization and performance. The familiarization stage involved two practice runs for each piece of music in each trial. The first practice run involved the participant listening to the piece and following the melody in section "A" on the score *without* playing the piano. The melody in section A was played by a grand piano in the recording in addition to the musical accompaniment outlined above. The second practice run required the participant to play the displayed melody in section "A" with their right hand. The purpose of the familiarization stage was to ensure that participants were familiar with the piece of music in each trial.

After the familiarization stage of each trial, the performance stage commenced. The performance stage involved two complete attempts at each trial. The first performances in each trial in this stage were sent to the expert judges for adjudication, except for one trial from one participant who made significant errors in their improvisation and stopped playing. The purpose of allowing the participant to complete a second attempt

at each trial was to reduce performance anxiety. The grand piano that played the melody in section “A” during the familiarization stage was removed in the performance stage. Each participant was instructed to play the melody in section “A” as accurately as possible. This enabled us to evaluate indicators of sight-reading accuracy such as timing (asynchrony of each note played relative to expected timing as stipulated in the score) and pitch-note accuracy (whether a correct note was played relative to each note in the score). They were instructed to perform their right-handed improvisations in section “B.”

After completing five trials in the no-stimulation Block 1, participants were administered the online bihemispheric tDCS montage specific to their allocated condition (Anodal-Left M1/Cathodal-Right M1 or Cathodal-Left M1/Anodal-Right M1). The first 30 s of stimulation involved a “ramp-up” period. All participants were stimulated for two and a half minutes (including ramp-up) before completing the final five trials in Block 2. This duration was to ensure a considerable level of stimulation was reached before performance began. The final 30 s of stimulation involved a “ramp-down” period. Participants were stimulated between a range of 15 and 21 min in total. This variation in stimulation time was due to the difference in time participants required to work through the familiarization stage of each trial in Block 2. Nevertheless, the stimulation duration and level of tDCS used in the present study remained well within safe limits (Bikson et al., 2009).

To ensure that participants were familiar with the experimental procedure, two complete practice trials were administered before the 10 experiment trials. The pieces of music in the practice trials were not used in the experiment trials. All performances were recorded using *GarageBand* (Version 10.2.0) on the MacBook Air and audio recordings were all formatted to ACC audio, de-identified, and randomly placed in a list of 160 performances for each judge to adjudicate.

Expert Adjudication of Performances

The judges were provided with specific instructions and definitions for creativity and technical fluency to minimize ambiguity in judging. Creativity was defined as the quality of being *novel* and *appropriate* within a specific context. Technical fluency was defined as the level of accuracy and musicianship of the performances that may include accuracy in pitch and rhythm, articulation, and phrasing. Judges were instructed to rate the creativity and technical fluency of each of the 160 performances on two separate Likert scales ranging from 1 to 10. A score of 1 represented a low score on creativity or technical fluency; a score of 10 represented a very high level of creativity or technical fluency. Adjudicators were blind to the experimental conditions associated with each performance and did not know the true aim of the experiment or details about the participants’ musical background and training.

Statistical Approach

To assess the consistency of ratings for creativity and technical fluency, a multiple-raters, consistency, 2-way mixed effects

intra-class correlation coefficient (ICC) model was computed for the three independent judges across 16 participants. We conducted statistical tests to assess the reliability of differences in ratings of creativity and technical fluency between the two tDCS groups (excitatory vs. inhibitory). This comparison was first done for block one (no stimulation to either group) and again for block two (excitatory vs. inhibitory stimulation). To account for potential differences in the three judges’ assessments and differences as a function of the five consecutive attempts to improvise in each tDCS condition, we conducted two $2 \times 3 \times 5$ mixed-ANOVAs, with Stimulation Group as the between-subjects factor (hereafter Group: excitatory or inhibitory), and Judge (1–3) and Take (1–5) as repeated measures factors. The first mixed-ANOVA analyzing the data from Block 1 was designed to check whether performances were similar across groups when no tDCS was administered. The second mixed-ANOVA analyzing the data from Block 2 was conducted to assess whether excitatory tDCS over the left M1 region resulted in performances that were rated by adjudicators as more creative and technically fluent than for those who received inhibitory tDCS. This approach ensured that all 960 data points from 16 participants were included in the analyses (160 performances rated by three adjudicators on creativity and technical fluency).

A Pearson’s r correlation coefficient was calculated to examine the association between mean creativity and technical fluency ratings averaged across the three judges. Structural analyses of improvisations were conducted and independent samples t -tests were computed to investigate any differences between the two tDCS groups with respect to the following three performance features: number of notes, pitch range, and number of different notes. Two multiple linear regressions were also computed to determine whether there was an association between these three performance features and ratings of creativity and technical fluency for the improvisations produced under conditions of tDCS in Block 2 (excitatory and inhibitory). Lastly, two components of sight-reading accuracy – pitch and timing accuracy – were recorded for all performances during the “sight-reading” stage of each trial. Timing accuracy was measured in milliseconds as an asynchrony between each performed note and the specific timing of each note as stipulated by the score. Pitch-note accuracy was coded as “0” each time participants pressed the correct piano key corresponding to each pitch in the score. Pitch-note accuracy was coded as “1” each time participants pressed the incorrect piano key relative to each note in the score. Therefore, the higher the score, the more inaccurate the sight-reading performance. Two independent samples t -tests were computed to analyze the sight-reading accuracy for both tDCS groups.

RESULTS

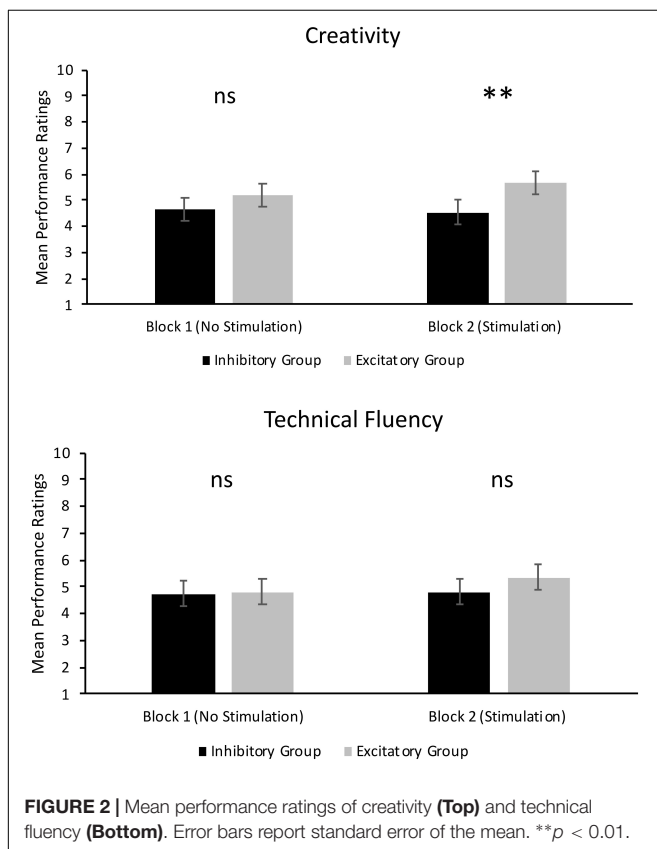
Ratings of Creativity

The mean ICC for creativity was 0.507 with a 95% confidence interval from 0.358 to 0.626, $F(159,318) = 2.029$, $p < 0.001$. Therefore, inter-rater reliability for ratings of creativity across

the three judges can be considered “fair” (Cicchetti, 1994). The first mixed-ANOVA analyzing the data from Block 1 was designed to assess whether ratings of performances were similar across groups when no tDCS was administered, as well as to monitor any differences between adjudicators or between the five consecutive improvisations in each condition. The second mixed-ANOVA analyzing the data from Block 2 assessed whether excitatory tDCS over the left M1 region resulted in performances that were rated by the adjudicators as more creative than for those who received inhibitory tDCS.

Block 1 (No Stimulation)

As can be seen in the top panel of **Figure 2**, there was no significant main effect of Group in Block 1, $F(1,14) = 1.21$, $p = 0.290$, $\eta_p^2 = 0.08$. Thus, ratings of creativity in the excitatory tDCS group ($M = 5.18$, $SD = 1.69$) were not significantly different to ratings of creativity in the inhibitory tDCS group ($M = 4.67$, $SD = 1.73$) under conditions where no tDCS was administered. There was, however, a significant main effect of Judge in Block 1, $F(2,28) = 20.97$, $p < 0.001$, $\eta_p^2 = 0.60$. The mean rating of creativity from Judge 2 ($M = 3.78$, $SD = 2.07$) was significantly lower than that of Judge 1 ($M = 5.43$, $SD = 1.63$) and Judge 3 ($M = 5.58$, $SD = 1.47$, $p < 0.001$). There was no significant difference between mean ratings from Judge 1 and 3 ($p = 0.511$). There were no other significant effects.

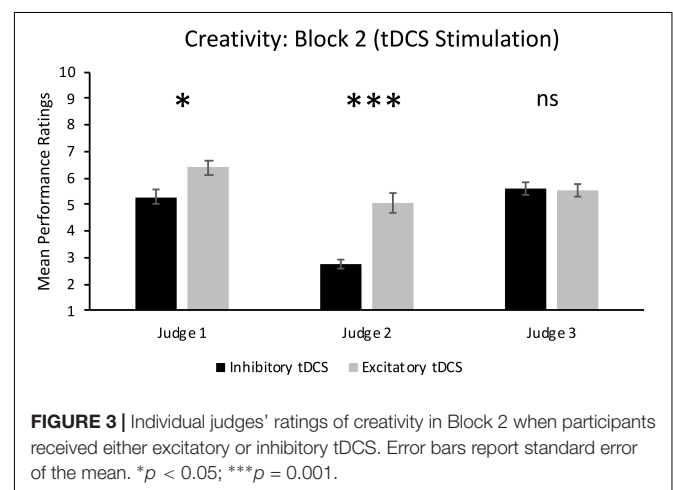


Block 2 (Stimulation)

As can also be seen in the top panel of **Figure 2**, there was a significant main effect of Group in Block 2, $F(1,14) = 10.50$, $p = 0.006$, $\eta_p^2 = 0.43$. This result supports our hypothesis and shows that jazz improvisation performances by participants who received excitatory tDCS were rated significantly more creative ($M = 5.68$, $SD = 1.80$) than performances by participants who received inhibitory tDCS ($M = 4.55$, $SD = 1.91$). However, there was a significant Group \times Judge interaction, $F(2,28) = 10.35$, $p < 0.001$, $\eta_p^2 = 0.43$. As can be seen in **Figure 3**, mean ratings of creativity were significantly greater in the excitatory tDCS condition relative to the inhibitory tDCS condition from Judge 1, $t(14) = 2.35$, $p = 0.034$, 95% CI [0.097, 2.153], and from Judge 2, $t(14) = 4.46$, $p = 0.001$, 95% CI [1.206, 3.444], but not from Judge 3, $t(14) = -0.20$, $p = 0.844$, 95% CI [-0.870, 0.720]. Overall, these findings appear to reflect both excitatory and inhibitory effects: six of the eight participants who received excitatory tDCS in Block 2 exhibited an absolute increase in rated creativity relative to Block 1 (no-stimulation), and four of the eight participants who received inhibitory tDCS in Block 2 exhibited an absolute decrease in rated creativity relative to Block 1.

Ratings of Technical Fluency

The mean ICC for technical fluency was 0.475 with a 95% confidence interval from 0.317 to 0.602, $F(159,318) = 1.906$, $p < 0.001$. This result suggests that inter-rater reliability for ratings of technical fluency across the three judges can also be considered “fair” (Cicchetti, 1994). Similar to analyses of creativity, the first mixed-ANOVA analyzing the technical fluency data from Block 1 was designed to assess whether performances were similar across groups when no tDCS was administered, as well as to monitor any differences between adjudicators or between the five consecutive improvisations in each condition. The second mixed-ANOVA analyzing the data from Block 2 assessed whether excitatory tDCS over the left M1 region resulted in performances that were rated by the adjudicators as more technically fluent than for those who received inhibitory tDCS.



Block 1 (No Stimulation)

As can be seen in the bottom panel of **Figure 2**, there was no significant main effect of Group in Block 1, $F(1,14) = 0.05$, $p = 0.832$, $\eta_p^2 = 0.00$. Thus, ratings of technical fluency were not significantly different between the excitatory tDCS group ($M = 4.82$, $SD = 1.81$) and the inhibitory tDCS group ($M = 4.74$, $SD = 1.76$) under conditions where no tDCS was administered. There was also a significant main effect of Judge, $F(2,28) = 44.22$, $p < 0.001$, $\eta_p^2 = 0.76$. Similar to ratings of creativity, the mean rating of technical fluency from Judge 2 ($M = 3.28$, $SD = 1.94$) was significantly lower than Judge 1 ($M = 5.69$, $SD = 1.58$) and Judge 3 ($M = 5.34$, $SD = 1.32$, $p < 0.001$). There was no significant difference between Judge 1 and Judge 3 ($p = 0.100$). There were no other significant effects.

Block 2 (Stimulation)

As can be seen in the bottom panel of **Figure 2**, there was no significant main effect of Group in Block 2, $F(1,14) = 2.28$, $p = 0.153$, $\eta_p^2 = 0.14$. Thus, the type of tDCS administered to participants did not differentially affect the technical fluency of their performances. However, there was a significant main effect of Judge, $F(2,28) = 38.59$, $p < 0.001$, $\eta_p^2 = 0.73$, which followed the same trend in results as the aforementioned main effects of Judge, and a significant Judge \times Take interaction, $F(8,112) = 2.39$, $p = 0.021$, $\eta_p^2 = 0.15$. Examination of mean ratings suggests that this interaction may be driven by differences in ratings between judges at take 5. Judges 1 and 3 assigned similar overall ratings of technical fluency (and creativity) across takes 1–4. However, mean ratings of technical fluency by Judge 3 dropped below that of Judge 1 in take 5.

Correlation Between Ratings of Creativity and Technical Fluency

All trials were analyzed irrespective of tDCS stimulation to investigate the relationship between creativity and technical fluency. For Judge 1, there was a significant positive correlation between technical fluency and creativity, $r = 0.72$, 95% BCa CI [0.621, 0.794], $p = 0.01$. For Judge 2, there was also a significant positive correlation between technical fluency and creativity, $r = 0.74$, 95% BCa CI [0.633, 0.850], $p = 0.01$. Finally, for Judge 3 there was a significant positive correlation between technical fluency and creativity, $r = 0.67$, 95% BCa CI [0.578, 0.741], $p = 0.01$. All reported correlations are considered to reflect a large effect size (Babchishin and Helmus, 2016).

Melodic Performance Features

Total Number of Notes Used

The analysis revealed a significant difference between the excitatory tDCS group ($M = 30.28$, $SD = 5.26$) and the inhibitory tDCS group ($M = 21.23$, $SD = 4.31$), $t(14) = 3.763$, $p = 0.002$. This result shows that with tDCS stimulation to the M1, the mean total number of notes used in the improvisation stage was significantly greater for those who experienced excitatory tDCS when compared to inhibitory tDCS.

Number of Different Notes Used

The analysis revealed a significant difference between the excitatory tDCS group ($M = 9.00$, $SD = 0.76$) and the inhibitory group ($M = 7.83$, $SD = 1.05$), $t(14) = 2.569$, $p = 0.022$. This result shows that when tDCS stimulation is applied to the M1, the mean number of different notes used in the improvisation stage was significantly greater for those who experienced excitatory tDCS when compared to inhibitory tDCS.

Pitch Range

The analysis also revealed a significant difference between the excitatory tDCS group ($M = 19.93$, $SD = 5.53$) and the inhibitory group ($M = 14.20$, $SD = 1.54$), $t(8) = 2.288$, $p = 0.022$. This result shows that when tDCS stimulation is applied to the M1, the mean pitch range used in the improvisation stage was significantly larger for those who experienced excitatory tDCS when compared to inhibitory tDCS.

Association Between Creativity, Technical Fluency, and Melodic Performance Features

Ratings of Creativity

For the *excitatory* tDCS group in Block 2, the three melodic performance features (total number of notes, number of different notes, and pitch range) were significant predictors of creativity, $F(3,4) = 8.381$, $p = 0.034$, Adjusted $R^2 = 0.760$. For the *inhibitory* tDCS group in Block 2, the three melodic performance features were not significant predictors of creativity, $F(3,4) = 2.632$, $p = 0.186$, adjusted $R^2 = 0.412$.

Ratings of Technical Fluency

For the *excitatory* tDCS group in Block 2, the three melodic performance features were not significant predictors of technical fluency $F(3,4) = 3.149$, $p = 0.148$, adjusted R^2 of 0.479. For the *inhibitory* tDCS group in Block 2, the three melodic performance features were also not significant predictors of technical fluency, $F(3,4) = 0.1479$, $p = 0.906$, adjusted R^2 of -0.543 .

Sight-Reading Performance Accuracy

Timing Accuracy

In the sight-reading stage of performances in Block 1 (section A in each score), the analysis revealed no significant difference in timing accuracy between the excitatory tDCS group ($M = 27.81$ ms, $SD = 49.22$) and the inhibitory tDCS group ($M = 87.74$ ms, $SD = 140.95$), $t(14) = -1.135$, $p = 0.287$. In the sight-reading stage of performances in Block 2, the analysis also revealed no significant difference between the excitatory tDCS group ($M = 12.70$ ms, $SD = 49.43$) and the inhibitory tDCS group ($M = 93.15$ ms, $SD = 139.93$), $t(14) = -1.533$, $p = 0.161$. These results show that tDCS stimulation did not significantly affect timing accuracy in the sight-reading stage.

Pitch-Note Accuracy

In the sight-reading stage of performances in Block 1, the analysis revealed no significant difference in pitch-note accuracy between the excitatory tDCS group ($M = 0.55$, $SD = 0.72$) and the

inhibitory tDCS group ($M = 1.24$, $SD = 1.68$), $t(14) = -1.062$, $p = 0.314$. In the sight-reading stage of performances in Block 2, the analysis also revealed no significant difference between the excitatory tDCS group ($M = 0.42$, $SD = 0.36$) and the inhibitory tDCS group ($M = 1.18$, $SD = 2.02$), $t(14) = -0.950$, $p = 0.385$. This result shows that the type of tDCS stimulation did not affect pitch-note accuracy in the sight-reading section.

DISCUSSION

The aim of this investigation was to determine whether the M1 plays a role in creative and technically fluent musical improvisations. Expert jazz pianists received either excitatory or inhibitory tDCS over the left hemispheric M1 while completing right-handed jazz piano performances that comprised a sight-reading stage and an improvisation stage. Performances were adjudicated by expert musicians who judged creativity and technical fluency. We hypothesized that improvisations performed by participants who received excitatory tDCS would be more creative and technically fluent than improvisations performed by those who received inhibitory tDCS. This hypothesis was supported for ratings of creativity: improvisations by participants who received excitatory tDCS were rated as significantly more creative than those who received inhibitory tDCS. Interestingly, we observed no significant differences between excitatory and inhibitory tDCS for ratings of technical fluency. Follow-up analyses revealed that melodic performance features such as the total number of notes played, number of different notes played, and pitch range were significant predictors of creative performances for those in the excitatory tDCS group. The type of tDCS did not differentially affect sight-reading accuracy as measured by timing and pitch-note accuracy in the sight-reading stage.

One possible explanation for the results is that the M1 mediates the potential for a creative motor action associated with a pre-planned creative idea. Specifically, the foundations of a creative idea may form in brain areas associated with higher-order creative processes such as attention, planning, working memory, cognitive flexibility, and imagination, and then flow in part via the M1 to be realized as a creative motor action (Dietrich, 2004; Lucchiari et al., 2018). Research suggests that networks in the PFC are responsible for higher-order cognitive functions (specifically the ECN and DMN) associated with creative processes in all domains including music (Bengtsson et al., 2007; Gonen-Yaacovi et al., 2013; Boccia et al., 2015). It is also clear that the PFC and the M1 are functionally linked (e.g., Hasan et al., 2013). Thus, exciting the M1 may have increased the potential for converting a preplanned creative idea into a creative motor action. In the context of piano improvisations in the present study, stimulating the M1 may have facilitated the flow of creative “content” (notes) from the pre-planned creative idea into motor output (piano performance). As a result, improvisations during excitatory tDCS were more creative than inhibitory tDCS because

they reflected an increased output of creative performance features.

The data reported here provide some support for this interpretation. Participants who received excitatory tDCS performed improvisations with a significantly greater number of notes and greater number of different notes, as well as a wider pitch range than participants who received inhibitory tDCS. Furthermore, results from multiple regression analyses showed that these three performance features were significant predictors of creativity for the excitatory tDCS group, explaining 76% of the variance. This was not the case for the inhibitory tDCS group.

Interestingly, a parallel effect of tDCS on technical fluency was not observed, even though ratings of creativity and technical fluency were positively and significantly correlated. Brain stimulation may have facilitated the flow of creative ideas from higher levels of processing through to motor planning and motor actions, releasing a low-level neural “gateway” for high level creative ideas. Technical fluency, in contrast, may operate independently of that process of disinhibition and may instead rely on over-learned, automated processes of action control that are comparatively fixed through training and less susceptible to transient changes from stimulation. Alternatively, it may be that task demands for technical fluency were such that there was less opportunity for performers to differ in technical fluency than in creativity. For performers to display fluency, they needed to play syntactically plausible pitches on plausible metric subdivisions. Although timing and pitch errors occurred, the task demands might have afforded less opportunity for variability in technical fluency.

Finally, judges evaluated the inherently creative musical task of improvisation. As a result, they may have focused more attention and greater cognitive resources on their judgments of creativity and fewer resources on the adjudication of technical fluency, thus resulting in less reliability in judgments of fluency. Future research could alleviate this possibility by recruiting two groups of judges: one that adjudicates the creative element of each performance, and the other that adjudicates technical fluency. Indeed, the difference in results between creativity and technical fluency will need to be replicated in future studies with greater statistical power by including more expert performers and adjudicators. Nevertheless, there was a strong positive correlation between creativity and technical fluency ratings from all three adjudicators irrespective of the type of tDCS stimulation participants received. This result suggests that creativity and technical fluency are related phenomena in adjudication of musical improvisation, even though both are differentially affected by stimulation of the M1.

To investigate the M1 with greater localization specificity, future studies could also use a rTMS paradigm. rTMS is a non-invasive brain stimulation technique that facilitates or inhibits neural activity by modulating MEPs in the M1 (Romero et al., 2002; Peinemann et al., 2004). This is accomplished by varying the frequency of pulses (pulses per second), number of total pulses, and the inter-train interval (period where TMS is not administered). rTMS has the potential to modulate neural activity for a prolonged period (20–60 min) and with greater localization specificity than tDCS (Huang et al., 2005; Rotenberg et al., 2014).

Replicating the present study with rTMS will allow more causal inferences to be made regarding the role of the M1 region in creative and technically fluent piano improvisations.

CONCLUSION

To conclude, our findings illustrate an important role for the M1 in musical creativity. Indeed, the M1 may not only act as a gateway for translating creative cognition into action, but likely mediates the potential for maximizing such creative output. Although more research is needed to link such an association to applied contexts such as performance pedagogy, the results imply that programs emphasizing movement and rhythm have the potential to benefit creative musicianship. Technical fluency, on the other hand, may operate independently of this process and instead rely on learned automated actions that are comparatively fixed through music training. Future research is needed to evaluate these proposals with a greater number of expert musician participants and adjudicators. Nevertheless, the current findings suggest that the M1 should receive greater consideration in the already complex neural network that mediates creativity, especially in the context of movement-based expertise.

REFERENCES

- Adhikari, B. M., Norgaard, M., Quinn, K. M., Ampudia, J., Squire, J., and Dhamala, M. (2016). The brain network underpinning novel melody creation. *Brain Connect.* 6, 772–785. doi: 10.1089/brain.2016.0453
- Anic, A., Thompson, W.F., and Olsen, K.N. (2017). “Stimulation of the primary motor cortex enhances creativity and technical fluency of piano improvisations,” in *Proceedings of the 10th International Conference of Students of Systematic Musicology (SysMus17)*, London.
- Babchishin, K. M., and Helmus, L. (2016). The influence of base rates on correlations: an evaluation of proposed alternative effect sizes with real-world data. *Behav. Res. Methods*, 48, 1021–1031. doi: 10.3758/s13428-015-0627-7
- Bashwiler, D. M., Wertz, C. J., Flores, R. A., and Jung, R. E. (2016). Musical creativity “revealed” in brain structure: interplay between motor, default mode, and limbic networks. *Sci. Rep.* 6:20482. doi: 10.1038/srep20482
- Batey, M., and Furnham, A. (2006). Creativity, intelligence, and personality: a critical review of the scattered literature. *Genet. Soc. Gen. Psychol. Monogr.* 132, 355–429. doi: 10.3200/MONO.132.4.355-430
- Bendetowicz, D., Urbanski, M., Garcin, B., Foulon, C., Levy, R., Brechemier, M., et al. (2017). Two critical brain networks for generation and combination of remote associations. *Brain* 141, 217–233. doi: 10.1093/brain/awx294
- Benedek, M., Jauk, E., Sommer, S., Arendasy, M., and Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: the common and differential involvement of executive functions in intelligence and creativity. *Intelligence* 46, 73–83. doi: 10.1016/j.intell.2014.05.007
- Bengtsson, S. L., Csikszentmihályi, M., and Ullén, F. (2007). Cortical regions involved in the generation of musical structures during improvisation in pianists. *J. Cogn. Neurosci.* 19, 830–842. doi: 10.1162/jocn.2007.19.5.830
- Berkowitz, A. L., and Ansari, D. (2008). Generation of novel motor sequences: the neural correlates of musical improvisation. *Neuroimage* 41, 535–543. doi: 10.1016/j.neuroimage.2008.02.028
- Biasutti, M. (2015). Pedagogical applications of cognitive research on musical improvisation. *Front. Psychol.* 6:614. doi: 10.3389/fpsyg.2015.00614
- Biasutti, M. (2017). Teaching improvisation through processes. Applications in music education and implications in general education. *Front. Psychol.* 8:911. doi: 10.3389/fpsyg.2017.00911
- Biasutti, M., and Frezza, L. (2009). Dimensions of musical improvisation. *Creat. Res. J.* 21, 232–242. doi: 10.1080/10400410902861240
- Bikson, M., Datta, A., and Elwassif, M. (2009). Establishing safety limits for transcranial direct current stimulation. *Clin. Neurophysiol.* 120, 1033–1034. doi: 10.1016/j.clinph.2009.03.018
- Boccia, M., Piccardi, L., Palermo, L., Nori, R., and Palmiero, M. (2015). Where do bright ideas occur in our brain? Meta-analytic evidence from neuroimaging studies of domain-specific creativity. *Front. Psychol.* 6:1195. doi: 10.3389/fpsyg.2015.01195
- Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms*, 2nd Edn. London: Routledge. doi: 10.4324/9780203508527
- Brinkman, J., and Kuypers, H. G. (1973). Cerebral control of contralateral and ipsilateral arm, hand, and finger movements in the split-brain rhesus monkey. *Brain* 96, 653–674. doi: 10.1093/brain/96.4.653
- Chouinard, P. A., and Paus, T. (2006). The primary motor and premotor areas of the human cerebral cortex. *Neuroscientist* 12, 143–152. doi: 10.1177/1073858405284255
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardised assessment instruments in psychology. *Psychol. Assess.* 6, 284–290. doi: 10.1037/1040-3590.6.4.284
- Colombo, B., Bartsaghi, N., Simonelli, L., and Antonietti, A. (2015). The combined effects of neurostimulation and priming on creative thinking. A preliminary tDCS study on dorsolateral prefrontal cortex. *Front. Hum. Neurosci.* 9:403. doi: 10.3389/fnhum.2015.00403
- de Manzano, O., and Ullén, F. (2012). Activation and connectivity patterns of the presupplementary and dorsal premotor areas during free improvisation of melodies and rhythms. *Neuroimage* 63, 72–280. doi: 10.1016/j.neuroimage.2012.06.024
- Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750
- Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychon. Bull. Rev.* 11, 1011–1026. doi: 10.3758/BF03196731
- Fink, A., and Benedek, M. (2014). EEG alpha power and creative ideation. *Neurosci. Biobehav. Rev.* 44, 111–123. doi: 10.1016/j.neubiorev.2012.12.002
- Gonen-Yacovi, G., de Souza, L. C., Levy, R., Urbanski, M., Josse, G., and Volle, E. (2013). Rostral and caudal prefrontal contribution to creativity: a meta-analysis of functional imaging data. *Front. Hum. Neurosci.* 7:465. doi: 10.3389/fnhum.2013.00465
- Hasan, A., Galea, J. M., Casula, E. P., Falkai, P., Bestmann, S., and Rothwell, J. C. (2013). Muscle and timing-specific functional connectivity between the

AUTHOR CONTRIBUTIONS

AA coordinated testing and data collection. AA and KO were responsible for data analysis and all authors contributed to data interpretation. AA wrote the first draft of the manuscript and all authors contributed to further revisions. All authors approved the final version of the manuscript, and contributed to the design and development of the study.

FUNDING

This research was funded by the Australian Research Council Centre of Excellence in Cognition and its Disorders (CE110001021), and by a Discovery grant from the Australian Research Council awarded to WFT (DP160101470).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01758/full#supplementary-material>

- dorsolateral prefrontal cortex and the primary motor cortex. *J. Cogn. Neurosci.* 25, 558–570.
- Hoshi, E., and Tanji, J. (2007). Distinctions between dorsal and ventral premotor areas: anatomical connectivity and functional properties. *Curr. Opin. Neurobiol.* 17, 234–242. doi: 10.1016/j.conb.2007.02.003
- Huang, Y., Edwards, M. J., Rounis, E., Bhatia, K. P., and Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron* 45, 201–206. doi: 10.1016/j.neuron.2004.12.033
- Jauk, E., Benedek, M., and Neubauer, A. C. (2014). The road to creative achievement: a latent variable model of ability and personality predictors. *Eur. J. Personal.* 28, 95–105. doi: 10.1002/per.1941
- Jauk, E., Neubauer, A. C., Dunst, B., Fink, A., and Benedek, M. (2015). Gray matter correlates of creative potential: a latent variable voxel-based morphometry study. *NeuroImage* 111, 312–320. doi: 10.1016/j.neuroimage.2015.02.002
- Karok, S., and Witney, A. G. (2013). Enhanced motor learning following task-concurrent dual transcranial direct current stimulation. *PLoS One* 8:e85693. doi: 10.1371/journal.pone.0085693
- Kenett, Y. N., Medaglia, J. D., Beaty, R. E., Chen, Q., Betzel, R. F., Thompson-Schill, S. L., et al. (2018). Driving the brain towards creativity and intelligence: a network control theory analysis. *Neuropsychologia* doi: 10.1016/j.neuropsychologia.2018.01.001 [Epub ahead of print].
- Kuhn, S., Ritter, S. M., Muller, B. C. N., Baaren, R. B. V., Brass, M., and Dijksterhuis, A. (2013). The importance of the default mode network in creativity: a structural MRI study. *J. Creat. Behav.* 48, 152–163. doi: 10.1002/jocb.45
- Limb, C. J., and Braun, A. R. (2008). Neural substrates of spontaneous musical performance: an fMRI study of Jazz improvisation. *PLoS One* 3:e1679. doi: 10.1371/journal.pone.0001679
- Lucchiari, C., Sala, P. M., and Vanutelli, M. E. (2018). Promoting creativity through transcranial direct current stimulation (tDCS): a critical review. *Front. Behav. Neurosci.* 12:167. doi: 10.3389/fnbeh.2018.00167
- Nitsche, M. A., Schauenburg, A., Lang, N., Liebetanz, D., Exner, C., Paulus, W., et al. (2003). Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. *J. Cogn. Neurosci.* 15, 619–626. doi: 10.1162/089892903321662994
- Peinemann, A., Reimer, B., Loer, C., Quartarone, A., Munchau, A., Conrad, B., et al. (2004). Long-lasting increase in corticospinal excitability after 1800 pulses of subthreshold 5 Hz repetitive TMS to the primary motor cortex. *Clin. Neurophysiol.* 115, 1519–1526. doi: 10.1016/j.clinph.2004.02.005
- Pinho, A. L., de Manzano, O., Fransson, P., Eriksson, H., and Ullén, F. (2014). Connecting to create: expertise in musical improvisation is associated with increased functional connectivity between premotor and prefrontal areas. *J. Neurosci.* 34, 6156–6163. doi: 10.1523/JNEUROSCI.4769-13.2014
- Romero, J. R., Ansel, D., Sparing, R., Gangitano, M., and Pascual-Leone, A. (2002). Subthreshold low frequency repetitive transcranial magnetic stimulation selectively decreases facilitation in the motor cortex. *Clin. Neurophysiol.* 113, 101–107. doi: 10.1016/S1388-2457(01)00693-9
- Rotenberg, A., Horvath, J. C., and Pascual-Leone, A. (2014). *Transcranial Magnetic Stimulation*. New York, NY: Springer. doi: 10.1007/978-1-4939-0879-0
- Schwab, D., Benedek, M., Papousek, I., Weiss, E. M., and Fink, A. (2014). The time-course of EEG alpha power changes in creative ideation. *Front. Hum. Neurosci.* 8:310. doi: 10.3389/fnhum.2014.00310
- Simonton, D. K. (2010). Creative thought as blind-variation and selective-retention: combinatorial models of exceptional creativity. *Phys. Life Rev.* 7, 156–179. doi: 10.1016/j.plrev.2010.02.002
- Sosnik, R., Flash, T., Sterkin, A., Hauptmann, B., and Karni, A. (2014). The activity in the contralateral primary motor cortex, dorsal premotor and supplementary motor area is modulated by performance gains. *Front. Hum. Neurosci.* 8:201. doi: 10.3389/fnhum.2014.00201
- Sowden, P. T., Pringle, A., and Gabora, L. (2015). The shifting sands of creative thinking: connections to dual-process theory. *Think. Reason.* 21, 40–60. doi: 10.1080/13546783.2014.885464
- Tan, C.-S., Lau, X.-S., Kung, Y.-T., and Kailsan, R. A. L. (2016). Openness to experience enhances creativity: the mediating role of intrinsic motivation and the creative process engagement. *J. Creat. Behav.* doi: 10.1002/jocb.170
- van den Berg, F. E., Swinnen, S. P., and Wenderoth, N. (2011). Involvement of the primary motor cortex in controlling movements executed with the ipsilateral hand differs between left- and right-handers. *J. Cogn. Neurosci.* 23, 3456–3469. doi: 10.1162/jocn_a_00018
- Vines, B. W., Cerruti, C., and Schlaug, G. (2008a). Dual-hemisphere tDCS facilitates greater improvements for healthy subjects' non-dominant hand compared to uni-hemispheric stimulation. *BMC Neurosci.* 9, 1–7. doi: 10.1186/1471-2202-9-103
- Vines, B.W., Nair, D., and Schlaug, G. (2008b). Modulating activity in the motor cortex affects performance for the two hands differently depending upon which hemisphere is stimulated. *Eur. J. Neurosci.* 28, 1667–1673. doi: 10.1111/j.1460-9568.2008.06459.x
- Waters-Metenier, S., Husain, M., Wiestler, T., and Diedrichsen, J. (2014). Bihemispheric transcranial direct current stimulation enhances effector-independent representations of motor synergy and sequence learning. *J. Neurosci.* 34, 1037–1050. doi: 10.1523/JNEUROSCI.2282-13.2014
- Zhu, W., Chen, Q., Xia, L., Beaty, R. E., Yang, W., Tian, F., et al. (2017). Common and distinct brain networks underlying verbal and visual creativity. *Hum. Brain Mapp.* 38, 2094–2111. doi: 10.1002/hbm.23507

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 Anic, Olsen and Thompson. 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.



Collaborative Musical Creativity: How Ensembles Coordinate Spontaneity

Laura Bishop*

Austrian Research Institute for Artificial Intelligence (OFAI), Vienna, Austria

OPEN ACCESS

Edited by:

William Forde Thompson,
Macquarie University, Australia

Reviewed by:

Matthew Rodger,
Queen's University Belfast,
United Kingdom
Daniel Bangert,
Georg-August-Universität Göttingen,
Germany

*Correspondence:

Laura Bishop
laura.bishop@ofai.at

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 19 April 2018

Accepted: 04 July 2018

Published: 24 July 2018

Citation:

Bishop L (2018) Collaborative Musical
Creativity: How Ensembles Coordinate
Spontaneity. *Front. Psychol.* 9:1285.
doi: 10.3389/fpsyg.2018.01285

Music performance is inherently social. Most music is performed in groups, and even soloists are subject to influence from a (real or imagined) audience. It is also inherently creative. Performers are called upon to interpret notated music, improvise new musical material, adapt to unexpected playing conditions, and accommodate technical errors. The focus of this paper is how creativity is distributed across members of a music ensemble as they perform these tasks. Some aspects of ensemble performance have been investigated extensively in recent years as part of the broader literature on joint action (e.g., the processes underlying sensorimotor synchronization). Much of this research has been done under highly controlled conditions, using tasks that generate reliable results, but capture only a small part of ensemble performance as it occurs naturalistically. Still missing from this literature is an explanation of how ensemble musicians perform in conditions that require creative interpretation, improvisation, and/or adaptation: how do they coordinate the production of something new? Current theories of creativity endorse the idea that dynamic interaction between individuals, their actions, and their social and material environments underlies creative performance. This framework is much in line with the embodied music cognition paradigm and the dynamical systems perspective on ensemble coordination. This review begins by situating the concept of collaborative musical creativity in the context of embodiment. Progress that has been made toward identifying the mechanisms that underlie collaborative creativity in music performance is then assessed. The focus is on the possible role of musical imagination in facilitating performer flexibility, and on the forms of communication that are likely to support the coordination of creative musical output. Next, emergence and group flow—constructs that seem to characterize ensemble performance at its peak—are considered, and some of the conditions that may encourage periods of emergence or flow are identified. Finally, it is argued that further research is needed to (1) demystify the constructs of emergence and group flow, clarifying their effects on performer experience and listener response, (2) determine how constrained musical imagination is by perceptual experience and understand people's capacity to depart from familiar frameworks and imagine new sounds and sound structures, and (3) assess the technological developments that are supposed to facilitate or enhance musical creativity, and determine what effect they have on the processes underlying creative collaboration.

Keywords: creativity, ensemble performance, embodiment, emergence, mental imagery, communication

1. INTRODUCTION

Music performance is a social task. Most of the world's music is performed in groups, and even soloists are subject to influence from (real or imagined) audiences. Music perception is social too: audiences recognize social relationships and communicative behavior between members of a performing ensemble (Moran et al., 2015; Aucouturier and Canonne, 2017), and they infer human agency when hearing music—even without visual confirmation of a performer (Launay, 2015; Olsen and Dean, 2016), making sounded performances a means of interpersonal communication.

Music performance is also creative. In some musical traditions, new musical material is created via improvisation, while in others, sounded performances are created from loosely-defined visual notation. Performers across traditions adjust their playing to accommodate new performance environments as well as errors that result from imperfect technique (e.g., missed notes) or attentional lapses (e.g., missed repeats; Glowinski et al., 2016).

This paper addresses the question of how creativity is distributed across members of a music ensemble during performance. The focus is on processing that occurs online (i.e., during performance), though it is acknowledged that many offline musical tasks are creative as well (e.g., composing, structuring practice sessions, preparing an interpretation of a piece across successive rehearsals, evaluating other's performances, etc.). The real-time nature of music performance differentiates it from many other everyday tasks that require creative collaboration, such as brainstorming solutions to a problem with colleagues or jointly writing a report. There are also constraints on ensemble musicians' communication that do not exist in the context of many other tasks—for example, verbal discussion goes against performance conventions in many musical traditions, and gesturing may be hampered by the physical presence of instruments. For these reasons, music performance provides a particularly useful context for investigating how creative ideas emerge in real-time from the interactions between members of a group.

The structure of this paper is as follows. First, definitions of creativity and collaborative creativity in the context of music performance are outlined. I then present some theoretical perspectives on creativity and musical interaction that conceptualize creativity as distributed between interacting individuals and their social and material environments, and I argue that a refined framework drawing on these ideas is needed to guide ongoing research efforts. Next, some of the key mechanisms that are thought to support collaborative creativity in music ensembles are outlined, including flexible and persistent idea generation, musical imagination, communication, and empathetic attunement. The concepts of flow and emergence, which are central to discussions of collaborative creativity, are then considered, and some of the conditions that encourage group flow are discussed. Finally, some topics that are important to address in future research are identified.

2. CREATIVITY

Creativity describes the component of human cognition that enables generation of output (an object, idea, performance) that is both novel and significant (Dietrich, 2004). In research contexts, creative output is typically evaluated on the basis of its originality and appropriateness. In artistic domains such as music performance, negotiating a balance between originality and appropriateness means maintaining flexibility within a given set of stylistic constraints. It is important not to confuse creativity with either originality, defined as the degree of novelty of a creative output relative to a given sample of related outputs, or value, the quality assigned to a creative output by a receiving audience (Williamon et al., 2006). Creativity is a component of cognition, while originality and value are evaluations made by others in the context of their own cultural experiences.

Recent theoretical frameworks include this evaluation process as a critical component of the overarching creativity construct. Fischer et al. (2005) describe four components of creativity: (1) originality, (2) expression, the externalization of the creative idea, (3) social evaluation, the process by which others consider the creative output and judge its value and (4) social appreciation, the process of encouraging or discouraging further creative efforts. As discussed in the next section, current theories endorse the idea that creativity does not function in a vacuum or within the confines of an individual mind, but rather, is shaped continually, in real-time, by past, present and anticipated interactions with the external world. This view is in contrast to earlier work on creativity, which focused on the internal cognitive processes of individuals, and treated these processes as separable from external influences.

2.1. Collaborative Creativity: Forms and Levels

Collaborative creativity refers to the distribution of creativity across members of a group as they collaborate to solve a shared problem. It is in contrast to a division of labor, where each group member is assigned a part of the task and the collective outcome is equal to the sum of individual contributions. Collaborative creativity involves more complex interaction between group members and can yield an outcome that is greater than the sum of individual contributions. This greater collective outcome arises because the difference in task conditions prompted by group members working together, instead of individually, allows for the occurrence of ideas that cannot be attributed to any one person—a phenomenon referred to as emergence (see section 4; Fischer et al., 2005).

Creative collaboration between people can be (1) serial, if an individual creates something in isolation, then presents their creation to others who can build on it, (2) parallel, if group members create things separately, then bring them together to combine them into something new, or (3) simultaneous, if group members create something together, at the same time (Fischer et al., 2005). Simultaneous collaboration is of primary interest to the current discussion, though serial and parallel collaboration can be observed in the context of music performance as

well. In particular, rehearsal of ensemble music often involves parallel collaboration, as ensemble members may do preliminary preparation of their own parts of a piece before playing it together as a group.

Seddon and Biasutti (2009) observed a professional string quartet and a student jazz sextet in rehearsal and noted three levels of interaction between ensemble members: instruction, which occurred when one group member communicated to another what to do; cooperation, which occurred when group members communicated to ensure that output was cohesive; and collaboration, which occurred when group members took creative risks, leading to the emergence of something new. Of particular interest to the current discussion is how performers move from the level of cooperation to the level of collaboration: what conditions prompt or prevent a higher level of interaction? How do interactions aimed at cooperation and interactions aimed at collaboration differ? These are among the questions explored in the later sections of this paper.

2.2. Creativity as Embodied and Distributed

Contemporary theories of creativity incorporate ideas from distributed cognition and dynamical systems theory, emphasizing the role of the social and material environments in which creative processes are carried out (Schiavio and Høffding, 2015; Linson and Clarke, 2017). This is in contrast to early studies of creativity, which focused on individuals' internal cognitive processes. This section of the paper discusses three theoretical approaches—"5 A's" creativity framework, the extended mind thesis, and the embodied music cognition paradigm—and how these approaches might be applied to an explanation of collaborative creativity. The aim is to define a theoretical conceptualization of collaborative creativity as involving a network of interactive, embodied, socially-situated, and externalizable processes.

Proposed by Glăveanu (2013), the "5 A's" creativity framework defines five components: (1) Actor(s) who engage in (2) Actions (i.e., creative thinking externalized) that bring about an (3) Artifact (creative output) in the context of (4) an Audience (the social environment) and (5) Affordances (the material environment). This framework is a reworking of the earlier "4 P's" framework (comprising Person, Process, Product, and Press; Rhodes, 1961), re-designed with the aim of emphasizing the interdependence of the five components (whereas the "4 P's" were conceptualized as separable, and often studied independently).

According to this theory, interaction between the actor(s) and audience is critical, as it is the audience, who, presented with output produced by the actors, determines it to be creative (Csikszentmihalyi, 1999). That is, the creative quality of the actor's output is not realized until others have recognized it as such. This emphasis on evaluation places the audience in almost as important a role for achieving creative output as the actors (Glăveanu, 2013). It should be noted that an audience may be real and present, providing live feedback to the actors (e.g., as when people attend a concert) or imagined/anticipated, in which case the actors may become the audience vicariously by assuming an audience perspective (e.g., when students consider how a performance might be received by the judges at their

upcoming exam). During collaborative creativity, individual collaborators can be said to fulfill the roles of actor and audience simultaneously, as they are continuously judging the creative quality of each other's output while also producing creative output themselves. The extent to which collaborators judge each other's output to be creative can either encourage or discourage their own continued participation in the task and influence their willingness to take creative risks.

Also emphasizing the interdependence of actors, actions, and the environment, the extended mind thesis proposes that some cognitive processes are partially composed of actions made within the "we-space," a dynamically structured physical space surrounding a person in which interaction with others is possible (Krueger, 2010). Multiple levels of we-space exist: personal space is taken up by the body; peripersonal space immediately surrounds the body and is accessible via auditory, visual, and tactile perception; and extrapersonal space is beyond the person's immediate reach and accessible only via auditory and visual perception. Gestures are an important means of navigating interactions in the we-space; they are an externalization of the gesture giver's cognitive-affective processes and involved in driving those processes, while simultaneously facilitating the task of the gesture receiver by narrowing the range of responses they have to choose from.

Interpersonal coordination in the we-space is thought to be partially a process of co-regulation, or continuous adaptation to one another's expressive behavior (e.g., automatic mimicking of facial expressions during conversation). Co-regulation distinguishes "focused interaction" (i.e., collaboration with a shared focus of attention) from "unfocused interaction" (i.e., co-presence without shared attention). For example, members of a music ensemble may adapt to each other's behavior during collaborative performance, which constitutes focused interaction, but not while practicing individually in the same rehearsal space, which constitutes unfocused interaction. The idea of co-regulation is much in line with the idea of coordination emerging dynamically from pre-reflective interactions at the level of body movement, which has been discussed in the music cognition literature (e.g., Maes, 2016). This idea is explored in greater depth in Section 3.3.2.

Like the "5 A's" and extended mind frameworks, the embodied music cognition (EMC) paradigm conceptualizes cognition as distributed between a person's brain, body, and environment. The body is thought to mediate interactions between subjective experiences and the external world, during both music performance (as meaning is transformed into sound) and music perception (as meaning is constructed from sounded stimuli; Leman and Maes, 2014; Maes et al., 2014; Moran, 2014). Perceptual-motor coupling has been proposed as a possible mechanism for body-mediated meaning formation. Perceptual-motor coupling occurs when perceptual events and the motor commands needed to produce them share overlapping neural representations. This shared coding creates an association that can be activated bidirectionally—actions can prompt expectations for specific perceptual effects, and perceived or anticipated effects can prime related actions (Prinz, 1990; Jeannerod, 2003).

The EMC paradigm posits that during group interaction, perceptual-motor coupling functions at multiple levels simultaneously (Leman, 2012; van der Wel et al., 2016; MacRitchie et al., 2017). At a lower level, motor activity and sensory input are in continuous interaction, enabling automatic regulation of performance technique and entrainment between ensemble members. At a higher level, performers draw on a repertoire of learned gestures to control their own playing and achieve more deliberate coordination with co-performers. In Section 3.3, these coordination modes are discussed in greater depth.

The theoretical perspectives outlined here suggest that a music ensemble should be thought of as a system in which all components, including individual performers, their instruments, the audience, the performance space, are interdependent and dynamically interacting. Empirical study of the processes underlying ensemble performance increasingly reflects this perspective. Still, most research thus far has focused on the processes involved in achieving and maintaining interpersonal synchronization in situations with relatively high temporal predictability (Keller and Appel, 2010; Loehr and Palmer, 2011; Ragert et al., 2013; Repp and Su, 2013; Zamm et al., 2014). To a large extent, designing controlled experimental conditions has meant reducing the demands on performer's creativity as much as possible. Whether the mechanisms that underlie performance on such controlled tasks generalize to performance under normal conditions – when the demands on creativity are high – is unclear. As discussed in the next section, it seems likely that additional mechanisms must be activated for ensembles to play coherently under conditions that demand creativity.

3. MECHANISMS FOR MUSICAL CREATIVITY

Creative accomplishments are seen in a wide range of domains, from everyday problem-solving to interpersonal to artistic to scientific. The question of how domain-general and domain-specific processes combine to support performance on creative tasks is still an open question. Exceptional creative ability within a given domain usually depends on a person having extensive domain-specific knowledge and, in some cases, specialized motor skills (Ericsson, 1998). On the other hand, neuroimaging studies have shown that while the patterns of brain activation seen in people engaged in creative behavior are largely task-specific, certain regions (specifically, the lateral prefrontal cortex, inferior parietal cortex, and lateral posterior temporal cortex) are activated consistently regardless of the task (Gonen-Yaacovi et al., 2013). These regions may support a general network of creative abilities.

Barbot and Tinio (2015) argued that while evidence of a unitary, domain-general creativity capacity (i.e., similar to the “g” factor of intelligence) is limited, there seems to be a set of general creative resources that combine in different ways to support performance on a range of tasks. These resources include different processing strategies, such as associative thinking, selective combination, perseverance, and elaboration, as well as

general intelligence, motivation (An et al., 2016) and mindset (Bittner and Heidemeier, 2013). Creative performance on a given task is facilitated when an optimal combination of resources is drawn upon. There is likely to be an ideal “fit” between the resources that are activated and the demands of the situation. For instance, De Dreu et al. (2011) found that trait behavioral activation—the tendency to carry out goal-directed behavior and respond with positive feelings to signs of an impending reward—potentiates creativity on tasks that afford flexible and global processing, but impedes creativity on tasks that afford local processing.

In performing creatively, ensemble musicians face two primary challenges: generating original (but stylistically appropriate) ideas and maintaining coordination while translating these ideas into musical output. Meeting these challenges draws on a large network of cognitive processes, likely a combination of general and task-specific. Three processes proposed to be central to creative collaboration are highlighted in the current paper: potential mechanisms for generating ideas (spreading activation), elaborating and evaluating ideas (musical imagination), and coordinating the implementation of ideas (communication). These processes are discussed individually in the following three sections.

Much of the research referenced in these sections—especially in relation to idea generation and musical imagery—adopts an individualistic perspective, focusing on cognitive processes within individuals. This is in contrast to the theoretical perspective endorsed in this paper, that collaborative creativity is embodied and distributed. Currently, little of the published research on creativity focuses on collaboration, and a theory of collaborative creativity has not yet been proposed. Therefore, this paper aims to identify theoretical concepts and empirical observations from individual-focused research that may be applicable to collaborative contexts, and highlight gaps in what these theories and observations are able to explain.

3.1. Flexible and Persistent Modes of Idea Generation

Theories of creativity commonly distinguish between two contrasting processing modes – cognitive persistence and cognitive flexibility (Dietrich, 2004; Nijstad et al., 2010). Cognitive persistence involves sustained attention and controlled, incremental, and structured exploration of ideas. Cognitive flexibility, in contrast, involves divergent thinking, a global focus, the use of broad cognitive categories, and frequent switching between categories.

Dietrich (2004) defines four subtypes of creative processing by crossing flexibility and persistence (which he refers to as spontaneous and deliberate modes of thinking) with cognitive and emotional knowledge domains. Most creative tasks are said to engage a combination of these modes. For example, Dietrich suggests that creativity in the arts derives from emotional responses to environmental stimuli, and that artistic inspiration is thus largely the result of a flexible-emotional mode of processing. It should be added that artistic creativity can also draw substantially on cognitive knowledge domains (e.g., music

theory, mathematics) and cognitive persistence (especially in non-real-time tasks, e.g., composition). Creative insights—defined as the conscious realization of an idea in working memory—can occur via any of the processing modes (Dietrich, 2004).

More recently, the dual pathway to creativity model was developed, positing the existence of persistence and flexibility pathways (Nijstad et al., 2010; De Dreu et al., 2011, 2012). The “persistence pathway” is critically supported by working memory, while the “flexibility pathway”, characterized inhibition, defocused attention, and automatic spreading of activation, is only minimally dependent on working memory. The authors behind the dual pathway model posit a role for emotion in creative performance, arguing that performance on any type of creative task can be influenced by the performer’s emotional state. Both trait (personality associated) and state (temporarily activated) related mood characteristics are thought to mediate processing along persistence and flexibility pathways (De Dreu et al., 2008; Nijstad et al., 2010). “Activating” moods that are positive in tone (e.g., happiness) can improve creative performance by promoting cognitive flexibility, while activating moods that are negative in tone (e.g., fear) improve creative performance by promoting cognitive persistence. “Deactivating” moods (low in arousal; e.g., relaxation, sadness) seem to offer no such benefit for creative performance.

Schubert (2012) proposed a model for musical creativity that explains how processing might proceed along the flexibility pathway, in particular—though a similar explanation might also describe processing along the persistence pathway. Based on spreading activation theory, the model posits that activation spreads between nodes – which are abstract units representing knowledge and emotions—via links representative of learned associations.

Schubert describes the process of spreading activation as automatic, proceeding with or without conscious attention, and suggests that creative inspiration occurs when new paths form spontaneously between previously unconnected nodes. The process of spreading activation is driven by a desire to activate “pleasurable” nodes and inhibit “painful” nodes. As an example, improvisation involves constructing new musical sequences that fit within a given framework: musicians are guided in this task by the pleasure that comes from alighting upon ideas for patterns that fit, while simultaneously avoiding patterns that break from the framework. Central to Schubert’s model is the idea that maintaining positive feelings is a critical component of musical creativity. As a potential extension to the model, it might be argued that processing along the persistence pathway involves controlled, incremental exploration through the network of nodes, driven by the same desire to achieve pleasing results.

The models presented in this section were developed to explain creative performance on individual, not collaborative tasks. How generalizable are these ideas to collaborative situations? The pool of cognitive resources available to a group is greater and more varied than would be the case for individuals performing a similar task alone. As a result, there is the potential for a wider variety of associations between ideas to be made. This could lead to more creative performance – but it could also lead to a lack of cohesion between outputs. There is also potential

for conflicts to arise within the pool of cognitive resources that either facilitate or impair the creative process. For example, if members of an ensemble were to have different concepts of how a performance should progress (i.e., how the structure should unfold), then in terms of Schubert’s spreading activation model, an idea that is “pleasurable” for one performer might be “painful” for another. Further investigation of how ensemble members negotiate specific, open-ended problems (e.g., interpretation of particularly ambiguous passages of a new piece) might clarify how conflicts are resolved and facilitate development of a model that accounts for collaborative creativity.

3.2. Using Musical Imagination to Elaborate and Evaluate Ideas

The idea of spreading activation is closely linked to the idea of imagery. Indeed, they could be said to describe two parts of the same process: spreading activation is the mechanism through which nodes in a knowledge network are selected, and imagery is the activation of those nodes in memory. This section of the paper discusses the potential role of musical imagery in facilitating the search, selection, and evaluation of ideas during music performance.

Musical imagination has been suggested to underlie creativity in both music perception and performance (Hargreaves, 2012). *Musical imagination* refers to the human capacity to experience music in a way that is not a direct and immediate consequence of having perceived it. In the current paper, the term *musical imagery* is used to refer to the process of experiencing music in this way. While musical imagery has traditionally been defined as a form of mental imagery, I will avoid characterizing it as a specifically and exclusively mental process, as it might also be said to involve activation of the motor system, even if no overt movement is apparent (Aleman and Wout, 2004; Chen et al., 2008; Bernardi et al., 2013a; Bishop et al., 2014).

Musical imagery involves the multimodal activation of musical knowledge and the (re-)construction of musical stimuli in working memory. It is to be distinguished from the process of remembering details about music: recalling that Rachmaninoff’s Piano Concerto No. 3 begins in the key of D minor is different from imagining the sound of the first chords or the feel of playing the piano line. The pitch (Aleman et al., 2000), timing (Janata and Paroo, 2006; Jakubowski et al., 2016), dynamics (Wu et al., 2011; Bishop et al., 2014), and timbre (Halpern et al., 2004) of perceived music can be imagined with high veridicality. Emotion is also perceived similarly in sounded and imagined music (Lucas et al., 2010).

Musical imagery is sometimes—but not always—a controlled process, and people are sometimes—but not always—aware of it. It should therefore be described as a process that is accessible to attention. Sometimes mental images are the focus of attention; for instance, during mental rehearsal (Bernardi et al., 2013b; Bach et al., 2014) or when distracted by an earworm (Müllensiefen et al., 2014; Floridou et al., 2017). Such instances are most likely to occur offline (i.e., not concurrent with overt performance, though still evolving in real-time). Online, many concurrent processes compete for a performer’s attention, so even though imagery

can be an important part of the action planning process, it might proceed largely without the performer's awareness (Keller and Appel, 2010; Bishop et al., 2013). Musical imagery is less often referred to in the context of music perception, but can nonetheless be said to contribute. Sounded music unfolds over time, and listeners must maintain some evolving representation of it in memory in order to make sense of the structure. Evidence of this process can be seen in the way listeners are able to re-interpret previously-established tonal contexts when incongruous chords are added to a progression (Bailes et al., 2013).

If musical imagery involves re-activating musical knowledge in memory, what is its relation to creativity, defined as the generation of something new? In other words, what is the difference between creative imagery and recall? Benedek et al. (2014) examined brain activity while people were engaged in a divergent thinking task (the alternate uses task), and observed different patterns of activation during recall of known ideas and generation of new ideas. In particular, the generation of new ideas involved activation of the left inferior parietal cortex, which has previously been linked to imagery and mental simulation. Creativity on this task was demonstrated through the recall of known ideas and application of those ideas to a novel situation: giving "swing" as a possible use of a tire was considered a recalled idea, as participants had seen it before, while giving "picture frame" as a possible use was considered a new idea.

Imagery allows people who are engaged in creative tasks to evaluate the appropriateness and originality of activated ideas before expending energy in externalizing them. It may also play a critical role in the type of controlled and structured idea generation that is associated with the persistence pathway. As described above, the persistence pathway draws on working memory: performance on tasks that encourage controlled generation and evaluation of ideas has been shown to suffer under high cognitive load conditions (De Dreu et al., 2012). This study also showed that cellists with high working memory capacity performed increasingly creative improvisations across several trials, while cellists with low working memory capacity performed decreasingly creative improvisations. The authors suggest that improvisation requires a great deal of planning and mental structuring, especially in cases where several rounds of improvisation will be required, and a high working memory capacity helps with maintaining a representation of that structure. When musical ideas are maintained in working memory (i.e., imagined attentively), they are accessible for reflection and evaluation. Musicians may, therefore, use imagery to structure their search for creative ideas and reflect on possible outputs.

In addition to its roles in idea generation and evaluation, musical imagery allows for manipulation of recalled material without interference from externalized sounds or movements. Composers, in particular, report using imagery to evaluate and elaborate on their ideas. Some claim that this is critical to do before trying to translate those ideas to an instrument or score, as creative thinking becomes more constrained and ideas become harder to change after that point (Agnew, 1922; Bailes and Bishop, 2012). For performers, the process of deliberately

manipulating or elaborating on images often occurs offline (e.g., during mental rehearsal, or when deciding how a piece should sound).

Online, there is not usually time to imagine different variations of an idea before implementing it. However, the malleability of musical images – the fact that they can be disrupted by incoming signals or deliberately manipulated – may be critical for creative performance. This malleability may help performers to be flexible in their playing, allowing them to adjust for errors (Glowinski et al., 2016) and accommodate new ideas in real-time (either their own or their co-performer's). The use of anticipatory imagery as a means of guiding musical performance has been studied empirically (Keller and Appel, 2010; Keller et al., 2010; Bishop et al., 2013) and described anecdotally by highly skilled musicians (Trusheim, 1993). Anticipatory imagery involves activating evolving expectations of how musical output should sound, feel, and/or look, and facilitates selection of the action parameters needed to achieve the desired output by way of inverse perceptual-motor activation (see perceptual-motor coupling in section 2.2). That these expectations are accessible to attentive reflection (even if not always attended to) is important: this feature of the imagery process enables deliberate revision of plans as well as constant monitoring of performance success.

As stated above, imagery contributes to listeners' abilities to make sense out of musical performances as they unfold over time. This role of imagery in music listening is central to ensemble performance, because a large part of the task of performing with a group is listening to and taking cues from each other. Inter-performer communication is discussed in the next section of this paper, but here, I want to emphasize how important it is for ensemble musicians to listen to each other with "open ears" in order to perform creatively. That is, while hearing the combined output of the group, they must be open to receiving new ideas, changing their interpretation of already-performed structures, and pursuing deviations from the prescribed script that is guiding their performance. This openness requires a strong awareness of the group's current and previous output, which, I would hypothesize, takes the form of a flexible guiding image.

3.3. Communication Drives Alignment of Ideas

The term "communication" refers broadly to the transfer of information that occurs between members of a group. Communication between ensemble members can take many forms: fluctuations in audio signals produced by an instrument, audible breathing, shifts in eye gaze, changes in posture, overt gestures, or facial expressions. The information that is transferred might relate to performer's interpretation of the music, their engagement in the task, a shift in roles, or an acknowledgment of a mistake, among other things.

Some communication between musicians is necessary for ensembles to perform coherently. This is clearly shown by studies testing musicians' success at playing with disrupted communication channels—while eliminating visual communication between performers has relatively minor effects, eliminating audio communication leads to substantial temporal

misalignment (Bishop and Goebel, 2015). Delays in audio communication likewise impair coordination, even rendering performance non-interactive if the delays are large enough (Bartlette et al., 2006). Here, research that has been done on communication in music ensembles is considered, along with some criticisms of the assumptions that underlie this research and some recent studies that attempt to test these assumptions.

3.3.1. Sharing Intentions: Simulation and Prediction

Widespread in the literature on ensemble performance—and in the broader literature on joint action—is the idea that collaborating members of a group each have individual intentions regarding their own contribution to a task, as well as shared intentions regarding how their individual contribution will fit into the group's combined output (e.g., Keller, 2001). Performers' *intentions* encompass their action-oriented anticipatory imagery as well as knowledge relating to the expressive constructs that they plan to implement. The intentions performers have are multi-leveled and exist in parallel across overlapping time scales. High-level intentions, which span a long time frame, might relate to the overall structure of the performance (e.g., formal structure as notated in a score) or general expressive content. In contrast, low-level intentions, which are directly involved in action planning, unfold rapidly and often without conscious control. Coordinating a joint performance successfully requires individual performers to share clues to their own intentions while also monitoring the signals given by others.

Musicians' low-level intentions are often studied by manipulating the expectedness of the sounds that their movements generate. When manipulated sound output induces performance errors or other compensatory behavior, we conclude that those manipulations were not in line with the musicians' intentions, and that their action planning system is trying to correct for the "error." Responses to unexpected sound output can also be observed in readings of brain activity. Research using these methods has shown that when playing duets, pianists anticipate the sounds of their own and their partner's key-presses, as well as the combined output (Loehr et al., 2013). For instance, novice pianists who learn to play a simple melody with live accompaniment perform better at test with accompaniment than without, suggesting that they learn their own melody in terms of how it fits into the combined output (Loehr and Vesper, 2016).

Action simulation is thought to underlie musicians' anticipation of others' sound output during music performance (Jeannerod, 2003). This process of covert action representation engages coupled perceptual-motor brain networks without necessitating overt movement (Patel and Iversen, 2014). Simulation is facilitated when the action and its resulting sound are strongly coupled in the brain. More effective simulation leads to better anticipation and improved temporal coordination between performers (Keller et al., 2007; Wöllner and Cañal-Bruland, 2010).

Communication between performers is thought to support action prediction processes by providing cues to initiate the simulation process. While auditory communication in the form

of a musical sound signal is usually sufficient for performers to maintain temporal coordination, they sometimes supplement their audio signals with visual signals (Badino et al., 2014; Kawase, 2014; Bishop and Goebel, 2017). Ensemble musicians are better able to predict the course of observed gestures when those gestures fall within their practiced repertoire (Wöllner and Cañal-Bruland, 2010; Bishop and Goebel, 2014), and better able to predict such gestures than are novice musicians (Luck and Nte, 2008; Petrini et al., 2009; Lee and Noppeney, 2014). It seems that for ensemble musicians, simulating co-performers' actions in response to a visual cue is a well-practiced task.

3.3.2. Sharing Intentions: When Is It Necessary?

The idea that successful ensemble performance necessarily involves performers communicating their individual intentions to each other and, ultimately, constructing shared intentions, has been a source of debate in the literature. Under some conditions, it is argued, coordination can emerge from local (often pre-reflective) responses to the gradually unfolding musical output, making a shared global plan and explicit communication unnecessary (Hutchins, 1990; Linson and Clarke, 2017). This is the perspective generally endorsed by the EMC approach (Schiavio and Høffding, 2015; Maes, 2016).

The description of coordination as emerging dynamically from local interactions is in line with musician's descriptions of group flow—as discussed in Section 4.2, group flow is characterized by joint feelings of effortlessness, a lack of self-awareness, and non-reflective patterns of thought. On the other hand, ensemble musicians also communicate with each other reflectively through overt body gestures and deliberate manipulation of sound output, particularly when working together to construct an interpretation of notated music (Williamon and Davidson, 2002; Davidson, 2012). Such evidence suggests that ensemble performance may be supported by different types of communication under different conditions (MacRitchie et al., 2017). Relevant to the current discussion is how much ensemble musicians draw on reflective and pre-reflective types of communication when performing naturally and creatively.

Ensemble musicians have been shown to exchange communicative gestures deliberately at critical moments in their performances, as a way of facilitating note coordination. Such cueing gestures often take the form of hand/arm movements or head nods (Bishop and Goebel, 2018). Breathing gestures are likely used as well and have the benefit of providing an audiovisual cue, but they are more difficult to measure experimentally. Cueing gestures can be exchanged at moments of sudden tempo or meter change (Kawase, 2014) or at piece entrances or re-entrances that require synchronization between performers (Bishop and Goebel, 2015, 2017). These are ambiguous, isolated moments when co-performer's expectations about how to play might not otherwise align. Performers might be expected to make greater use of communicative gestures during the early stages of rehearsal, when still unclear on how they would like the music to sound, than when performing well-practiced pieces. On the other hand, in some cases, gestures at structurally or expressively significant

moments are retained through rehearsals and integrated into the performance script (Williamson and Davidson, 2002).

Visual communication between ensemble members may be particularly important when the demands on creativity are high, and a number of temporally ambiguous moments arise in the music. Note, however, that cueing gestures serve primarily to clarify irregular timing; whether or how they contribute to the coordination of other parameters remains unclear. Thus, greater use of visual communication during some types of creative performance (e.g., playing notated music with no meter) but not others (e.g., improvisation of temporally-regular music) might be expected.

Some recent studies, outlined below, have started searching for evidence that ensemble performance is supported—at least partially—by low-level interaction between members. These studies reject the assumption that coordination between ensemble members is necessarily dependent on the construction of shared intentions.

In an attempt to determine whether shared intentions are truly needed for a coordinated ensemble performance, some study has been made of collective free improvisation (CFI), a form of improvisation drawn upon in several musical genres. With other forms of improvisation, it is standard for performers to identify a framework to help structure their playing by reducing the range of possible contributions they could make. Such a framework, or “referent,” might include aspects of large-scale structure (e.g., in jazz, how many choruses to cycle through), melodic/harmonic content (e.g., themes, chord progressions, keys to use), leader/follower roles (e.g., a pre-arranged order of solos), and perhaps also some expressive content. Musicians engaging in CFI, in contrast, deliberately eschew the use of a shared referent, instead constructing musical structure in real time (Canonne and Garnier, 2015).

Canonne and Aucouturier (2015) tested for the presence of “shared mental models” (i.e., schemas) among musicians who regularly perform CFI. In particular, the hypothesis that musicians would have overlapping concepts of the CFI task and overlapping interpretations of certain musical elements was investigated. Musicians categorized musical excerpts from CFI performances based on how they would respond musically. Response similarity was calculated between participants and subjected to a nearest neighbor classification algorithm, which predicted familiarity between participants with higher than chance accuracy: musicians who performed together tended to interpret the musical excerpts similarly. Such a shared understanding of the music could (unintentionally) give collaborating musicians a common language with which to exchange ideas.

Pachet et al. (2017) tested the hypothesis that ensemble performance is partially driven by low-level interactions that emerge as relationships in the acoustic features of collaborating performers. These relationships are distinct from those that emerge as a result of performers adhering in parallel to a prescribed structural framework (“score effects”), and instead attributable to real-time interaction. A number of acoustic features were extracted from six improvised performances recorded by a five-member jazz bebop band, and

comparisons were made between individual performers. No pair of features correlated reliably across performances, so even though significant correlations occurred within performances, the possibility that these were attributable to score effects could not be ruled out. As the authors point out, whether signs of pre-reflective interaction might be seen with higher-level information in performer’s audio signals (e.g., rhythm patterns) is yet to be tested.

Interaction between performers might also emerge as relationships in features of their body movements—in particular, their ancillary body movements, or those not directly involved in sound production. Some studies of piano duet performance have shown evidence of coordination in patterns of pianists’ head movements (Goebel and Palmer, 2009) and body sway (Keller and Appel, 2010). In a study by Ragert et al. (2013), pianists learned either one or both parts of piano duets, which they then performed for recording with another pianist, as their body movements were tracked. Pairs of pianists who knew both parts of the duets displayed a steady high degree of coordination in their body movements throughout the performances, while pairs who had learned one part were less coordinated at the start, but increased their coordination as the experiment progressed. The authors suggested that practicing both parts of a duet allowed pianists to construct a more thorough image of the piece structure, which facilitated timing predictions at the relatively long time scales at which head and torso movements unfold, improving movement coordination. This explanation implies that pianists intend, at some level, to coordinate their body movements, however, which may not be the case. An alternative explanation is that pairs of pianists who knew the full pieces were more likely to share an interpretation of it than were pairs who each knew a different part, and tended to display similar patterns of motion as a result of their overlapping interpretations.

On the other hand, a study by Badino et al. (2014) provides some evidence of ensemble musicians influencing each other’s movements, indicating that coordinated patterns of ancillary movement can emerge as a result of performers’ interactions. Head movements were tracked for members of a professional string quartet during performance under normal and perturbed conditions (in which the first violinist introduced unexpected expressive changes). Across takes, the first violinist exerted the strongest influence over the other musicians (measured with Granger causality), though his influence was reduced during the perturbation segments. Musicians’ combined influence over each other was highest during technically complex sections of the piece, suggesting an increase in the communicative value of their movements during these sections.

The function of coordination in performers’ ancillary movements is not clear. It could be an aesthetic aim, for the benefit of an observing audience, or meant to facilitate note coordination. It might also serve a motivational function by enhancing the feeling of interaction and engagement. Recent research on visual attention suggests that duo performers look at each other more often than we would expect if they were seeking only to clarify irregular timing (Bishop and Goebel, 2017). Instances of two-way eye contact also occur at predictable points in the performance, indicating that performers are not

solely driven to look toward a “leader” for timing cues; instead, both performers monitor each other, perhaps as a means of communicating and confirming each other’s engagement and understanding.

The literature described here paints a still-unclear picture of the nature of the communication processes that drive ensemble coordination – particularly the processes that drive the real time coordination of new and spontaneous ideas. The field is especially in need of further systematic study of low-level communication mechanisms. Until recently, it was difficult to capture low-level features of visual communication, in particular, in meaningful detail. However, developments in motion capture and motion analysis techniques—especially techniques that enable us to quantify the influence that collaborating musicians have over each other’s movement patterns (e.g., Badino et al., 2014; Walton et al., 2017)—provide a promising means to understanding emergent coordination.

4. EMERGENCE AND GROUP FLOW

In the Western classical music tradition, musicians prepare for public performances of a piece with extensive rehearsal and careful study of the score. Yet at the same time, they value creativity and spontaneity, as do their audiences (Repp, 1997b; Chaffin et al., 2007). A series of studies by Chaffin et al. (2006, 2007, 2010) have investigated how skilled musicians maintain enough control over their performances to be able to make spontaneous interpretive decisions, despite simultaneously drawing on highly automatized movements. The results of these studies suggest that creativity in performance depends on where musicians focus their attention. If attention is directed away from the music (e.g., focused on a distracting audience member or the performer’s own anxiety symptoms), performance is likely to be automatic and uncreative; if attention is directed toward the music but focused on errors, performance is likely to be uncreative and cautious. Skilled performers seem to construct a structure of attention cues during rehearsal that relate to different aspects of expression and technique. These cues help musicians focus their attention during performance and allow for conscious interpretive decisions to be made.

For music ensembles, spontaneity in interpretation can manifest as emergence. Occurrences of emergence, along with group flow, seem to characterize ensemble performance at its peak. In this section of the paper, the concepts of emergence and group flow are addressed and conditions that encourage their occurrence are identified. In particular, an external focus of attention (i.e., toward the musical output and away from the self), which Chaffin has shown to be key for managing creativity in interpretive decisions, seems to be critical for achieving flow; likewise, shared knowledge of an intended guiding framework for the performance is thought to be important.

4.1. Emergence as a Function of Group Interaction

Emergence, as defined in section 2.1, is a phenomenon that occurs when the collective output of the group amasses to greater than the sum of individual contributions. In some ways, ensemble

performance is necessarily emergent, as individual contributions combine to form cumulative units with distinct structural meaning (e.g., three notes played by three performers combine to form a chord, which as a complete unit has meaningful harmonic implications that none of the three notes have independently). More relevant to the current paper, however, is emergence that corresponds to flexibility in interpretation of a prescribed structure (in the case of notated music) or the construction of substructures (in the case of improvisation within a set framework).

An alternate way of defining emergence is to say that it occurs when a group performs in a way that cannot be attributed to any one individual contributor. It can be argued that ensemble performance is not always emergent in this way. For example, social factors (e.g., skill level, age, position in a social hierarchy, etc.) or piece structure can combine to encourage performers to fall into leader/follower roles, which can result in one person making most of the interpretive decisions. Furthermore, ensembles do not always achieve what they set out to achieve, and while the goal might be a performance that is original and spontaneous, the outcome is sometimes poorly coordinated or uninspired.

A study by Hart et al. (2014), examined performance on the “mirror game” (Noy et al., 2011), a task for dyads that involves moving a pair of horizontal sliders back and forth along a track to create coordinated patterns of improvised movement. Periods of smooth, highly-synchronized motion emerged, which a subsequent study found to coincide with increases in heart rate and increases in correlation of heart rates between performers (Noy et al., 2015). Suggestive of emergence, these periods were less likely than other performance segments to carry the signatures of either performer’s individual style. Notably, between-group overlap in motion characteristics was high, suggesting that these periods of emergent coordination were supported by predictable, rather than idiosyncratic, movement. As discussed in section 3.3.2, when necessary, performers can manipulate aspects of the audio and visual signals that they exchange to increase their predictability to each other.

The relationship of emergence to flow states, described below, is unclear. Is emergence more likely during periods of flow? Emergence in music performance, as defined in this paper, is potentially complicated to identify because it requires comparing the combined output of an ensemble to the output that individuals would produce if performing their part alone. A reliable method of quantifying differences between individual and group interpretations (e.g., similar to that used by Noy et al., 2011; Hart and Di Blasi, 2015) has yet to be defined. In future research, it will be necessary to investigate how often periods of emergence occur during ensemble performance, what prompts them, and how they shape audience members’ perception of performance quality and expressivity.

4.2. External Focus of Attention and Shared Knowledge Support Group Flow

Musicians sometimes find themselves in a state of acute absorption: wholly focused on the task of performing, they feel an intense connection to the music, which flows out seemingly

effortlessly. This rare but rewarding experience is called flow, and is generally thought to arise from an optimal match between task demands and the performer's skills, which fuels a sense of intrinsic motivation (Keller et al., 2011). The concept of "flow" was originally identified at the individual level (Csikszentmihalyi, 1990), and has only more recently been found to occur at a group level (Sawyer, 2006). It is important to note that group flow is an emergent quality of groups engaged in creative performance; it is not reducible to study at an individual level, and it is not the same as individual flow in a group setting (Sawyer, 2006).

In a qualitative study involving interviews with regularly-performing (improvising) musicians, (Hart and Di Blasi, 2015) identified some common themes in musicians' descriptions of their flow experiences. Musicians described the conditions that they thought were necessary to build up to a state of flow, including being able to establish and maintain a sense of individuality and dismiss feelings of self-consciousness. Another theme that came through was the idea that flow states require a lack of awareness of the self and less reflective patterns of thought. The musicians spoke about not appreciating (reflecting on) the performance as it happens, and being unable to remember afterwards what they had played.

The term "mutual engagement" has been used to describe interperformer interaction during periods of group flow (Bryan-Kinns and Hamilton, 2009; Bryan-Kinns, 2013). Performers in this state are engaged with each other and with the music they are producing. Some conditions are posited to underlie performers' achievement of a group flow state, including a mutual awareness of each other's actions (i.e., who is contributing what and when), shared representations of the intended outcome, equal access to musical output, the possibility of modifying each other's output (e.g., by responding to it), and the possibility of communicating around the output (rather than exclusively through it; e.g., visually, through body gestures; Bryan-Kinns and Hamilton, 2009).

According to the Networked Flow model, group flow develops through three stages, which draw on successively higher levels of empathy (Gaggioli et al., 2013). Central to this model is the concept of social presence, described as an individual's ability to interact with others by understanding and sharing their intentions. At an initial stage, "proto-social presence" involves performers recognizing each other's motor intentions. The second stage, "interactive social presence," involves each performer individually recognizing those intentions that are directed toward him/her. At the final stage, "shared social presence" involves performers entering into resonance with each other. Some support for the model was offered by a study of performance quality and self-report measures of group flow and social presence among rehearsing (3–7 member) bands. A positive relationship was observed between self-reported measures of group flow and social presence. Flow also related to self-ratings of performance quality, though not to expert ratings (Gaggioli et al., 2017).

A point of overlap between the Networked Flow model and the mutual engagement paradigm is the idea that shared knowledge of individual intentions is needed for group flow to emerge. What constitutes "intentions" is not entirely clear

(see also section 3.3.2), but at a minimum, it is likely that performers must at least agree over the intended structure of the performance. Musicians' descriptions of their group flow experiences suggest that individual group members need to feel that they have a specific and valuable role to play—that is, they need to be able to conceptualize how their own contribution will fit into the collective outcome (Hart and Di Blasi, 2015). On the other hand, musicians may also benefit from having few constraints to limit the possible contributions that they can make (Canonne and Aucouturier, 2015). In a study by Walton et al. (2017), musical duos reported a greater sense of freedom when improvising over a drone backing track than when improvising over a swing bass line. They felt that the drone encouraged a greater degree of interaction. Indeed, their coordination in sound output and body movement was higher during improvisation over the drone.

In addition to shared structural intentions, a shared emotional state might also promote group flow. Seddon (2005) distinguished between sympathetic and empathetic levels of attunement, positing that sympathetic attunement between performers supports coordination of a cohesive performance, while empathetic attunement is necessary for flow states and the "spontaneous musical utterances" that characterize emergence (see also Seddon and Biasutti, 2009). Progression to empathetic attunement can be impaired by interperformer conflicts in musical style or skill. Empathetic attunement requires performers to assume each other's musical perspectives, and is therefore thought to draw on their capacity for empathy. Indeed, prior research has shown a correlation between duet performer's scores on measures of empathy and the strength with which they represent their co-performer's part (Novembre et al., 2012).

Empathy is defined on two dimensions: cognitive empathy relates to capacity for perspective-taking, while emotional empathy relates to the flow of feelings between people (Babiloni et al., 2011). The process by which emotional states spread from one person to another—called emotional contagion—is speculated to occur during creative collaboration, and could potentially help to support emergent coordination. Emotional contagion seems to occur between performers and listeners (Lundqvist et al., 2009) and empathy has been shown to mediate the process (Egermann and McAdams, 2013). However, whether this also occurs within performing ensembles has not yet been confirmed. In one study, ensemble musicians reporting on completed performances showed less overlap in their experienced affective states than in their perceptions of leadership (Morgan et al., 2015).

As a final point, group flow in music ensembles could be encouraged by a shared cooperative, rather than competitive, mindset. Outside the music domain, in the context of verbal divergent thinking tasks, the effects of cooperative vs. competitive mindsets are mediated by regulatory focus—that is, the tendency to attend to either promotion goals (aiming to achieve an "ideal self" through growth and development) or prevention goals (aiming to achieve an "ought self" by preventing failure; Bittner and Heidemeier, 2013). Activating a promotion focus seems to prompt people to adopt a cooperative strategy, which improves performance on the task, while activating a prevention focus prompts people to adopt a competitive strategy, which

worsens performance. As a general rule, we can assume that most ensemble musicians intend to cooperate with their co-performers; however, it is possible that some performance situations prompt a prevention focus and/or an intention to compete. For instance, a student ensemble participating in a competition might be preoccupied with preventing technical errors or outperforming other groups, and in doing so constrain their own creative processes.

5. RESEARCH DIRECTIONS

Coordination is a broad and multilevelled construct. In this paper, the focus has been on coordination during creative performance – a high-level task that requires alignment of spontaneously-generated ideas in real-time, without prior practice. The processes that support lower levels of coordination (e.g., synchronizing periodic taps or regularly-timed duets) may be insufficient to explain the high-level coordination of creative ideas that ensemble musicians can achieve. I have highlighted some of the processes that could account for aspects of during ensemble performance. However, our discussion has raised a number of issues that are still relatively unexplored. Below, three lines of research are outlined that would benefit from further attention.

5.1. Explaining Emergence and Group Flow

Further study of emergence and group flow will be critical to identify the mechanisms engaged by collaborative creativity during performance. Earlier, I made reference to some interview studies with ensemble musicians; these have been useful for obtaining descriptions of flow experiences from a first-person perspective, and have lent support to models that propose explanations for how flow and emergent coordination develop including models by Bryan-Kinns and Hamilton (2009) and Gaggioli et al. (2013). Largely absent from the literature, however, are systematic, empirical studies that test these models. We have some idea of the conditions that are necessary for group flow to develop, but what triggers the onset of a flow state? What conditions trigger emergence? How do group flow and emergence relate, and how these states maintained? How resistant is flow to perturbations resulting from technical errors or environmental disruptions?

At this stage, the answers to these questions seem largely theoretical. The literature might benefit from further efforts to manipulate potentially relevant factors and induce flow experimentally. The importance of musicians' focus (self-reflective vs. external) and interaction in real-time via auditory and visual channels might be tested this way. Focus could be manipulated by catering the instructions that performers receive: instructions that direct attention toward individual success or accuracy should encourage a self-reflective focus, while instructions that direct attention toward a particular expressive goal might encourage a external focus. Ideally, performances should be given under as naturalistic conditions as possible (e.g., established ensembles playing familiar under self-selected constraints). Flow would be best assessed using a combination of self-report, physiological, and behavioral (e.g., performance

output, body movement) measures. In such a study, it would also be useful to analyse performance data for evidence of emergence; for example, by comparing solo and ensemble performances of the same material (e.g., as in, Hart et al., 2014; Noy et al., 2015).

Investigation of how group flow emerges during performance in non-Western musical traditions would also improve our understanding of the phenomenon—particularly in cases where performances are occasions for widespread participation, and there is not a strict performer/audience separation (Hill, 2012). In such cases, musicians may tend less toward a self-reflective focus than do musicians in Western traditions, who are often preoccupied with individual success and audience judgments (Hart and Di Blasi, 2015).

The relevance of musical imagery to group flow is still also a source of debate. As argued in section 3.2, imagery could facilitate flexibility during creative performance. According to Cochrane (2017), flexibility in performance means being able to choose between several responses to a given stimulus, and should only be possible for performers who can represent the possible responses before carrying one out. This should especially be the case when the musical structure is complex and requires sophisticated interpretation. Some authors have argued that imagery, or more generally, private intentions, are not necessary for ensembles to coordinate a cohesive performance (see section 3.3); however, it is unclear what other mechanisms could account for the flexibility seen in skilled performance. Cochrane (2017) goes on to explain how performers' intentions may critically underlie their flow experiences. While playing, performers monitor the disparities between their intended and output sound; disparities create a sense of tension, which is alleviated when the intended and output sounds match. The alleviation of tension enables a reduction in self-consciousness and perceived effort, allowing performers to focus on musical output in a way that is characteristic of flow. Thus, maintaining (and overtly realizing) intentions could enable the development of flow states. Further study would be needed to test this hypothesis.

Ultimately, many musicians could benefit from a clarified understanding of what causes group flow and how to encourage it. At present, research is still needed to identify the effects of flow states on musical output. As mentioned in section 4, Gaggioli et al. (2017) found that ensemble member's ratings of their own performance quality related to measures of group flow, while ratings of performance quality given by independent experts did not. Thus, the perception of success that motivates performers and fuels their sense of effortlessness may not relate reliably to the quality of musical output as perceived by an audience. On the other hand, audiences are sensitive to aspects of the interaction that occurs between ensemble performers. Aucouturier and Canonne (2017), for instance, showed that listeners use cues relating to temporal and harmonic coordination to decode social intentions (attitudes such as domineering, disdainful, or conciliatory) in improvised duo performances. Attentive audiences may pick up on evidence of group flow, and their perception or engagement with the performance might be enhanced as a result.

The literature would also benefit from more thorough investigation of the physiological and social effects of flow. Physiologically, flow has been shown to share an inverted u-shaped relationship with stress-induced sympathetic arousal, and a positive linear relationship with parasympathetic heart rate control (Peifer et al., 2014). Cohen and Bodner (2018) observed a strong negative relationship between the occurrence of flow and performance anxiety among classical orchestral musicians, and suggest that devising means of encouraging flow might help reduce the effects of performance anxiety. Socially, some of the factors that support group flow, including joint attention (Wolf et al., 2015) and rhythmic synchronization (Hove and Risen, 2009), are also thought to underlie the heightened affiliation that has been shown to develop between musical partners. We might hypothesize that the bonding effects that are seen generally as a result of ensemble playing (Tarr et al., 2014; Pearce et al., 2016) are exaggerated in instances of group flow.

5.2. Recalling and Creating: Can we Imagine What we Have Not Perceived?

Though “free” because it is not driven by incoming stimuli, musical imagery is simultaneously constrained by the perceptually-shaped cognitive space in which it is carried out (Leman, 2001). As discussed in Section 3.2, imagery involves reconstructing elements of previously-perceived material. The process of reconstruction can be fairly accurate, yielding musical images that retain many of the parameters of the original percepts. Relevant to the issue of creativity, however, is the question of how free people are to manipulate or elaborate previously-perceived material. To what extent can people imagine what they have never perceived?

This question is particularly relevant to collaborative musical creativity, where, in optimal cases, the music that is produced is distinct from what individual group member would have produced alone (i.e., emergence occurs). As this paper has discussed, to achieve emergence in a collaborative performance, individual group members must be flexible enough to accommodate and elaborate on novel ideas. Sometimes – if the group includes members with vastly different musical backgrounds, or the musical genre encourages experimentation with sound and structure – the range of ideas that arise might be broad. In such cases, an ability to imagine musical structures (e.g., tone qualities, meters, pitch intervals) outside the performer’s prior experience would be beneficial, if not critical.

It is important to note that the process of imagining music is an imperfect one, even if the aim is a precise reconstruction of a specific stimulus (Large et al., 1995; Dowling et al., 2002). Details of a musical experience can be erroneously perceived or encoded, or insufficiently embedded in a network of associations, making them difficult to retrieve. The use of heuristics and schemas in facilitating reconstruction can also lead to errors (Vuvan et al., 2014). Thus, people almost never imagine music precisely as it was perceived. More important for creative thinking is the ability that people have to selectively recall elements of prior perceptual experiences and recombine them into something new (“combinatorial play”). This is what we

assume happens when musicians imagine a new improvisation or a new interpretation for a practiced piece: details relating to pitch, timing, instrumental tone, and dynamics are drawn from well-established networks of musical knowledge and re-assembled in a new way.

These images can then be manipulated. It was in the visual domain that evidence of “emergent properties” in imagery was first found—that is, evidence that images can be reinterpreted, allowing patterns to emerge that were not noticed at the time of perception. People can reinterpret simple geometric shapes in memory (e.g., identify new shapes formed by imagining a capital “H” superimposed on a capital “X”); complex shapes prove more difficult, probably because they require more resources to maintain in working memory in sufficient detail (Finke et al., 1989). In the musical domain, trained musicians using “notational audiation” to imagine music from a score are able to extract familiar melodies hidden in embellished phrases (Brodsky et al., 2003). Foster et al. (2013) tested musicians’ abilities to imagine pitch and timing transformations (i.e., pitch transpositions and melody reversals) on simple melodies. The transformation task was found to activate parts of the bilateral intraparietal sulcus, a region that has been previously associated with visuospatial transformation and calculation.

Imagery may also facilitate translations between musical stimuli and sensations or perceived events. Music is an effective and versatile means of nonverbal communication, in part, because it activates so many associations for those involved in producing or hearing it. These associations often relate to emotion, and as such, the communication of emotion has received a great deal of attention in the literature (Juslin and Laukka, 2004; Molnar-Szakacs and Overy, 2006; Lundqvist et al., 2009; Lucas et al., 2010). Other constructs are known to be communicated as well, though, including sensations of motion (Eitan and Timmers, 2009; Olsen and Dean, 2016) and interpretations of musical structure (Clarke, 1993; Toiviainen et al., 2010). Even complex environmental events, such as animal behavior, changes in season, landscapes, or city life, can be communicated musically (without the aid of lyrics). Wong and Lim (2017) found imagery to facilitate children’s creativity on a music composition task. Young children were instructed to construct audiovisual images of animals before composing short melodies in which the animals “came alive.” Scores of creativity (judged by experienced music teachers) were higher for participants in the imagery condition than for participants who did not receive imagery instructions. Thus, imagery may have helped participants translate between knowledge of animal characteristics and acoustic representations of those characteristics. Using imagery as a means of translation between modalities and representations arguably constitutes imagining what we have not perceived.

In sum, people have the ability to manipulate musical images in ways that deviate substantially from music they have perceived in the past. Whether people can create new tone qualities in their imagination (e.g., when constructing a new instrument or synthesizing a new sound) remains unclear. This is a question that should be addressed, especially given the increasing popularity of music that uses non-traditional methods

of sound production or sound modification (e.g., electric guitars, synthesizers, digital musical interfaces, algorithm-based voices, etc.). Do musicians imagine the tone quality that they want to achieve before attempting to match it acoustically? Likewise, during group performance of music incorporating synthesized/digital sounds, how do performers adapt their sound to match (or compliment) their co-performers' sounds?

It will also be important to continue investigation of the motor aspects of musical imagery. Specifically, to what extent can people replicate in imagination what they have not previously performed (or do not know how to perform)? Section 3.3.1 discussed the role that motor simulation might play in interpersonal coordination. As ensemble musicians make increasing use of non-acoustic instruments and/or perform alongside algorithmically-controlled co-performers, will they draw on the same mechanisms for creativity and coordination as they do when playing acoustic instruments with human co-performers?

5.3. Does Technology Facilitate or Constrain Creativity?

Creativity is widely valued in Western music traditions. In some other traditions, this is not the case, and performers are instead expected to replicate the ideal performance of a piece with as much precision as possible. It has been suggested that the preoccupation with creativity that exists in Western society is maintained by the commercial benefits of musicians distinguishing themselves with a personal identity (Clarke, 2012). Alongside the drive for creativity and individuality has come an upsurge in the number of technologies available for producing and hearing music. These have led to some marked changes in the way music is experienced, and could have either facilitatory or impairing effects on musical creativity.

For example, since audio recording of music performances became possible, more and more of the music that people hear is “disembodied,” comprising only audio, with no visual cues and no possibility of real-time performer-audience interaction. Today, most people have ready access to a vast collection of recordings from a wide range of musical styles. As a result, present-day musicians are exposed to far more musical ideas than would have been the case if they had been born in an era where the only access to music was via live performance. The potentially rich networks of musical knowledge that they have constructed in memory could facilitate their creative musical thinking by providing numerous possibilities for new associations to be made.

On the other hand, over-familiarity with popular interpretations or conventions could constrain either performers' abilities to consider more unusual ideas or listeners' willingness to accept more idiosyncratic performances (see Repp, 1997a). Today, music is everywhere—playing in the background while we work, shop, exercise, travel, and relax—and most people receive a great deal of passive (and often unsought) exposure to certain genres, which might affect their openness to new styles or interpretations. More broadly in the expertise literature, an inverted-U relationship is hypothesized to exist between

formal knowledge and creativity, with highly-knowledgeable people sometimes struggling to break away from established frameworks and generate novel ideas (Weisberg, 1999). Future research might investigate collaborative creativity in ensembles comprising professional musicians who are at different stages of their careers.

Some technologies, like music notation software, are designed to make the process of creating music easier and more generally available, including for people who wish to compose collaboratively. These programs usually convert MIDI information into musical scores, so musicians can compose at a keyboard without having to attend to notating their ideas. Alternatively, those who lack technical performance skills can enter notes using a mouse or computer keyboard and hear their ideas played back to them – the ability to play or audiate their own compositions is not necessary. It could be argued that while such programs do simplify the task of composing, they also constrain composers' creativity by minimizing their reliance on imagination and potentially impeding cognitive flexibility. On the other hand, notation software could be seen as a means of composers extending their own working memory capacity. Fewer resources spent maintaining a single idea in working memory means that more resources are available for elaborating on that idea or drawing new associations. Whether the net effect is enhanced or impaired creativity, however, requires some investigation.

Other technologies, like new digital musical interfaces (DMIs), broaden the range of sounds and sound-producing gestures that can be part of a music performance. In some cases, they also reduce the extent to which music performance depends on highly practiced technical skills, making them potential means of music-making for a large number of people. A critical difference between DMIs and traditional instruments concerns how directly gestures and sound output relate. For DMIs, gesture-sound relations are indirect – and sometimes complex: gestures activate electronic signals, which pass through several layers of algorithmic mappings before triggering sound output (Jensenius, 2013). As research has already shown, audience members are sometimes unable to make sense of complex gesture-sound mappings, and show little appreciation for the performed music as a result (Emerson and Egermann, 2017). Do ensemble members also struggle to make sense of each other's gestures when performing with DMIs? When their sound-producing gestures carry little communicative value, what other communication techniques do they use to ensure successful collaboration? Future research should consider whether different mechanisms support collaborative creativity in DMI and traditional instrument contexts.

6. CONCLUSIONS

Driving our discussion has been the question of how musicians coordinate their performance under conditions that encourage creativity. Despite extensive research into ensemble coordination mechanisms, the literature on music performance has largely avoided the topic of creativity, focusing instead on simplified

musical contexts that lack the ambiguity, unpredictability, and variety of real-world music. In recent years, however, the field of music cognition has seen increased interest in studies of music outside the Western classical repertoire e.g., (Freeman and van Troyer, 2011; Marandola, 2014; Clayton, 2017), which has prompted questions about how generalizable our current understanding of coordination processes may be. At the same time, theoretical perspectives have shifted away from treating cognition as individual and internal, moving instead toward embodiment and distributed cognition paradigms. A growing number of studies now focuses on constructs such as group flow, in many cases attempting to develop conceptual models based on investigation of performers' experiences.

Researchers have long shied away from the scientific study of creativity in music performance, presumably because the idea of artistic creativity seems ill-defined and difficult to quantify. I have not ventured into any discussion of how musical creativity or creative abilities should be evaluated, and would argue that the evaluation of creative output is a different issue from describing the underlying processes. The creative processes involved in ensemble performance can be probed objectively and systematically by investigating musician's real-time adaptability and flexibility, testing for differences in behavior or musical output between solo and ensemble playing conditions, monitoring the (multilevelled) audiovisual signals that pass between them, or measuring the patterns of leader-follow influence that come and go throughout a performance.

REFERENCES

- Agnew, M. (1922). The auditory imagery of great composers. *Psychol. Monogr.* 31, 279–287. doi: 10.1037/h0093171
- Aleman, A., Nieuwenstein, M., Bocker, K., and de Haan, E. (2000). Music training and mental imagery ability. *Neuropsychologia* 38, 1664–1668. doi: 10.1016/S0028-3932(00)00079-8
- Aleman, A., and Wout, M. (2004). Subvocalization in auditory-verbal imagery: Just a form of motor imagery? *Cogn. Process.* 5, 228–231. doi: 10.1007/s10339-004-0034-y
- An, D., Song, Y., and Carr, M. (2016). A comparison of two models of creativity: Divergent thinking and creative expert performance. *Person. Indiv. Differ.* 90, 78–84. doi: 10.1016/j.paid.2015.10.040
- Aucouturier, J. J., and Canonne, C. (2017). Musical friends and foes: The social cognition of affiliation and control in improvised interactions. *Cognition* 161, 94–108. doi: 10.1016/j.cognition.2017.01.019
- Babiloni, C., Buffo, P., Vecchio, F., Marzano, N., Del Percio, C., Spada, D., et al. (2011). Brains “in concert”: frontal oscillatory alpha rhythms and empathy in professional musicians. *Neuroimage* 60, 105–116. doi: 10.1016/j.neuroimage.2011.12.008
- Bach, P., Allami, B. K., Tucker, M., and Ellis, R. (2014). Planning-related motor processes underlie mental practice and imitation learning. *J. Exper. Psychol. Gen.* 143, 1277–1294. doi: 10.1037/a0035604
- Badino, L., D'Ausilio, A., Glowinski, D., Camurri, A., and Fadiga, L. (2014). Sensorimotor communication in professional quartets. *Neuropsychologia* 55, 98–104. doi: 10.1016/j.neuropsychologia.2013.11.012
- Bailes, F., and Bishop, L. (2012). *Musical Imagery in the Creative Process*. Farnham: Ashgate.
- Bailes, F., Dean, R. T., and Pearce, M. T. (2013). Music cognition as mental time travel. *Sci. Reports* 3, 1–4. doi: 10.1038/srep02690
- Barbot, B., and Tinio, P. P. L. (2015). Where is the “g” in creativity? a specialization - differentiation hypothesis. *Front. Hum. Neurosci.* 8:1041. doi: 10.3389/fnhum.2014.01041
- Bartlette, C., Headlam, D., Bocko, M., and Velick, G. (2006). Effect of network latency on interactive musical performance. *Music Percept.* 24, 49–62. doi: 10.1525/mp.2006.24.1.49
- Benedek, M., Jauk, E., Fink, A., Koschutnig, K., Reishofer, G., Ebner, F., and Neubauer, A. C. (2014). To create or to recall? Neural mechanisms underlying the generation of creative new ideas. *Neuroimage* 88, 125–133. doi: 10.1016/j.neuroimage.2013.11.021
- Bernardi, N. F., De Buglio, M., Trimarchi, P. D., Chielli, A., and Bricolo, E. (2013a). Mental practice promotes motor anticipation: Evidence from skilled music performance. *Front. Hum. Neurosci.* 7:451. doi: 10.3389/fnhum.2013.00451
- Bernardi, N., Schories, A., Jabusch, H. C., Colombo, B., and Altenmüller, E. (2013b). Mental practice in music memorization: An ecological-empirical study. *Music Percept.* 30, 275–290. doi: 10.1525/mp.2012.30.3.275
- Bishop, L., Bailes, F., and Dean, R. T. (2013). Musical imagery and the planning of dynamics and articulation during performance. *Music Percept.* 31, 97–116. doi: 10.1525/mp.2013.31.2.97
- Bishop, L., Bailes, F., and Dean, R. T. (2014). Performing musical dynamics: How crucial are musical imagery and auditory feedback for expert and novice musicians? *Music Percept.* 32, 51–66. doi: 10.1525/mp.2014.32.1.51
- Bishop, L., and Goebel, W. (2014). Context-specific effects of musical expertise on audiovisual integration. *Front. Cogn. Sci.* 5:1123. doi: 10.3389/fpsyg.2014.01123
- Bishop, L., and Goebel, W. (2015). When they listen and when they watch: Pianist's use of nonverbal audio and visual cues during duet performance. *Music. Sci.* 19, 84–110. doi: 10.1177/1029864915570355
- Bishop, L., and Goebel, W. (2017). “Mapping visual attention of ensemble musicians during performance of “temporally-ambiguous” music,” in *Conference on Music and Eye-Tracking* (Frankfurt).

I have highlighted some potential mechanisms for collaborative creativity, including musical imagery, which could facilitate performance flexibility and adaptability, and multilevelled reflective and prereflective communication processes that could help performers align their constantly-evolving intentions in real-time. I have also discussed the potential importance of empathy in facilitating perspective-taking and coordinating of emotional states. In future research, particular attention should be paid to demystifying concepts such as emergence and flow, perhaps through systematic study of how often they arise and how substantially they affect audience members' perceptions of a performance.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

FUNDING

This research was supported by Austrian Science Fund grant P29427.

ACKNOWLEDGMENTS

Many thanks to Assoc. Prof. Werner Goebel for insightful comments on an earlier version of this manuscript.

- Bishop, L., and Goebel, W. (2017, Online First). Communication for coordination: gesture kinematics affect coordination success in piano duos. *Psychol. Res.* doi: 10.1007/s00426-017-0893-3
- Bishop, L., and Goebel, W. (2018). Beating time: how ensemble musician's cueing gestures communicate beat position and tempo. *Psychol. Music* 46, 84–106. doi: 10.1177/0305735617702971
- Bittner, J. V., and Heidemeier, H. (2013). Competitive mindsets, creativity, and the role of regulatory focus. *Think. Skills Creativ.* 9, 59–68. doi: 10.1016/j.tsc.2013.03.003
- Brodsky, W., Henik, A., Rubinstein, B., and Zorman, M. (2003). Auditory imagery from musical notation in expert musicians. *Percept. Psychophys.* 65, 602–612. doi: 10.3758/BF03194586
- Bryan-Kinns, N. (2013). Mutual engagement and collocation with shared representations. *Int. J. Hum. Comput. Stud.* 71, 76–90. doi: 10.1016/j.ijhcs.2012.02.004
- Bryan-Kinns, N., and Hamilton, F. (2009). Identifying mutual engagement. *Behav. Inform. Technol.* 31, 101–125. doi: 10.1080/01449290903377103
- Canonne, C., and Aucouturier, J. (2015). Play together, think alike: Shared mental models in expert music improvisers. *Psychol. Music* 44, 544–558. doi: 10.1177/0305735615577406
- Canonne, C., and Garnier, N. (2015). Individual decisions and perceived form in collective free improvisation. *J. New Music Res.* 44, 145–167. doi: 10.1080/09298215.2015.1061564
- Chaffin, R., Lemieux, A. F., and Chen, C. (2006). *Spontaneity and Creativity in Highly Practised Performance*. Hove: Psychology.
- Chaffin, R., Lemieux, A. F., and Chen, C. (2007). “It is different each time I play”: Variability in highly prepared musical performance. *Music Percept.* 24, 455–472. doi: 10.1525/mp.2007.24.5.455
- Chaffin, R., Lisboa, T., Logan, T., and Begosh, K. T. (2010). Preparing for memorized cello performance: The role of performance cues. *Psychol. Music* 38, 3–30. doi: 10.1177/0305735608100377
- Chen, J., Penhune, V., and Zatorre, R. (2008). Listening to music rhythms recruits motor regions of the brain. *Cereb. Cortex* 18, 2844–2854. doi: 10.1093/cercor/bhn042
- Clarke, E. (1993). Imitating and evaluating real and transformed musical performances. *Music Percept.* 10, 317–341. doi: 10.2307/40285573
- Clarke, E. (2012). *Creativity in Performance*. Oxford: Oxford University.
- Clayton, M. (2017). *The Ethnography of Embodied Music Interaction*. London, UK: Routledge.
- Cochrane, T. (2017). *Group Flow*. London: Routledge.
- Cohen, S., and Bodner, E. (2018, Online First). The relationship between flow and music performance anxiety amongst professional classical orchestral musicians. *Psychol. Music*. doi: 10.1177/0305735618754689
- Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, NY: Harper Collins.
- Csikszentmihalyi, M. (1999). *Implications of a Systems Perspective for the Study of Creativity*. Cambridge: Cambridge University.
- Davidson, J. (2012). Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies. *Psychol. Music* 40, 595–633. doi: 10.1177/0305735612449896
- De Dreu, C. K., Baas, M., and Nijstad, B. A. (2008). Hedonic tone and activation level in the mood-creativity link: Toward a dual pathway to creativity model. *Attit. Soc. Cogn.* 94, 739–756. doi: 10.1037/0022-3514.94.5.739
- De Dreu, C. K., Nijstad, B. A., and Baas, M. (2011). Behavioral activation links to creativity because of increased cognitive flexibility. *Soc. Psychol. Person. Sci.* 2, 72–80. doi: 10.1177/1948550610381789
- De Dreu, C. K. W., Nijstad, B. A., Baas, M., Wolsink, I., and Roskes, M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Person. Soc. Psychol. Bull.* 38, 656–669. doi: 10.1177/0146167211435795
- Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychon. Bull. Rev.* 11, 1011–1026. doi: 10.3758/BF03196731
- Dowling, W. J., Tillmann, B., and Ayers, D. F. (2002). Memory and the experience of hearing music. *Music Percept.* 19, 249–276. doi: 10.1525/mp.2001.19.2.249
- Egermann, H., and McAdams, S. (2013). Empathy and emotional contagion as a link between recognized and felt emotions in music listening. *Music Percept.* 31, 139–156. doi: 10.1525/mp.2013.31.2.139
- Eitan, Z., and Timmers, R. (2009). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition* 114, 405–422. doi: 10.1016/j.cognition.2009.10.013
- Emerson, G., and Egermann, H. (2017). Gesture-sound causality from the audience's perspective: Investigating the influence of mapping perceptibility on the aesthetic perception of new digital musical instruments. *Psychol. Aesthet. Creativ. Arts* 12, 96–109. doi: 10.1037/aca0000114
- Ericsson, K. A. (1998). The scientific study of expert levels of performance: general implications for optimal learning and creativity. *High Abil. Stud.* 9, 75–100. doi: 10.1080/1359813980090106
- Finke, R. A., Pinker, S., and Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cogn. Sci.* 13, 51–78. doi: 10.1207/s15516709cog1301_2
- Fischer, G., Giaccardi, E., Eden, H., Sugimoto, M., and Ye, Y. (2005). Beyond binary choices: integrating individual and social creativity. *Int. J. Hum. Comp. Stud.* 63, 482–512. doi: 10.1016/j.ijhcs.2005.04.014
- Floridou, G. A., Williamson, V. J., and Stewart, L. (2017). A novel indirect method for capturing involuntary musical imagery under varying cognitive load. *Q. J. Exper. Psychol.* 70, 2189–2199. doi: 10.1080/17470218.2016.1227860
- Foster, N. E. V., Halpern, A., and Zatorre, R. (2013). Common parietal activation in musical mental transformations across pitch and time. *Neuroimage* 75, 27–35. doi: 10.1016/j.neuroimage.2013.02.044
- Freeman, J., and van Troyer, A. (2011). Collaborative textual improvisation in a laptop ensemble. *Comp. Music J.* 35, 8–21. doi: 10.1162/COMJ_a_00053
- Gaggioli, A., Chirico, A., Mazzoni, E., Milani, L., and Riva, G. (2017). Networked flow in musical bands. *Psychol. Music* 45, 283–297. doi: 10.1177/0305735616665003
- Gaggioli, A., Riva, G., Milani, L., and Mazzoni, E. (2013). *Networked Flow: Towards an understanding of Creative Networks*. Dordrecht: Springer.
- Glowinski, D., Bracco, F., Chiorri, C., and Grandjean, D. (2016). Music ensemble as a resilient system: Managing the unexpected through group interaction. *Front. Psychol.* 7:1548. doi: 10.3389/fpsyg.2016.01548
- Glăveanu, V. P. (2013). Rewriting the language of creativity: The Five A's Framework. *Rev. Gen. Psychol.* 17, 69–81. doi: 10.1037/a0029528
- Goebel, W., and Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Percept.* 26, 427–438. doi: 10.1525/mp.2009.26.5.427
- Gonen-Yaacovi, G., Cruz de Souza, L., Levy, R., Urbanski, M., Josse, G., and Volle, E. (2013). Rostral and caudal prefrontal contribution to creativity: A meta-analysis of functional imaging data. *Front. Hum. Neurosci.* 7:465. doi: 10.3389/fnhum.2013.00465
- Halpern, A. R., Zatorre, R. J., Bouffard, M., and Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia* 42, 1281–1292. doi: 10.1016/j.neuropsychologia.2003.12.017
- Hargreaves, D. J. (2012). *Imagination and Creativity in Music Listening*. Oxford: Oxford University.
- Hart, E., and Di Blasi, Z. (2015). Combined flow in musical jam sessions: a pilot qualitative study. *Psychol. Music* 43, 275–290. doi: 10.1177/0305735613502374
- Hart, Y., Noy, L., Feniger-Schaal, R., Mayo, A. E., and Alon, U. (2014). Individuality and togetherness in joint improvised motion. *PLoS ONE* 9:e87213. doi: 10.1371/journal.pone.0087213
- Hill, J. (2012). *Imagining Creativity: An Ethnomusicological Perspective on How Belief Systems Encourage or Inhibit Creative Activities in Music*. Oxford: Oxford University.
- Hove, M. J., and Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Soc. Cogn.* 27, 949–961. doi: 10.1521/soco.2009.27.6.949
- Hutchins, E. (1990). *The Technology of Team Navigation*. Hillsdale, NJ: Lawrence Erlbaum.
- Jakubowski, K., Farrugia, N., and Stewart, L. (2016). Probing imagined tempo for music: effects of motor engagement and musical experience. *Psychol. Music* 44, 1274–1288. doi: 10.1177/0305735615625791
- Janata, P., and Paroo, K. (2006). Acuity of auditory images in pitch and time. *Percept. Psychophys.* 68, 829–844. doi: 10.3758/BF03193705
- Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behav. Brain Res.* 142, 1–15. doi: 10.1016/S0166-4328(02)00384-4
- Jensenius, A. R. (2013). An action-sound approach to teaching interactive music. *Organ. Sound* 18, 178–189. doi: 10.1017/S1355771813000095

- Juslin, P., and Laukka, P. (2004). Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. *J. New Music Res.* 33, 217–238. doi: 10.1080/0929821042000317813
- Kawase, S. (2014). Assignment of leadership role changes performer's gaze during piano duo performances. *Ecol. Psychol.* 26, 198–215. doi: 10.1080/10407413.2014.929477
- Keller, J., Bless, H., Blomann, F., and Kleinböhl, D. (2011). Physiological aspects of flow experiences: Skills-demand-compatibility effects on heart rate variability and salivary cortisol. *J. Exp. Soc. Psychol.* 47, 849–852. doi: 10.1016/j.jesp.2011.02.004
- Keller, P. E. (2001). Attentional resource allocation in musical ensemble performance. *Psychol. Music* 29, 20–38. doi: 10.1177/0305735601291003
- Keller, P. E., and Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Percept.* 28, 27–46. doi: 10.1525/mp.2010.28.1.27
- Keller, P. E., Dalla Bella, S., and Koch, I. (2010). Auditory imagery shapes movement timing and kinematics: evidence from a musical task. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 508–513. doi: 10.1037/a0017604
- Keller, P. E., Knoblich, G., and Repp, B. (2007). Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Conscious. Cogn.* 16, 102–111. doi: 10.1016/j.concog.2005.12.004
- Krueger, J. (2010). Extended cognition and the space of social interaction. *Conscious. Cogn.* 20, 643–657. doi: 10.1016/j.concog.2010.09.022
- Large, E. W., Palmer, C., and Pollack, J. B. (1995). Reduced memory representations for music. *Cogn. Sci.* 19, 53–93. doi: 10.1207/s15516709cog1901_2
- Launay, J. (2015). Musical sounds, motor resonance, and detectable agency. *Emp. Musicol. Rev.* 10, 30–40. doi: 10.18061/emr.v10i1-2.4579
- Lee, H., and Noppeney, U. (2014). Temporal prediction errors in visual and auditory cortices. *Curr. Biol.* 24, R309–R310. doi: 10.1016/j.cub.2014.02.007
- Leman, M. (2001). *Modeling Musical Imagery in a Framework of Perceptually Constrained Spatio-Temporal Representations*. New York, NY: Taylor & Francis.
- Leman, M. (2012). “Musical gestures and embodied cognition,” in *Journées d'Informatique Musicale*, eds T. Dutoit, T. Todoroff, and N. d'Alessandro (Mons: University of Mons), 5–7.
- Leman, M., and Maes, P. J. (2014). The role of embodiment in the perception of music. *Emp. Musicol. Rev.* 9, 236–246. doi: 10.18061/emr.v9i3-4.4498
- Linson, A., and Clarke, E. F. (2017). *Distributed Cognition, Ecological Theory and Group Improvisation*. New York, NY: Oxford University.
- Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., and Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *J. Cogn. Neurosci.* 25, 1049–1061. doi: 10.1162/jocn_a_00388
- Loehr, J. D., and Palmer, C. (2011). Temporal coordination between performing musicians. *Q. J. Exp. Psychol.* 64, 2153–2167. doi: 10.1080/17470218.2011.603427
- Loehr, J. D., and Vesper, C. (2016). The sound of you and me: novices represent shared goals in joint action. *Q. J. Exp. Psychol.* 69, 535–547. doi: 10.1080/17470218.2015.1061029
- Lucas, B. J., Schubert, E., and Halpern, A. R. (2010). Perception of emotion in sounded and imagined music. *Music Percept.* 27, 399–412. doi: 10.1525/mp.2010.27.5.399
- Luck, G., and Nte, S. (2008). An investigation of conductor's temporal gestures and conductor-musician synchronization, and a first experiment. *Psychol. Music* 36, 81–99. doi: 10.1177/0305735607080832
- Lundqvist, L. O., Carlsson, F., Hilmersson, P., and Juslin, P. (2009). Emotional responses to music: experience, expression, and physiology. *Psychol. Music* 37, 61–90. doi: 10.1177/0305735607086048
- MacRitchie, J., Varlet, M., and Keller, P. E. (2017). *Embodied Expression Through Entrainment and Co-representation in Musical Ensemble Performance*. New York, NY: Routledge.
- Maes, P. J. (2016). Sensorimotor grounding of musical embodiment and the role of prediction: a review. *Front. Psychol.* 7:308. doi: 10.3389/fpsyg.2016.00308
- Maes, P. J., Leman, M., Palmer, C., and Wanderley, M. M. (2014). Action-based effects on music perception. *Front. Psychol.* 4:1008. doi: 10.3389/fpsyg.2013.01008
- Marandola, F. (2014). *Expressiveness in the Performance of Bedzan Pygmy's Vocal Polyphonies: When the Same is Never the Same*. Oxford: Oxford University.
- Molnar-Szakacs, I., and Overy, K. (2006). Music and mirror neurons: from motion to 'emotion. *Soc. Cogn. Affect. Neurosci.* 1, 235–241. doi: 10.1093/scan/nsl029
- Moran, N. (2014). Social implications arise in embodied music cognition research which can counter musicological “individualism”. *Front. Cognit. Sci.* 5:676. doi: 10.3389/fpsyg.2014.00676
- Moran, N., Hadley, L. V., Bader, M., and Keller, P. E. (2015). Perception of ‘back-channeling’ nonverbal feedback in musical duo improvisation. *PLoS ONE* 10:e0130070. doi: 10.1371/journal.pone.0130070
- Morgan, E., Gunes, H., and Bryan-Kinns, N. (2015). Using affective and behavioural sensors to explore aspects of collaborative music making. *Int. J. Hum. Comput. Stud.* 82, 31–47. doi: 10.1016/j.ijhcs.2015.05.002
- Müllensiefen, D., Fry, J., Jones, R., Jilka, S., Stewart, L., and Williamson, V. J. (2014). Individual differences predict patterns in spontaneous involuntary musical imagery. *Music Percept.* 31, 323–338. doi: 10.1525/mp.2014.31.4.323
- Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., and Baas, M. (2010). The dual pathway to creativity model: creative ideation as a function of flexibility and persistence. *Eur. Rev. Soc. Psychol.* 21, 34–77. doi: 10.1080/10463281003765323
- Novembre, G., Ticini, L., Schtz-Bosbach, S., and Keller, P. (2012). Distinguishing self and other in joint action: evidence from a musical paradigm. *Cereb. Cortex* 22, 2894–2903. doi: 10.1093/cercor/bhr364
- Noy, L., Dekel, E., and Alon, U. (2011). The mirror game as a paradigm for studying the dynamics of two people improvising motion together. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20947–20952. doi: 10.1073/pnas.1108155108
- Noy, L., Levit-Binun, N., and Golland, Y. (2015). Being in the zone: physiological markers of togetherness in joint improvisation. *Front. Hum. Neurosci.* 9:187. doi: 10.3389/fnhum.2015.00187
- Olsen, K. N., and Dean, R. T. (2016). Does perceived exertion influence perceived affect in response to music? Investigating the “FEELA” hypothesis. *Psychomusicol. Music Mind Brain* 26, 257–269. doi: 10.1037/pmu0000140
- Pachet, F., Roy, P., and Foulon, R. (2017). *Do Jazz Improvisers Really Interact?* New York, NY: Routledge.
- Patel, A. D., and Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front. Syst. Neurosci.* 8:57. doi: 10.3389/fnsys.2014.00057
- Pearce, E., Launay, J., van Duijn, M., Rotkirch, A., David-Barrett, T., and Dunbar, R. I. M. (2016). Singing together or apart: the effect of competitive and cooperative singing on social bonding within and between sub-groups of a university fraternity. *Psychol. Music* 44, 1255–1273. doi: 10.1177/0305735616636208
- Peifer, C., Schulz, A., Schächinger, H., Baumann, N., and Antoni, C. H. (2014). The relation of flow-experience and physiological arousal under stress – can u shape it? *J. Exp. Soc. Psychol.* 52, 62–69. doi: 10.1016/j.jesp.2014.01.009
- Petrini, K., Russell, M., and Pollick, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition* 110, 432–439. doi: 10.1016/j.cognition.2008.11.015
- Prinz, W. (1990). *A Common Coding Approach to Perception and Action*. Berlin/Heidelberg: Springer.
- Ragert, M., Schroeder, T., and Keller, P. E. (2013). Knowing too little or too much: the effects of familiarity with a co-performer's part on interpersonal coordination in musical ensembles. *Front. Audit. Cogn. Neurosci.* 4:368. doi: 10.3389/fpsyg.2013.00368
- Repp, B. (1997a). The aesthetic quality of a quantitatively average music performance: two preliminary experiments. *Music Percept.* 14, 419–444. doi: 10.2307/40285732
- Repp, B. (1997b). Expressive timing in a Debussy Prelude: a comparison of student and expert pianists. *Music. Sci.* 1, 257–268. doi: 10.1177/102986499700100206
- Repp, B. H., and Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. Bull. Rev.* 20, 403–452. doi: 10.3758/s13423-012-0371-2
- Rhodes, M. (1961). An analysis of creativity. *Phi Delta Kappan* 42, 305–310.
- Sawyer, R. K. (2006). Group creativity: musical performance and collaboration. *Psychol. Music* 34, 148–165. doi: 10.1177/0305735606061850
- Schiavio, A., and Høffding, S. (2015). Playing together without communicating? A pre-reflective and enactive account of joint musical performance. *Music. Sci.* 19, 366–388. doi: 10.1177/1029864915593333

- Schubert, E. (2012). *Spreading Activation and Dissociation: A Cognitive Mechanism for Creative Processing in Music*. Oxford: Oxford University.
- Seddon, F. (2005). Modes of communication during jazz improvisation. *Br. J. Music Educ.* 22, 47–61. doi: 10.1017/S0265051704005984
- Seddon, F., and Biasutti, M. (2009). A comparison of modes of communication between members of a string quartet and a jazz sextet. *Psychol. Music* 37, 395–415. doi: 10.1177/0305735608100375
- Tarr, B., Launay, J., and Dunbar, R. I. M. (2014). Music and social bonding: “self-other” merging and neurohormonal mechanisms. *Front. Psychol.* 5:1096. doi: 10.3389/fpsyg.2014.01096
- Toiviainen, P., Luck, G., and Thompson, M. R. (2010). Embodied meter: hierarchical eigenmodes in music-induced movement. *Music Percept.* 28, 59–70. doi: 10.1525/mp.2010.28.1.59
- Trusheim, W. H. (1993). Audiation and mental imagery: implications for artistic performance. *Q. J. Music Teach. Learn.* 2, 139–147.
- van der Wel, R. P. R. D., Sebanz, N., and Knoblich, G. (2016). “A joint action perspective on embodiment,” in *Conceptual and Interactive Embodiment: Foundations of Embodied Cognition, Vol. 2*, eds M. H. Fischer and Y. Coello (Routledge, NY), 165–181.
- Vuvan, D. T., Podolak, O. M., and Schmuckler, M. A. (2014). Memory for musical tones: the impact of tonality and the creation of false memories. *Front. Psychol.* 5:582. doi: 10.3389/fpsyg.2014.00582
- Walton, A. E., Washburn, A., Langland-Hassan, P., Chemero, A., Kloos, H., and Richardson, M. J. (2017). Creating time: social collaboration in music improvisation. *Topics Cogn. Sci.* 10, 95–119. doi: 10.1111/tops.12306
- Weisberg, R. W. (1999). *Creativity and Knowledge: A Challenge to Theories*. Cambridge: Cambridge University.
- Williamon, A., and Davidson, J. W. (2002). Exploring co-performer communication. *Music. Sci.* 6, 53–72. doi: 10.1177/102986490200600103
- Williamon, A., Thompson, S., Lisboa, T., and Wiffen, C. (2006). *Creativity, Originality, and Value in Music Performance*. Hove: Psychology.
- Wolf, W., Launay, J., and Dunbar, R. I. M. (2015). Joint attention, shared goals, and social bonding. *Br. J. Psychol.* 107, 322–337. doi: 10.1111/bjop.12144
- Wöllner, C., and Cañal-Bruland, R. (2010). Keeping an eye on the violinist: motor experts show superior timing consistency in a visual perception task. *Psychol. Res.* 74, 579–585. doi: 10.1007/s00426-010-0280-9
- Wong, S. S., and Lim, S. W. (2017). Mental imagery boosts music compositional creativity. *PLoS ONE* 12:e0174009. doi: 10.1371/journal.pone.0174009
- Wu, J., Yu, Z., Mai, X., Wei, J., and Luo, Y. (2011). Pitch and loudness information encoded in auditory imagery as revealed by event related potentials. *Psychophysiology* 48, 415–419. doi: 10.1111/j.1469-8986.2010.01070.x
- Zamm, A., Pfordresher, P., and Palmer, C. (2014). Temporal coordination in joint music performance: effects of endogenous rhythms and auditory feedback. *Exper. Brain Res.* 233, 607–618. doi: 10.1007/s00221-014-4140-5

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 Bishop. 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.



Learning Choreography: An Investigation of Motor Imagery, Attentional Effort, and Expertise in Modern Dance

Katy Carey, Aidan Moran* and Brendan Rooney

School of Psychology, University College Dublin, Dublin, Ireland

OPEN ACCESS

Edited by:

Philip A. Fine,
University of Buckingham,
United Kingdom

Reviewed by:

Emma Redding,
Trinity Laban Conservatoire of Music
and Dance, United Kingdom
Matthew Woolhouse,
McMaster University, Canada

*Correspondence:

Aidan Moran
Aidan.Moran@ucd.ie

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 26 March 2018

Accepted: 12 February 2019

Published: 01 March 2019

Citation:

Carey K, Moran A and Rooney B
(2019) Learning Choreography: An
Investigation of Motor Imagery,
Attentional Effort, and Expertise
in Modern Dance.
Front. Psychol. 10:422.
doi: 10.3389/fpsyg.2019.00422

The study of choreography in dance offers researchers an intriguing window on the relationship between expertise, imagination, and attention in the creative process of learning new movements. The present study investigated an unresolved issue in this field – namely, the effects of expertise on motor imagery (MI; or the mental rehearsal of actions without engaging in the actual movements involved) and attentional effort (as measured by pupil dilation) on dancers while they engaged in the processes of learning, performing, and imagining a dance movement. Participants were 18 female dancers (mean age = 23, $SD = 5.85$) comprising three experience levels (i.e., novice, intermediate and expert performers) in this field. Data comprised these participants' MI scores as well as their pupil dilation while they learned, performed, and imagined a 15 s piece of choreography. In addition, the time taken both to perform and to imagine the choreography were recorded. Results showed no significant effect of dance expertise on MI but some differences between beginners and intermediate dancers in attentional effort (pupil dilation) at the start of the performance and the imagined movement conditions. Specifically, the beginners had the highest pupil dilation, with the experts having the second highest, while intermediates had the lowest dilation. Further analysis suggested that the novice dancers' pupil dilation at the start of the performance may have been caused, in part, by the initial mental effort required to assess the cognitive demands of the dance task.

Keywords: dance, creativity, expertise, motor imagery, attention, pupillometry

INTRODUCTION

Dance is a form of artistic expression and communication involving “moving the body through time and space” (Cross and Ticini, 2012, p. 6). It is a cognitively and physically demanding art-form which elicits creativity in the dancer, who is required to be able to adapt movements that are rhythmical and esthetically pleasing (Kaufman and Baer, 2005). Since the early 2000s, it has attracted research attention from psychologists and neuroscientists because it provides a “real life” window into topics like expertise (the study of what makes people exceptionally knowledgeable about, or skilled in, a particular domain; Moran and Toner, 2017), embodied cognition (the theory that cognition is largely grounded in sensorimotor experience; Laakso, 2011) and creativity (the capacity to produce ideas and outputs that are novel and adaptive or functional;

Simonton and Damian, 2013). Dance research has addressed both theoretical and practical issues. For example, at a theoretical level, Cross et al. (2014) showed how the neuroscientific study of dance can elucidate the mechanisms by which the brain perceives and learns complex motor sequences. In addition, research on dance facilitates the study of inter-genre differences in creativity among performers. For example, Fink and Woschnjak (2011) discovered that experienced modern contemporary dancers have heightened figural and verbal creative abilities in comparison to dancers of other genres (such as ballet and jazz) as well as non-dancers. At an applied level, research on dancers' mental rehearsal techniques has provided fascinating insights into the cognitive process of "motor imagery" (MI) or "mentally simulating an intended action without actually producing it" (Smith and Kosslyn, 2007, p. 456). For example, Nordin and Cumming (2005) conducted in-depth interviews with professional dancers to find out where, when and why they used MI. One of their findings was that dancers reported using imagery in practice both as a "creative tool" (p. 401) and to help them to learn and remember steps. Furthermore, Kaufman and Baer (2005) argued that dancers are inherently creative due to the constant decision making that they require when improvising or creating their own choreography and also when learning and performing movements. When learning a new movement (i.e., one that is not in their behavioral repertoire), dancers are confronted with a problem. In such situations, according to Weisberg's (2018) "expertise view" of creativity, the "presentation of a problem results in retrieval of knowledge – i.e. expertise – from memory; *creative advances evolve out of attempts to apply that knowledge to the new situation*" (p. 813; italics ours). Interestingly, Kaufman and Baer (2005) also postulated that a creative dancer is one who can utilize MI to achieve a heightened awareness of performance, as images can incorporate physicality, emotion and expressiveness, which are three key components of a dance performance.

Unfortunately, despite the preceding research, little or nothing is known at present about the relationship between dancers' expertise and their use of cognitive processes such as MI and attention (or "focusing on specific features, objects or locations or on certain thoughts or activities"; Goldstein, 2011, p. 391) when attempting to master new choreography. According to Torrents et al. (2015), learning choreography elicits "possibility thinking" in dancers – a creative process that begins with artistic performers imaginatively asking "what if?" before they proceed to execute a novel action. Similarly, Kaufman and Baer (2005) state that throughout a performance, the dancer faces moment-by-moment decisions such as how and when to execute movements, making it an inherently creative process. Against this background, and in view of the dearth of research on cognitive psychological aspects of dance, the purpose of the present paper is to investigate the relationship between MI and attentional effort (or the allocation of mental resources to satisfy cognitive demands; Sarter et al., 2006) in dancers of differing expertise who engage in this possibility thinking while learning and performing choreography. Before explaining MI and attentional effort in more detail, however, it is important to understand the methodological approach that we have adopted in the present paper: namely, "process tracing" – a term borrowed

from Williams and Ericsson (2005) to refer to procedures (e.g., eye tracking technology or the computerized measurement of the location, duration and sequence of people's visual fixations when they inspect a given scene) that help to identify the processes or mechanisms underlying expert performance in a given field. As Ericsson (2018) put it, process tracing "is essential for uncovering detailed information about most of the important characteristics that are responsible for the superiority of ... experts' achievements" (p. 207). To the best of our knowledge, no previous published study has used this "process tracing" approach to investigate the learning, performing and mental simulation of choreographed dance movements and the potential creative thinking which underlies these processes.

Exploring Creativity in Dance – Toward a Process Tracing Approach

In order to explain the "process tracing" approach to the study of dance, we need to consider how creativity is approached in psychology. According to Simonton and Damian (2013), creativity can be studied from at least three different perspectives in psychology: namely, those of products, persons, or processes. Research on creative *products* focuses mainly on the "ideational development" that spawns creative outputs such as poems or paintings. Next, research on creative *persons* typically explores either how the originators concerned manage to acquire relevant domain-specific expertise or how their cognitive abilities (e.g., divergent thinking skills) and inclinations (e.g., cognitive style) facilitate the outputs under scrutiny. Finally, research on creative *processes* is largely concerned with identifying and tracing the neurocognitive mechanisms that are postulated to mediate creative thought or action. In this latter regard, a variety of methodological tools is available for this type of "process tracing" of psychological mechanisms. For example, electroencephalography (EEG; a technique that measures cortical activity by recording electrical signals generated by the brain using non-invasive electrodes placed at different points on the scalp in an elastic cap) has been used to explore the neural signature of creativity in dance. Thus, Fink et al. (2009) compared alpha-wave activity in expert professional dancers with that of a group of relative novices. Some notable differences were evident. For example, during a creative improvisation dance task, the professional performers displayed more right-hemispheric alpha synchronization in posterior parietal regions than did the novices. Alpha frequency EEG activity appears to be especially sensitive to creativity-related cognitive demands. Thus, synchronization of alpha has been observed to be stronger in response to tasks of creative thinking (such as generating unusual uses of everyday objects) in comparison with tasks requiring more convergent thinking (Fink et al., 2007). Interestingly, synchronization of alpha wave activity has been shown to increase as a result of creative thinking training (Fink et al., 2006). But do conventional teaching methods actually encourage a dancer's creative skills? Doubts about this issue were raised by Chappell et al. (2009) who found that because of increasing pressure on dancers to reach prescribed levels of attainment, certain formulaic styles of teaching,

and choreography have become popular in dance education. Unfortunately, these teaching styles may hinder a dancer's ability to generate movement solutions to the motor problems confronting them. To circumvent such difficulties, researchers have investigated the efficacy of novel teaching approaches on creativity in dance. For example, Torrents et al. (2015) discovered that when specific constraints were deliberately placed on dancers while improvising (e.g., by requiring them to keep one body-part in a designated position), the originality of their subsequent movement (as assessed by expert performers/choreographers) actually improved. Augmenting these studies, evidence has emerged to show that dance learners' mental imagery processes can facilitate the creative process of acquiring, or learning, new movements (Nordin and Cumming, 2007; Overby and Dunn, 2011; Heiland and Rovetti, 2013). This discovery leads us to consider the key variables in the present study – namely, MI and attentional effort.

Motor Imagery: Nature, Measurement, and Mechanisms

As mentioned earlier, motor imagery (MI: also known as “motor imagination”; Hanakawa, 2016) is the cognitive simulation of an action without actually executing it (see review by Moran et al., 2012). Research interest in MI is as old as the discipline of psychology itself. To illustrate, James (1890), in his prescient discussion of “motor images” (p. 708), suggested somewhat counter-intuitively that by anticipating experiences imaginatively, people actually learn to skate in the *summer* and to swim in the *winter*. Since the 1890s, hundreds of experimental studies have demonstrated the efficacy of MIP in improving skill-learning in a variety of performance domains (Moran et al., 2012). MI can be assessed using either subjective or objective measures. Whereas the former measures include psychometric instruments that require respondents to rate some aspect of their imagery experience (e.g., its vividness or clarity), the latter assess proficiency in imagery skills through the accuracy or speed with which respondents solve problems or complete tasks known to require imagery ability. A recent subjective measure of MI is the Movement Imagery Questionnaire-3 (MIQ-3; Williams et al., 2012) – which is an updated version of the movement imagery questionnaire (MIQ; Hall and Pongrac, 1983). The MIQ-3 is a 12-item questionnaire that assesses the ease or difficulty of generating images of four different movements (i.e., knee lift, jump, arm movement, and waist bend) from different imagery perspectives. For each item, participants are required to read a description of the movement, physically perform the movement, and then imagine that movement from the designated perspective. Respondents are then required to rate the resultant image on a 7-point Likert scale ranging from 1 (*very hard to see/feel*) to 7 (*very easy to see/feel*). Subscale scores range from 4 to 28 and higher scores reflect stronger imagery ability. According to its developers, the MIQ-3 displays good internal consistency. Turning to objective measures of MI, two main options are available at present. On the one hand, Madan and Singhal (2013, 2014) developed the test of ability in movement imagery (TAMI) which requires respondents to imagine a series

of bodily movements and then to select the correct option from a set of possible body-positioning images – including the appropriate one. Alternatively, MI can be measured objectively by comparing the time required to execute and imagine specific actions. To explain the rationale for this approach, if imagined and executed actions rely on similar motor representations and activate some common brain areas (as predicted by the “functional equivalence” hypothesis; discussed below), then their *temporal* organization should be equivalent. Accordingly, there should be a close correspondence between the time required to *mentally* perform a given action and that required for its *actual* execution. So, “mental chronometry” tasks measure MI by evaluating the correspondence between the actual and imagined duration required to perform a given action (see review by Guillot and Collet, 2005). Collet et al. (2011) also discussed factors which may mediate the correspondence between real and imagined movements, such as level of experience with the movement in question or the type of image the individual has generated, e.g., visual versus kinaesthetic.

Although there is a dearth of studies evaluating MIP programs in dancers (see Abraham et al., 2017), a growing research literature exists on other aspects of dance imagery (see reviews by Overby and Dunn, 2011; Pavlik and Nordin-Bates, 2016; Fisher, 2017). For example, Pavlik and Nordin-Bates (2016) reviewed 43 papers on dance imagery that had been published between 1990 and 2014. They concluded that dancers tend to use “technique imagery” (or mental rehearsal of movements or sequences) more frequently than other types of imagery – especially “to picture spatial relationships while simultaneously stimulating creativity and helping to plan the next steps” (p. 56). Interestingly, choreographers often use imagery to solve problems within a dance piece (Nordin and Cumming, 2005). In addition, Pavlik and Nordin-Bates (2016) concluded that dancers tend to use imagery before, during and after class, rehearsal, and performance. Other studies have examined the imagery abilities of dancers. Thus, Overby (1990) reported that experienced dancers tend to have stronger imagery abilities than novices – but, curiously, not for MI (as assessed by the Movement Imagery Questionnaire, MIQ; Hall and Pongrac, 1983). A possible explanation for this anomaly is that the MIQ is not an objective measure of MI. Subsequently, Jola and Mast (2005) compared the imagery abilities of dancers with those of non-dancers. Results showed that whereas dancers performed better than non-dancers on tests of imagined bodily rotation, they performed *worse* than non-dancers on tests assessing the rotation of inanimate objects – suggesting that dancers' imagery superiority may be domain-specific. Interestingly, Jola and Mast's (2005) study appears to be the only one in the dance imagery literature which used an objective measure of imagery (specifically, Shepard and Metzler's 1988, mental rotation test). Clearly, therefore, there is an urgent need for dance research on MI to combine subjective and objective measures – as we have done in the present paper.

Before we conclude this section, however, it is important to consider the possible theoretical mechanisms by which MI works. Perhaps the most influential account of these mechanisms is that offered by motor simulation theory

(MST; Jeannerod, 1994, 2001, 2006). According to MST (see critique by O'Shea and Moran, 2017), action planning and MI share a common mental representation. In other words, MI is based on the motor representation that underlies actual motor performance. Next, MST proposes that the motor system is part of a cognitive network that includes other psychological activities such as imagining actions, learning by observation, and attempting to understand the behavior of other people. Thirdly, Jeannerod (2001) claimed that actions involve a *covert* stage during which they are prepared or simulated mentally. This covert stage involves “a representation of the future, which includes the goal of the action, the means to reach it, and its consequences on the organism and the external world. Covert and overt stages thus represent a continuum, such that every overtly executed action implies the existence of a covert stage” (p. S103). Finally, combining these propositions, Jeannerod (2001) postulated that “MI ... should involve, in the subject's motor brain, neural mechanisms similar to those operating during the real action” (pp. S103-S104) – the so-called “functional equivalence” hypothesis. According to this hypothesis, imagined and executed actions share, to some degree, certain mental representations and underlying mechanisms (see brief review in Moran et al., 2012). For example, both overt and imagined actions share a motor representation of an intention to act. Whereas this intention is converted into an actual physical movement in the case of overt actions, it is inhibited in the case of imagined actions. Nevertheless, this shared motor representation facilitates certain forms of functional equivalence between actual and imagined actions. Thus, Héту et al. (2013) found that the neural network underlying MI includes several cortical regions known to control actual motor execution, such as the premotor cortex, parietal cortex and fronto-parietal regions such as the basal ganglia, putamen and pallidum. Having examined the nature and measurement of MI, and some of its key neurocognitive mechanisms, let us now turn to the second important variable in the present study – attentional effort.

Attentional Processes in Dance: Attentional Effort

The construct of attention has been invoked by cognitive psychologists for over a century to account for a range of mental phenomena such as selectivity of information processing, intensity of focus, and the allocation of limited mental resources to regulate concurrent task performance. Within attentional research, it has long been known that expert performance in any skilled domain depends significantly on the ability to focus selectively on task-relevant information (Moran, 1996). But apart from selectivity of information processing, another attentional process that seems crucial to skill learning is “attentional effort” (also known as “mental effort” or “cognitive effort”; Piquado et al., 2010; Burge et al., 2013). This rather loosely defined, if intuitively appealing, construct denotes the allocation of mental resources in order to satisfy task demands. For example, trying to multiply 36 by 49 in one's head requires more cognitive exertion than does multiplying 6 by 9. So, attentional effort captures the *intensive*, as distinct from the *selective*, nature of cognitive resource allocation.

To explain this distinction, Kahneman (1973) differentiated between “selective” and “intensive” aspects of attention. Whereas “selective” attention refers to the fact that we can assimilate only a fraction of all information available to us, “intensive” attention refers to the *intensity* with which one's attention is focused in a particular situation. For Kahneman (1973), therefore, “the intensive aspect of attention corresponds to effort” (p. 12).

One way of assessing attentional effort is through “pupillometry” – or the measurement of task-evoked changes in the diameter of the pupil of the eye as a function of cognitive processing (Mathôt and Van der Stigchel, 2015). To explain, pupil size changes in response to three different kinds of stimuli (Mathôt, 2018). Specifically, it constricts in response to brightness, constricts in response to near fixation, and dilates in response to increased cognitive activity, such as increased levels of arousal or mental effort. For example, Hess and Polt (1964) showed that pupil size is a reliable indicator of mental effort and arousal. They asked participants to perform mental calculations of varying complexity (e.g., 7×8 was deemed easy, whereas 16×23 was regarded as difficult) and discovered that pupil size reflected the difficulty of the calculation. The harder the calculation was to perform, the larger the pupil. Although space limitations preclude a review of research on pupillometry (but see Mathôt and Van der Stigchel, 2015), pupil dilation effects have been demonstrated reliably for cognitive tasks involving multiplication problems (Hess and Polt, 1964), visual search (Porter et al., 2007), and change detection (Unsworth and Robison, 2015) tasks. Furthermore, mounting evidence suggests that the pupil remains dilated throughout the expenditure of cognitive load (Granhölm et al., 1996). Unfortunately, apart from studies by Moran et al. (2016) and O'Shea and Moran (2016), pupillometry has rarely been investigated in sport, exercise and performance psychology despite its potential importance as a non-invasive, online measure of attentional effort. Clearly, as, Beatty and Lucero-Wagoner (2000) claimed, whatever activates the mind causes the pupil to dilate. According to Kahneman (1973), pupil dilation is “the best single index” (p. 18) of attentional effort. Supporting this view, recent evidence (e.g., Murphy et al., 2014) shows that pupil size predicts brain activity in the locus coeruleus-norepinephrine (LC-NE) system – the one that regulates the allocation of attentional resources to task engagement.

Some previous researchers have investigated attentional factors in dance. For example, Guss-West and Wulf (2016) surveyed a sample of expert ballet dancers to determine their preferred attentional focus while performing certain dance movements (e.g., *a pirouette en dehors*). Results showed that the dancers reported adopting either internal foci or a combination of internal and external foci most of the time when performing. Unfortunately, as this study relied on self-report data rather than objective measures, its results are limited to *perceived* rather than actual attentional processes. A different approach was adopted by Stevens et al. (2010) who used eye-tracking equipment to explore expert-novice differences in dancers' visual fixations and eye movements when watching a contemporary dance film. The hypothesis under investigation was that dance experts' expectations about dance would facilitate their

perception of dance movements. Corroborating this hypothesis, Stevens et al. (2010) discovered that the fixation times of dance experts watching a dance film were significantly shorter than those of novice counterparts – presumably reflecting a cognitive advantage (superior pattern recognition skills and more accurate expectations) of the former over the latter performers. But what of the level of attentional effort required by the creative, possibility thinking involved in learning and performing choreography? While it is understood that attention and focus are imperative in facilitating creativity (e.g., Kasof, 1997), less is known specifically about attentional effort in a dance setting. Kaufman and Baer (2005) identified attentiveness as an essential factor in facilitating creativity as well as stating that “creative performers of movement are those who maintain heightened awareness of and sensitivity to the creativity of the human body” (p. 89). Although this claim appears to support the role of attention in these creative processes, there is a lack of research which specifically examines the relationship between attentional effort and learning and performing choreography.

Unresolved Issues in Cognitive Psychological Research on Dance

From the preceding sections, it is evident that there are at least two major gaps in cognitive psychological research in dance. Firstly, few studies have examined the MI processes of dancers. Accordingly, the extent to which these processes vary with dancers' level of expertise is unknown. Secondly, no published studies could be located in which the attentional effort of dancers was objectively investigated while they engage in the creative or “possibility thinking” (Torrents et al., 2015) process that is hypothesized to aid the learning, performing and imagining of a new piece of choreography. Therefore, the purpose of the present study was to address these objectives.

The Present Study

The present study investigates the relationship between dancers' MI ability, attentional effort, and dance expertise (at three levels: novice, intermediate, and expert performer) while they learned, performed and imagined a piece of dance choreography. In order to measure dancers' MI abilities, we shall use a novel combination of subjective and objective measures described earlier – namely, the MIQ-3, the TAMI and the mental chronometry approach. Attentional effort will be assessed by the measurement of pupil dilation (as recorded by the Tobii Pro Glasses – a wearable eye-tracker; Tobii Technology, 2017).

HYPOTHESES

Hypothesis 1: That dancers' MI abilities will vary with their level of expertise.

Hypothesis 2: That the difference between actual and imagined time required to perform the choreography will vary indirectly with level of dance expertise – such that expert dancers will display the greatest congruence

between actual and imagined time and that novices will display the lowest congruence between these times.

Hypothesis 3: That there will be a significant interaction between level of dance expertise and level of pupil dilation at three time-points throughout the learning, performing and imagined movement conditions.

MATERIALS AND METHODS

Participants

Eighteen female ballet and modern dancers ($M = 23$ years; $SD = 5.85$) took part in this study, with 6 dancers recruited at each of three different levels of expertise (i.e., novice, intermediate and expert) based on the number of years of training that they had received. These levels were defined as follows. “Novice” dancers had received less than 5 years of continuous part-time training ($M = 3$ years; $SD = 1.86$). “Intermediate” dancers had received between 6 and 9 years of continuous part-time training ($M = 8$ years; $SD = 1.43$). Finally, “expert” dancers consisted of ballet or modern teachers who had gained at least 10 years of continuous part-time training and who had also obtained at least one dance teaching qualification with the imperial society of teachers of dance (ISTD) ($M = 14$ years; $SD = 3.01$).

Materials

A short 15 s video of a *tendu* exercise (a short movement of the leg), from the grade 6 modern syllabus (Imperial Society of Teachers of Dance (ISTD), 2017) was used as the piece of choreography to be learned, performed and imagined by participants. *Tendu*, meaning “stretched out” in French, is a foot exercise aimed at warming up and strengthening the feet. This segment was deemed appropriate for three reasons. Firstly, it is drawn from grade 6, which precedes vocational standard examinations (such as intermediate foundation and intermediate). This means that the standard of the segment is between novice and intermediate level and just below that of teaching (expert) level. Secondly, the segment was chosen because both ballet and modern dance contain *tendus*. Therefore, as the sample consisted of mixed experience with the two styles, the exercise was deemed to be equally accessible to all participants. Finally, the segment was selected because, at 15 s of choreography, it provided a significant amount of data to be recorded by the Tobii glasses. It was also of a manageable length so that dancers could learn it under experimental conditions (i.e., it was not feasible to have a full hour long class per person where they would learn a longer piece). No participants had previously learned this particular *tendu* exercise so it was not within their behavioral repertoire. This short exercise required the dancers to carry out a *tendu* to the front, side and back on the right leg followed by the left, with a bend and stretch of both knees in between legs.

Two MI questionnaires were administered to participants before the learning, performing, and imagined movement conditions in order to assess their imagery abilities. These were the objective test of ability in movement imagery

(TAMI; Madan and Singhal, 2013) and the subjective movement imagery questionnaire 3 (MIQ-3, Williams et al., 2012). The Tobii eye-tracking glasses (Tobii Technology, Stockholm, Sweden) were used to record participants' pupil dilation (an index of attentional/mental effort) throughout the creative processes of learning, performing and imagining the piece of choreography. Finally, a stopwatch was used to record how long it took each participant to perform and then imagine the choreography, so that the differences between these times could be analyzed. All data collection took place in the same dance studio and the level of artificial light in the studio was kept constant, in order to avoid unwanted pupil dilation effects. The studio had mirrors on at least one wall of the room.

Procedure

This research was first approved by the graduate research ethics committee, University College Dublin. The lead researcher contacted local dance schools where participants were recruited. Participation was voluntary and began only after the participant had provided informed consent. After such consent was obtained, the participants were provided with instructions and test materials for the TAMI and MIQ-3. Then, they completed these tests. In order to ensure anonymity, the participants were given an ID number which they wrote on their answer booklets and which was also used to label their pupillometry recordings. Upon completion of the questionnaires, participants wore the Tobii glasses and their pupils were calibrated. In order to do this, participants had to look at a light which was held at 9 different points at their eye level, about 2 m in front of them. The nine points are in the shape of a square, 3 points per line. The Tobii monitor is synced up to this light and indicates when calibration is complete at each point, what direction to move the light for the next point and when to move on to the next. It also indicates when calibration is complete.

Participants then watched a video of the *tendu* choreography three times, in order to learn it. They were told that they could mark the movements as they watched them if they wished. Such marking typically involves carrying out the choreography on a smaller scale rather than in full, perhaps using hand gestures to represent each movement. This is a common technique used by dancers when learning choreography (Nordin and Cumming, 2005) and increases the fidelity with which the learning condition represented a real-life, creative scenario. After three viewings, participants were asked to perform the piece in front of the mirrors, while still wearing the Tobii glasses so as to measure their attentional effort while performing. Their performance was also timed. In accordance with typical mental chronometry studies, participants were then asked to imagine themselves carrying out this same piece of choreography from a "third person" imagery perspective (i.e., they were asked to imagine it as if they were watching a video of themselves so it would be comparable to the learning condition). This was also timed so that timing of the performance and imagined movement conditions could be compared. Participants said "start" just before they began imagining it and "stop" when they were finished. They were then asked to remove the Tobii glasses and

were provided with a debrief information sheet and thanked for their participation.

Data Analysis

The data that were collected consisted of scores on the TAMI, MIQ-3 and also of times taken to perform and imagine the choreography. Scores on the TAMI were calculated using Madan and Singhal's (2014) weighted scoring method, whereby more difficult questions gave a higher score than did easier questions. Scores on the MIQ-3 were calculated according to Williams et al.'s (2012) guidelines, whereby each self-report rating out of 7 was added up to reach a total score. Descriptive statistics were calculated and a reliability analysis was also conducted. Thirdly, the time taken to perform and imagine the choreography was recorded using a stop watch. Finally, pupillometry data were recorded by the Tobii eye-tracking glasses. These data were recorded in terms of percentages, whereby 100% is considered typical pupil size. Anything below 100% is seen as the pupil shrinking and above 100% is the amount that the pupil has dilated (Tobii Technology, Stockholm, Sweden, 2017). The typical pupil size, or baseline data, is recorded for each individual when they first put on the glasses and pupils are calibrated as the individual is required to fixate 9 different points. Data from all measures were inputted into SPSS (2017) (SPSS Inc., Chicago, IL, United States) for analysis.

Hypothesis 1, which proposed that there would be a statistically significant difference between each level of expertise in terms of scores on the TAMI and MIQ-3, was tested using two one-way between-groups ANOVAs. The independent variable was level of expertise ($k = 3$) and the dependent variables were scores on the TAMI and MIQ-3, respectively. Hypothesis 2, which proposed that there would be a statistically significant difference between time taken to perform the choreography and time taken to imagine performing the choreography, based on level of expertise, was also tested using a one-way ANOVA. In this case, the times taken to perform and imagine the choreography were subtracted from each other in order to calculate the difference. This score was then used as the dependent variable while level of expertise was the independent variable ($k = 3$). Hypothesis 3 predicted that there would be a statistically significant interaction between pupil dilation at the start, middle and end of the (a) learning, (b) performance, and (c) imagined movement conditions, based on level of expertise. To test this, percentage pupil dilation change was sampled at 33 Hz and was averaged over 1 s at the three time points. The starting point for the learning and performance conditions was at 0–1 s while the middle was 7.5–8.5 s and the end was at 14–15 s. As each participant imagined the movement in their own time, the time points for this condition started at 0–1 s, while middle was exactly halfway between this and when participants told the lead researcher that they were finished (which was the end point). In order to test this hypothesis, a three-way repeated measures ANOVA was carried out whereby the three factors were level of expertise (i.e., beginner, intermediate, and expert), time point (start, middle, and end) and task (learning, performing, or imagining the movement). Sphericity and Levene's tests were

TABLE 1 | Mean scores for the TAMI and MIQ.

Level of dance expertise	TAMI score (M)	MIQ score (M)
Beginner	15.33	65.16
Intermediate	16.33	74.15
Expert	17.33	76.33

conducted and relevant assumptions for the analyses were checked and met.

RESULTS

In order to test hypothesis 1 (as stated above), a set of one-way, between groups ANOVAs were conducted on participants' imagery test scores. Although the apparent mean score differences would suggest an increase in performance on the TAMI and MIQ-3 as level of dance expertise increased (see **Table 1**), hypothesis 1 was not supported for the TAMI; $F(2, 17) = 0.63$, $p = 0.55$, $\eta_p^2 = 0.077$ or the MIQ-3; $F(2, 17) = 2$, $p = 0.17$, $\eta_p^2 = 0.211$. In order to test Hypothesis 2 (as stated above) imagined times were subtracted from movement times in order to calculate the difference. Then, a one-way ANOVA was carried out using this score for each participant. This hypothesis was also rejected, $F(2,17) = 0.12$, $p = 0.88$, $\eta_p^2 = 0.016$.

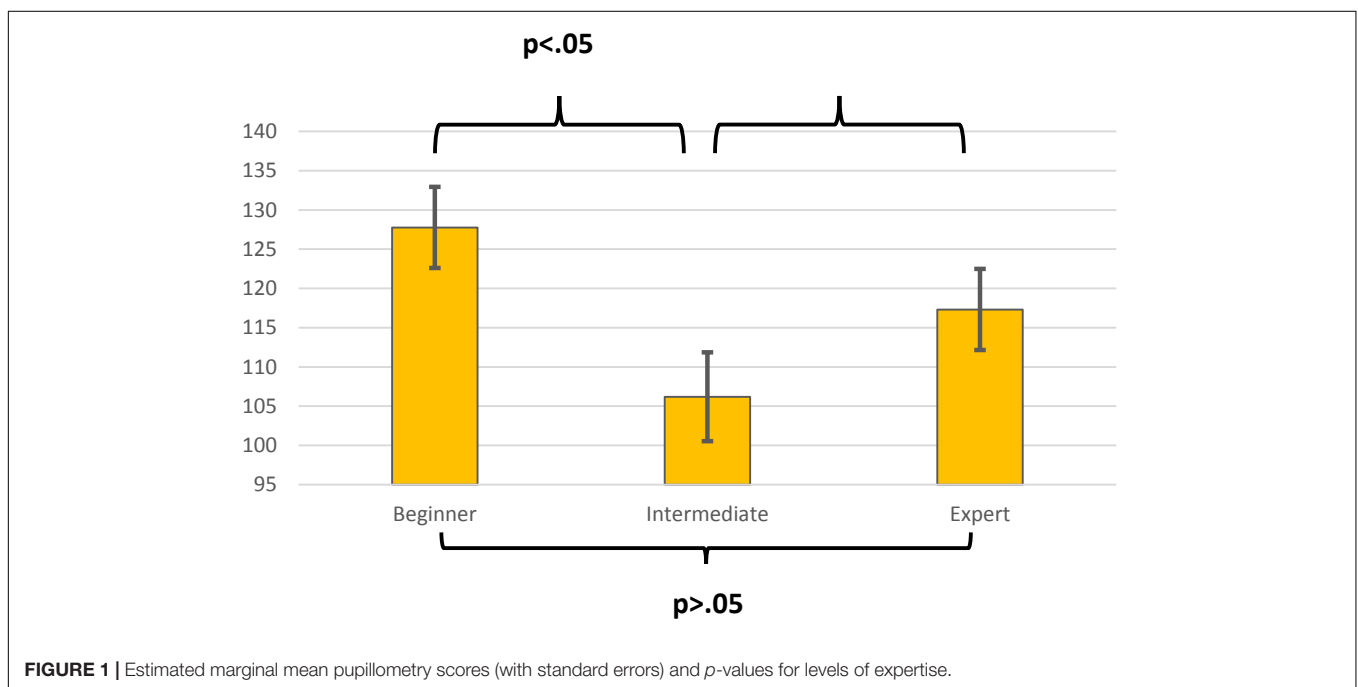
In order to test Hypothesis 3 (as stated above), a three-way repeated measures ANOVA was conducted. Hypothesis 3 was not supported as there was no statistically significant three-way interaction between dancers' pupil dilation at the start, middle and end of each condition based on level of expertise, $F(8,56) = 1.0123$, $p = 0.193$, $\eta_p^2 = 0.173$. Similarly, the two-way

interactions and main effects of time or task were also not significant. However, a significant main effect of expertise was found in terms of their levels of pupil dilation, $F(2,9) = 3.963$, $p = 0.043$, $\eta_p^2 = 0.362$. *Post hoc* Scheffe multiple comparisons of pupillometry scores indicated that the beginners and experts did not significantly differ from each other in pupil dilation, but beginners had significantly higher pupil dilation than the intermediates (see **Figure 1**).

A graphical portrayal of expertise-based differences in pupil dilation across three different time points was conducted for the performance condition (see **Figure 2**) and for the imagined movement condition (see **Figure 3**). Visual inspection of these two graphs suggest that the pattern of pupil dilation of experts over time is more consistent – particularly in the imagined condition – than for either the novice or intermediate performers. However, more fine-grained research is required to test the veracity of this observation. Additionally, an ANCOVA was carried out which explored whether or not the scores on the TAMI could account for the differences in pupillometry at each level of expertise. While TAMI score was not a significant covariate, $F(1,13) = 2.47$, $p = 0.14$, there was still difference between levels, $F(2,13) = 3.709$, $p = 0.05$. Thus, there is a significant difference between levels in terms of pupillometry scores, even when TAMI scores are controlled for.

DISCUSSION

The present study investigated the effect of expertise on dancers' MI abilities and attentional effort while learning, performing, and imagining a piece of choreography. Whereas the TAMI and mental chronometry paradigm provided objective measures of the dancers' MI abilities, the MIQ-3 provided a subjective



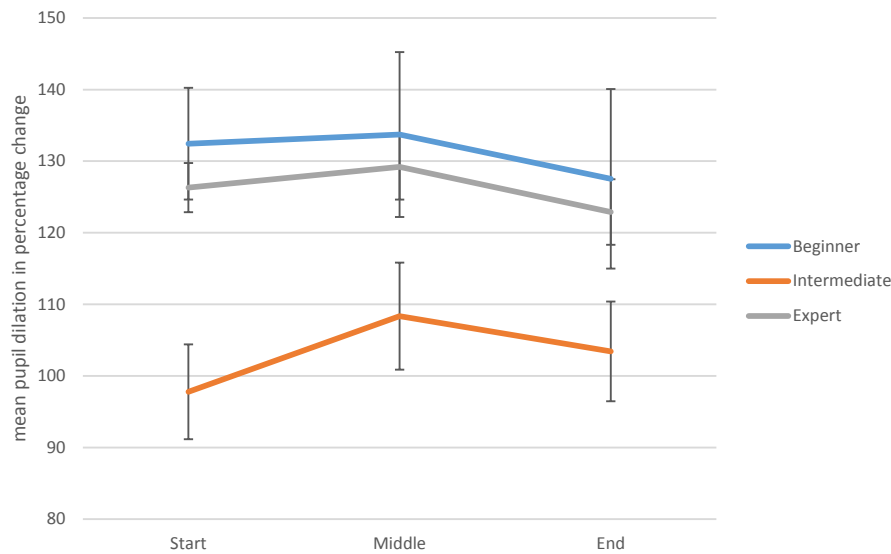


FIGURE 2 | Mean pupil dilation levels throughout the performance condition (with standard error bars). This includes beginner, intermediate, and expert mean pupil dilation levels at three time points.

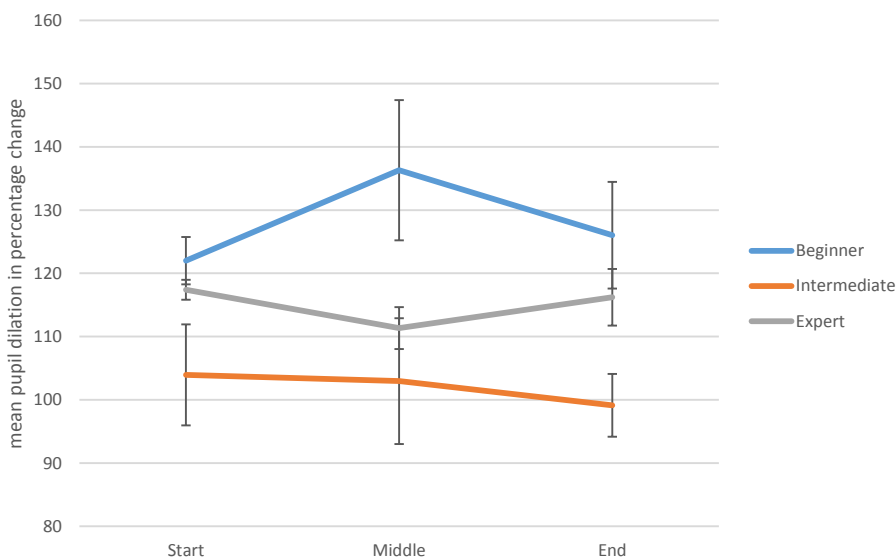


FIGURE 3 | Mean pupil dilation levels throughout the imagined movement condition (with standard error bars). This includes beginner, intermediate, and expert mean pupil dilation levels at three time points.

index of these skills. Pupil dilation (as recorded by eye-tracking equipment) was used to measure the level of attentional effort exerted by the dancers during the learning, performance and imagined execution of choreographed movements.

Let us begin our interpretation of the results by considering the relationships among the different measures of MI. As this study is the first of its kind to assess MI in dancers using a combination of psychometric tests (the TAMI and MIQ-3) and mental chronometry measures, the results are somewhat exploratory in nature. Previously, Collet et al. (2011) found that the temporal congruence between a real and imagined

movement was mediated by experience with the task in question. Unfortunately, our findings seem to contradict those of Collet et al. (2011). Some caution may be required when interpreting this inconsistency, however. This is so because the question of what precise aspect of MI the mental chronometry paradigm actually measures remains largely unresolved. Thus, Williams et al. (2015) raised the possibility that the MIQ-3 and chronometric tests may assess different components of MI. Specifically, they speculated that whereas the MIQ-3 may evaluate people's ability to *generate* a motor image (i.e., creating the initial image in one's minds eye), chronometric measures may assess

people's ability to *maintain* and *control* an image (i.e., retaining the image whilst also being able to manipulate different aspects of it). Collet et al. (2011) also make the point that it may be kinaesthetic imagery that relates to temporal congruence. As the participants in this study were specifically asked to create visual images, this may also explain the lack of significant differences.

Turning to the interpretation of dancers' performance on the TAMI, it may be helpful to review the only previous published study in which an objective measure of MI was administered to dancers. In this study, Jola and Mast (2005) found that dancers scored higher than non-dancers on a measure of similar nature to the TAMI – the mental body transformation task (MBTT; Parsons, 1987). Extrapolating from this research, it may be possible that dancers are more proficient than non-dancers in manipulating body images but that their MI skills do not significantly improve with expertise. Although similarities can be drawn between the MBTT and the TAMI (as they both require participants to manipulate images of their bodies), the relationship between the two measures has not been analyzed to date. Accordingly, we do not know the extent to which these measures overlap in their assessment of MI skills. The fact that we found no evidence of expert-novice differences in dancers' MIQ-3 scores is in line with results reported by Overby (1990). Recall that she found no significant differences between novice and experienced modern and ballet dancers on scores on the original MIQ. However, Williams et al. (2012) argued that the MIQ-3 is more likely than its predecessor to tap into differences in how easily one can generate movement images, due to the more specific demands it places on the participant (e.g., considering different image perspectives). One possible explanation for the absence of expert-novice differences in the present study is that the MIQ-3 may be too generic in its measurement of imagery ability. Thus, Pavlik and Nordin-Bates (2016) have argued that dance-specific imagery tools need to be developed because dancers use certain types of imagery (e.g., metaphorical imagery, where arms may be imagined as wings) that are not common among athletes. Clearly, it would be interesting to investigate the performance of dancers on dance-specific measures of MI. With regard to the pupillometry data (measuring attentional effort), a significant difference was discovered between experience levels at the start of the performance and imagined movement conditions. More specifically, this difference was detected between the beginners and intermediate-level dancers but not between the experts and intermediates or experts and beginners for both conditions. However, from inspection of **Figure 2**, the intermediates' and experts' pupils dilated slightly from the starting point to the middle stage, while the beginners' pupils shrank slightly in comparison to the starting point. This difference may indicate that the beginners' level of dilation at the start was due, in part, to the possibility that participants may have exerted some initial mental effort to work out the cognitive demands of this task. This could perhaps reflect the creative thinking processes required in order to navigate the cognitive challenge of co-ordinating dance movements, as described by Kaufman and Baer (2005) and Torrents et al. (2015). On the other hand, for the imagined movement condition, the beginners' and intermediates' pupils

dilated between the start and 3.5 s, while the experts' shrank slightly (see **Figure 3**). This may suggest that for beginners and intermediates it requires more mental effort to generate a motor image than it does for experts. Although there is currently no pupillometry data on dancers to compare this to, results in other areas of sport and performance psychology indicate that experts can generate motor images easier than can less skilled counterparts (Collet et al., 2011).

Let us now consider some methodological limitations of the present study. The first weakness concerns the absence of a kinematic performance measure for each dancer while attempting to master the choreographed movements. Although complex and time-consuming to implement, such a measure would have helped our study because it could have ensured that if any dancer had forgotten the choreography, the precise time point of this occurrence could have been noted and accounted for when analyzing subsequent pupillometry data. Additionally, it may have been useful to have a measure of how accurately each participant performed the choreography. This too could have been compared to pupillometry data for both the learning and performance conditions and could have contributed to our understanding of variances in pupil dilation across experience levels. It would be expected that the expert dancers who are more experienced in learning and performing, would perform and learn the choreography more accurately than would less skilled counterparts – which thus could subsequently affect levels of attentional effort. A second weakness concerns our interpretation of the pupil dilation data. According to Mathôt (2018), any information that activates the mind, or increases its “processing load” (Beatty, 1982; see also O'Shea and Moran, 2016), induces dilation of the pupil. In this paper, we have favored a mental effort-driven interpretation of pupil dilation. However, we must acknowledge that fluctuations in pupil size can occur for reasons other than as a function of the expenditure of mental effort. For example, Bouma and Baghuis (1971) speculated that they may be due simply to the waxing and waning of arousal. In a similar vein, Laeng et al. (2016) identified emotional engagement as a trigger for pupil dilation, whereas Hennessy and Amabile (2010) found emotion to be a hindering factor in creative processes. Unfortunately, as the present study lacked an independent measure of arousal and/or emotional engagement, we cannot exclude the possibility that these latter variables may have influenced our results. Nevertheless, our research is novel in being the first “process tracing” investigation of MI and attentional effort (as measured by pupil dilation) in dancers who are forced to engage in creative possibility thinking when learning and performing.

With regard to potentially fruitful directions for further psychological research on expert-novice differences in dance, several options are apparent. Firstly, future investigators of this topic may wish to include additional MI dimensions as imagery control (the ease with which a mental image can be manipulated by the person who creates it; Moran and Toner, 2017) and imagery accuracy or its “exactness of reference” (Denis, 1985). Secondly, it would be interesting to investigate the degree to which attentional effort affects the *accuracy* of dancers' mental simulation and/or recall of dance movements, as the accuracy

of performance may also reflect the extent to which the dancer could interpret and create these movements. Although MI and attentional effort may mediate the creative thinking required to learn and perform choreography, it may also be interesting to consider the effect of other factors in a dance setting which are known to interfere with creative thinking and the creation of one's own choreography, for example, motivation and the environment (Hennessy and Amabile, 2010). Additionally, further research is required to explore the extent to which prolonged experience of learning and performing dance movements affects multi-sensory integration (the ability to combine information from different sensory modalities; Grunbaum and Schram Christensen, 2018).

To conclude, the present study suggests that there is a significant difference between beginner and intermediate dancers in levels of pupil dilation when faced with the task of performing and imagining a short piece of choreography. This finding is beneficial in understanding the cognitive demands which face the dancer, as well as the mechanisms which may underlie the creative thinking proposed necessary to the performing and imagining of choreography. The present study also paves way for further development of this research, such as administering several MI measures with dancers and comparing results, comparing pupil dilation with measures of arousal or performance appraisals and looking at what exact cognitive skills may vary with different levels of dance expertise.

REFERENCES

- Abraham, A., Dunskey, A., and Dickstein, R. (2017). The effect of motor imagery practice on elev  performance in adolescent female dance students: a randomized controlled trial. *J. Imag. Res. Sport. Phys. Act.* 12, doi: 10.1515/jirspa-2016-0006.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psych. Bull.* 91, 276–292. doi: 10.1037/0033-2909.91.2.276
- Beatty, J., and Lucero-Wagoner, B. (2000). “The pupillary system,” in *Handbook of Psychophysiology*, 2nd Edn. eds. J. T. Cacioppo, L. G. Tassinari, and G. G. Berntson (New York, NY: Cambridge University Press), 142–162.
- Bouma, H., and Baghuis, L. C. J. (1971). Hippus of the pupil: periods of slow oscillations of unknown origin. *Vis. Res.* 11, 1345–1351. doi: 10.1016/0042-6989(71)90016-2.
- Burge, W. K., Ross, L. A., Amthor, F. R., Mitchell, W. G., Zotov, A., and Visscher, K. M. (2013). Processing speed training increases the efficiency of attentional resource allocation in young adults. *Front. Hum. Neuro.* 7:684. doi: 10.3389/fnhum.2013.00684.
- Chappell, K., Craft, A., Rolfe, L., and Jobbins, V. (2009). Dance partners for creativity: choreographing space for co-participative research into creativity and partnership in dance education. *Res. Dance Educ.* 10, 177–197. doi: 10.1080/14647890903324147.
- Collet, C., Guillot, A., Lebon, F., MacIntyre, T., and Moran, A. (2011). Measuring motor imagery using psychometric, behavioural, and psychophysiological tools. *Exerc. Sport Sci. Rev.* 39, 85–92. doi: 10.1097/JES.0b013e31820ac5e0
- Cross, E. S., Acquah, D., and Ramsey, R. (2014). A review and critical analysis of how cognitive neuroscientific investigations using dance can contribute to sport psychology. *Int. Rev. Sport Exerc. Psychol.* 7, 42–71. doi: 10.1080/1750984X.2013.862564.
- Cross, E. S., and Ticini, L. F. (2012). Neuroaesthetics and beyond: new horizons in applying the science of the brain to the art of dance. *Phenomenol. Cogn. Sci.* 11, 5–16. doi: 10.1007/s11097-010-9190-y
- Denis, M. (1985). Visual imagery and the use of mental practice in the development of motor skills. *Can. J. App. Sport Sci.* 10, 4s–16s.

ETHICS STATEMENT

The study was carried out in accordance with the recommendations from University College Dublin's Human Research Ethics Committee with written informed consent from all participants.

AUTHOR CONTRIBUTIONS

KC and AM contributed to the conception and design of the study. KC conducted the research and carried out the statistical analysis. KC and AM drafted sections of the first draft of the manuscript, while AM then edited, and critically revised it. BR contributed to the statistical analysis and interpretation of results in preparing the revised version of the manuscript and also to the responses to reviewers. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

We wish to acknowledge with gratitude the technical assistance provided by Colin Burke (School of Psychology, University College Dublin, Dublin, Ireland).

- Ericsson, K. A. (2018). “Capturing expert thought with protocol analysis: concurrent verbalizations of thinking during experts' performance on representative tasks,” in *The Cambridge Handbook of Expertise and Expert Performance*, eds K. A. Ericsson, R. R. Hoffman, A. Kozbelt, and A. M. Williams (Cambridge: Cambridge University Press), 192–212. doi: 10.1017/9781316480748.012
- Fink, A., Benedek, M., Grabner, R. H., Staudt, B., and Neubauer, A. C. (2007). Creativity meets neuroscience: experimental tasks for the neuroscientific study of creative thinking. *Methods* 42, 68–76. doi: 10.1016/j.ymeth.2006.12.001.
- Fink, A., Grabner, R. H., Benedek, M., and Neubauer, A. C. (2006). Divergent thinking training is related to frontal electroencephalogram alpha synchronization. *Eur. J. Neurosci.* 23, 2241–2246. doi: 10.1111/j.1460-9568.2006.04751.x
- Fink, A., Graif, B., and Neubauer, A. C. (2009). Brain correlates underlying creative thinking: EEG alpha activity in professional vs. novice dancers. *Neuroimage* 46, 854–862. doi: 10.1016/j.neuroimage.2009.02.036
- Fink, A., and Woschnjak, S. (2011). Creativity and personality in professional dancers. *Pers. Individ. Diff.* 51, 754–758. doi: 10.1016/j.paid.2011.06.024
- Fisher, V. J. (2017). Unfurling the wings of flight: clarifying ‘the what’ and ‘the why’ of mental imagery use in dance. *Res. Dance Educ.* 18, 252–272. doi: 10.1080/14647893.2017.1369508.
- Goldstein, E. B. (2011). *Cognitive Psychology*, 3rd Edn. Belmont, CA: Wadsworth/Cengage.
- Granholm, E., Asarnow, R. F., Sarkin, A. J., and Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology* 33, 457–461. doi: 10.1111/j.1469-8986.1996.tb01071.x
- Grunbaum, T., and Schram Christensen, M. (2018), *Sensation of Movement*, London: Routledge.
- Guillot, A., and Collet, C. (2005). Duration of mentally simulated movement: a review. *J. Mot. Behav.* 37, 10–20. doi: 10.3200/JMBR.37.1.10-20.
- Guss-West, C., and Wulf, G. (2016). Attentional focus in classical ballet: a survey of professional dancers. *J. Dance Med. Sci.* 20, 23–29. doi: 10.12678/1089-313X.20.1.23.
- Hall, C. R., and Pongrac, J. (1983). *Movement Imagery Questionnaire*. London: University of Western Ontario.

- Hanakawa, T. (2016). Organizing motor imageries. *Neuro. Res.* 104, 56–63. doi: 10.1016/j.neures.2015.11.003.
- Heiland, T., and Rovetti, R. (2013). Examining effects of franklin method metaphorical and anatomical mental images on college dancers' jumping height. *Res. Dance Educ.* 14, 141–161. doi: 10.1080/14647893.2012.712105.
- Hennessy, B. A., and Amabile, T. M. (2010). Creativity. *Ann. Rev. Psychol.* 61, 569–598. doi: 10.1146/annurev.psych.093008.100416
- Hess, E. H., and Polt, M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science* 140, 1190–1192. doi: 10.1126/science.143.3611.1190
- Hétu, S., Grégoire, M., Saimpont, A., Coll, M. P., Eugène, F., Michon, P. E., et al. (2013). The neural network of motor imagery: an ALE meta-analysis. *Neuro. Biobehav. Rev.* 37, 930–949. doi: 10.1016/j.neubiorev.2013.03.017.
- Imperial Society of Teachers of Dance (ISTD) (2017). Grade 6 modern syllabus. London: ISTD.
- James, W. (1890). *The Principles of Psychology*, Vol. 2. Cambridge, MA: Harvard University Press. doi: 10.1037/10538-000
- Jeannerod, M. (1994). The representing brain: neural correlates of motor intention and imagery. *Behav. Brain Sci.* 17, 187–202. doi: 10.1017/S0140525X00034026.
- Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage* 14, 103–109. doi: 10.1006/nimg.2001.0832.
- Jeannerod, M. (2006). *Motor Cognition*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780198569657.001.0001
- Jola, C., and Mast, F. W. (2005). Mental object rotation and egocentric body transformation: two dissociable processes? *Spatial Cogn. Comput.* 5, 217–237. doi: 10.1080/13875868.2005.9683804.
- Kahneman, D. (1973). *Attention and Effort*. New York, NY: Prentice-Hall.
- Kasof, J. (1997). Creativity and breadth of attention. *Creat. Res. J.* 10, 303–315. doi: 10.1207/s15326934crj1004_2
- Kaufman, J. C., and Baer, J. (eds.). (2005). *Creativity Across Domains: Faces of the Muse*. London: Psychology Press. doi: 10.4324/9781410611925
- Laakso, A. (2011). Embodiment and development in cognitive science. *Cogn. Brain Behav. Interdisc. J.* 4, 409–425.
- Laeng, B., Eidet, L. M., Sultvedt, U., and Panksepp, J. (2016). Music chills: the eye pupil as a mirror to music's soul. *Consc. Cogn.* 44, 161–178. doi: 10.1016/j.concog.2016.07.009.
- Madan, C. R., and Singhal, A. (2013). Introducing TAMI: an objective test of ability in movement imagery. *J. Mot. Behav.* 45, 153–166. doi: 10.1080/00222895.2013.763764.
- Madan, C. R., and Singhal, A. (2014). Improving the TAMI for use with athletes. *J. Sport. Sci.* 32, 1351–1356. doi: 10.1080/02640414.2014.889847.
- Mathôt, S. (2018). Pupillometry: psychology, physiology, and function. *J. Cogn.* 16, 1–23. doi: 10.5334/joc.18
- Mathôt, S., and Van der Stigchel, S. (2015). New light on the mind's eye: the pupillary light response as active vision. *Curr. Dir. Psychol. Sci.* 24, 374–378. doi: 10.1177/0963721415593725.
- Moran, A. (1996). *The Psychology of Concentration in Sport Performers: A Cognitive Analysis*. Hove: Psychology Press.
- Moran, A., Guillot, A., MacIntyre, T., and Collet, C. (2012). Re-imagining motor imagery: building bridges between cognitive neuroscience and sport psychology. *Br. J. Psychol.* 103, 224–247. doi: 10.1111/j.2044-8295.2011.02068.x
- Moran, A., Quinn, A., Campbell, M., Rooney, B., Brady, N., and Burke, C. (2016). Using pupillometry to evaluate attentional effort in quiet eye: a preliminary investigation. *Sport Exerc. Perform. Psychol.* 5:365. doi: 10.1037/spy0000066Issn.
- Moran, A., and Toner, J. (2017). *A Critical Introduction to Sport Psychology*, 3rd Edn. London: Routledge. doi: 10.4324/9781315657974
- Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., and Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum. Brain Map.* 35, 4140–4154. doi: 10.1002/hbm.22466.
- Nordin, S. M., and Cumming, J. (2005). Professional dancers describe their imagery: where, when, what, why, and how. *Sport Psychol.* 19, 395–416. doi: 10.1123/tsp.19.4.395.
- Nordin, S. M., and Cumming, J. (2007). Where, when, and how: a quantitative account of dance imagery. *Res. Quart. Exerc. Sport* 78, 390–395. doi: 10.1080/02701367.2007.10599437.
- O'Shea, H., and Moran, A. (2016). Chronometric and pupil-size measurements illuminate the relationship between motor execution and motor imagery in expert pianists. *Psychol. Music* 44, 1289–1303. doi: 10.1177/0305735615616286.
- O'Shea, H., and Moran, A. (2017). Does motor simulation theory explain the cognitive mechanisms underlying motor imagery? a critical review. *Front. Hum. Neuro.* 11:72. doi: 10.3389/fnhum.2017.00072
- Overby, L. Y. (1990). A comparison of novice and experienced dancers' imagery ability. *J. Men. Imag.* 14, 173–184.
- Overby, L. Y., and Dunn, J. (2011). The history and research of dance imagery: implications for teachers. *IADMS Bull. Teach.* 3, 9–11.
- Parsons, L. M. (1987). Imagined spatial transformations of one's hands and feet. *Cogn. Psychol.* 19, 178–241. doi: 10.1016/0010-0285(87)90011-9.
- Pavlik, K., and Nordin-Bates, S. (2016). Imagery in dance: a literature review. *J. Dance Med. Sci.* 20, 51–63. doi: 10.12678/1089-313X.20.2.51.
- Piquado, T., Isaacowitz, D., and Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology* 47, 560–569. doi: 10.1111/j.1469-8986.2009.00947.x
- Porter, G., Troscianko, T., and Gilchrist, I. D. (2007). Effort during visual search and counting: insights from pupillometry. *Quart. J. Exp. Psychol.* 60, 211–229. doi: 10.1080/17470210600673818
- Sarter, M., Gehring, W. J., and Kozak, R. (2006). More attention must be paid: the neurobiology of attentional effort. *Brain Res. Rev.* 51, 145–160. doi: 10.1016/j.brainresrev.2005.11.002
- Shepard, S., and Metzler, D. (1988). Mental rotation: effects of dimensionality of objects and type of task. *J. Exper. Psychol. Hum. Percept. Perform.* 14, 3–11. doi: 10.1037/0096-1523.14.1.3
- Simonton, D. K., and Damian, R. I. (2013). "Creativity," in *Handbook of Cognitive Psychology*, ed D. Reisberg (New York, NY: Oxford University Press), 795–807.
- Smith, E. E., and Kosslyn, S. M. (2007). *Cognitive Psychology: Mind and Brain (International Edition)*. Upper Saddle River, NJ: Pearson.
- SPSS (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM Corp.
- Stevens, C., Winskel, H., Howell, C., Vidal, L. M., Latimer, C., and Milne-Home, J. (2010). Perceiving dance schematic expectations guide experts' scanning of a contemporary. *J. Dance Med. Sci.* 14, 19–25. doi: 10.1136/bmj.c889
- Torrents, C., Ric, A., and Hristovski, R. (2015). Creativity and emergence of specific dance movements using instructional constraints. *Psychol. Aesth. Creat. Arts* 9:65. doi: 10.1037/a0038706
- Unsworth N., and Robison M. K. (2015). Individual differences in the allocation of attention to items in working memory: evidence from pupillometry. *Psychon. Bull. Rev.* 22, 757–765. doi: 10.3758/s13423-014-0747-6
- Weisberg, R. W. (2018). "Expertise and structured imagination in creative thinking: reconsideration of an old question," in *The Cambridge Handbook of Expertise and Expert Performance*, eds K. A. Ericsson, R. W. Hoffman, A. Kozbelt, and A. M. Williams (Cambridge: Cambridge University Press), 812–834. doi: 10.1017/9781316480748.041
- Williams, A. M., and Ericsson, K. A. (2005). Perceptual-cognitive expertise in sport: some considerations when applying the expert performance approach. *Hum. Mov. Sci.* 24, 283–307. doi: 10.1016/j.humov.2005.06.002
- Williams, S. E., Cumming, J., Ntoumanis, N., Nordin-Bates, S. M., Ramsey, R., and Hall, C. (2012). Further validation and development of the movement imagery questionnaire. *J. Sport Exerc. Psychol.* 34, 621–646. doi: 10.1123/jsep.34.5.621
- Williams, S. E., Guillot, A., Di Rienzo, F., and Cumming, J. (2015). Comparing self-report and mental chronometry measures of motor imagery ability. *Eur. J. Sport Sci.* 15, 703–711. doi: 10.1080/17461391.2015.1051133

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 Carey, Moran and Rooney. 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.



Expertise in Evaluating Choreographic Creativity: An Online Variation of the Consensual Assessment Technique

Lucie Clements^{1,2*}, Emma Redding², Naomi Lefebvre Sell² and Jon May³

¹ Department of Psychology and Counselling, University of Chichester, Chichester, United Kingdom, ² Dance Science, Trinity Laban Conservatoire of Music and Dance, London, United Kingdom, ³ School of Psychology, Cognition Institute, University of Plymouth, Plymouth, United Kingdom

OPEN ACCESS

Edited by:

Kathryn Friedlander,
University of Buckingham,
United Kingdom

Reviewed by:

Glenna Batson,
Wake Forest University, United States
Luke Stephen Hopper,
Edith Cowan University, Australia

*Correspondence:

Lucie Clements
l.clements@chi.ac.uk

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 29 April 2018

Accepted: 23 July 2018

Published: 24 August 2018

Citation:

Clements L, Redding E, Lefebvre
Sell N and May J (2018) Expertise
in Evaluating Choreographic
Creativity: An Online Variation of the
Consensual Assessment Technique.
Front. Psychol. 9:1448.
doi: 10.3389/fpsyg.2018.01448

In contemporary dance, experts evaluate creativity in competitions, auditions, and performances, typically through ratings of choreography or improvisation. Audiences also implicitly evaluate choreographic creativity, so dancers' livelihoods also hinge upon the opinions of non-expert observers. However, some argue that the abstract and often pedestrian nature of contemporary dance confuses non-expert audiences. Therefore, agreement regarding creativity and appreciation amongst experts and non-experts may be low. Finding appropriate methodologies for reliable and real-world creativity evaluation remains the subject of considerable debate within the psychology creativity research field. Although considerably variant in methodological operationalisation, the Consensual Assessment Technique (CAT) asks individuals to use an implicit definition to assess creativity in others' work. This study aimed to investigate the role of experience and expertise in the evaluation of choreographic creativity, with a secondary aim of testing the feasibility of an online snowballing methodology for large-scale dance-specific research, informed by the methodology of the CAT. We filmed 23 Contemporary Dance students each performing a 3-min peer-choreographed solo and then recruited 850 online evaluators with varying degrees of expertise and experience in dance and creativity. Evaluators viewed at least one randomly selected video and rated creativity, technical ability, appreciation and understanding of the work, each using a seven-point Likert scale. A one-way ANOVA showed a significant difference in creativity ratings across the 23 videos, and creativity correlated significantly with the other variables. We then categorized evaluators on nine aspects of their dance and creative experience and entered the data into a repeated-measures linear mixed model. Two of the fixed effects yielded differences in creativity evaluations: (i) contemporary choreographic experience and (ii) self-reported creative expertise, as did the random effect of the video. The results indicate that personal experience of the choreographic process impacts creativity assessment, above and beyond experience in dance class participation. Implications for creativity assessment within creativity research and practice are discussed.

Keywords: creativity, choreography, contemporary dance, expertise, audience, assessment

INTRODUCTION

'Contemporary dance' loosely refers to a range of dance styles that use the body to explore and express conceptual ideas or images (Strauss and Nadel, 2012). In contemporary dance, there are no set movement sequences to draw from so there is an expectation of finding new and inventive movement. The focus is, therefore, less on the formulaic construction of movement than in classical forms such as ballet, with an often-deliberate rebellion against codified technique. It is this freedom that supports the argument that contemporary dance is creative by nature (H'Doubler, 1998). Researchers commonly cite Guilford's (1950) presidential address to the American Psychological Association as the defining moment in persuading psychology researchers of the value and importance of scientific research into creativity (Kaufman and Sternberg, 2010; Runco, 2014). Psychological research has facilitated depth of understanding of the predictors, correlates and consequences of creativity, but typically focuses on general population research, with less research within specialist domains (Kaufman and Sternberg, 2010; Runco and Acar, 2012; Long et al., 2014; Runco, 2014; Simonton, 2015). Little has been published drawing on scientific methods within the domain of contemporary dance (Thomson and Jaque, 2017).

The lack of creativity research within the performing arts more broadly may be due to scientists' misinformed beliefs that performing artists are replicators who express work generated by others, rather than creators, and are therefore not a population of interest (Kogan, 2002; Sawyer, 2014; Thomson and Jaque, 2017). Butterworth (2004) notes that this traditional hierarchy of 'choreographer-as-creative' and 'dancer-as-reproducer' is no longer the sole means by which creativity occurs, citing numerous ways in which the dancer expresses their creativity in choreography. The boundary between dancer and choreographer is blurred, and dance students now learn both performance and creative skills. Professional contemporary dancers often contribute to the development of movement material, through 'exploring, selecting, and developing dance material' (Stevens and McKechnie, 2005, p. 40). The process is often guided by 'tasking', the use of a problem set by the choreographer, and solved by the dancers (May et al., 2011). Typically, each dancer's material will contribute in some way, through refinement of the movement and changes to timing, resulting in a creative product (Stevens et al., 2001). Farrer (2014) notes that whether improvising, choreographing, transforming a phrase of movement, or completing a task, dancers embody numerous creative roles, yet even dancers themselves do not recognize their creativity. These multiple perspectives highlight a broad lack of awareness of dancers' choreographic creativity, calling for greater scientific attention to this unique domain of creativity.

The purpose of our work was to investigate how experience in contemporary dance impacts assessment of choreographic creativity, because contemporary dance requires communication of creative ideas to an audience (Humphrey, 1959; Burrows, 2010; Risner, 2000). Thus, creativity in dance is a social phenomenon (Łuczniak, 2015). As Csikszentmihalyi (1999) notes, "*The underlying assumption is that an objective quality*

called 'creativity' is revealed in the products and that judges and raters can recognize it" (p. 314). Csikszentmihalyi (2014) argues that the interaction between three elements of a system constitute creativity. A culture contains symbolic rules for creativity, the individual brings that creativity into the domain, but creativity is only brought to fruition when experts from that domain recognize the creativity. Recognition of creativity occurs in contemporary dance education (for example, the ability to demonstrate creative engagement in improvisation is a typical entry requirement to higher education dance training), subsequent student assessments, and in reviewing professional work. Although experts are imperative to real-world creativity assessment, non-experts also play a role in the day-to-day sustenance of creative careers, and varying levels of expertise or knowledge may predict differences in assessment of creativity (Hong and Lee, 2015).

Since participation in contemporary dance is an increasingly popular recreational, educational and professional pursuit, one could argue that the audiences who engage with and see this creativity should also be increasing too. Burrows (2010) highlights that contemporary dance audiences seek novelty, but alternative research has also shown that some less experienced contemporary dance audiences report confusion, failure to understand the choreographic intention, and lack of enjoyment (Stevens et al., 2007, 2009; Van Dyke, 2010). Audiences of varying levels of expertise, levels or types of training, may, therefore, assess creativity differently. Research in dance indicates that non-expert dance audiences may fail to understand the meaning behind contemporary dance, perhaps because contemporary dance is detached from the 'magic' seen in dance which makes use of popular music, costumes and staging (Stevens et al., 2009). Contemporary dance has not become rooted in modern westernized culture in the same way other art forms or classical ballet have. For example, a dance director reports that his audiences mainly consist of friends, family or supporters of those directly involved in the performance rather than members of the public (Van Dyke, 2010). Contemporary dancers are often dressed in plain, everyday clothes or speaking directly to audiences; the movement is often pedestrian and effortless, or, hugely effortful. Often, dancers create movement without music, and the music is added later in the choreographic process. Thus contemporary dance may be a particularly unique and ripe area for novel research into creativity, and given this previous research we were interested in the broad role of expertise and understanding of contemporary dance in assessing creativity.

Williams et al. (2016) note that despite the growth of the psychology of creativity over the last 25 years, in particular, many fundamental complexities remain. One such challenge is finding appropriate methodologies for investigating previously underresearched domains of creativity. Problem solving approaches are perhaps the most common methodology seen in psychology research, where 'creativity' lies in the process or means by which an individual arrives at a solution (Lubart, 2001). Problem-solving measures predominantly investigate insight, also known as the 'aha moment' (e.g., the Remote Associates Tests, Wallas, 1926; Mednick and Mednick, 1971;

Runco and Jaeger, 2012). In these tests, problem solving involves a two-stage process of divergent and convergent thinking; restructuring the problem by reframing one's mental approach, to find the one appropriate answer (Guilford, 1956). A small number of research studies have used problem-solving approaches to dancers' creativity, using measures of divergent thinking (the ability to produce multiple responses to a problem) which is considered the 'backbone of creativity assessment' (Runco, 2014, p. 14). Stinson (1993) found students in Chinese dance education were significantly less creative (in divergent thinking) than a non-dancing control group. Fink and Woschnjak (2011) found differences in divergent thinking across contemporary, ballet and jazz dance, suggesting that creativity differs within dance genres. These studies suggest that differences in creativity occur at the microdomain level of dance, yet their generalized approach to assessing creativity may limit their usefulness.

There are reasons why traditional divergent thinking measures may be of limited use for choreographic creativity. Most importantly, some criticize the problem-solving approach to creativity assessment for constituting just one type of creativity, which assumes domain generality of the cognitive processes (i.e., attention, perception, memory, language, and intelligence) underpinning creativity (Kaufman and Baer, 2004; Runco, 2014). At this level, creativity is a nomothetic process shared by, and accessible to, all humans (Simonton, 1999; Glăveanu, 2010). This generalist perspective arguably lacks sensitivity to the individual nuances of creative specialization that manifest in different ways across different fields (Baer, 1998; Feist, 1998; Hu and Adey, 2002; Julmi and Scherm, 2015). Divergent thinking tests may assess only narrow ranges of ability and may not be conclusive about measuring 'creativity' itself. Instead, they indicate abilities related to creativity, which may not be as relevant in specialized domains (Amabile, 1982; Baer and McKool, 2009). Thus it is important also to develop methodologies that are sensitive to the individual nuances of creativity in each domain.

Choreographic creativity, for example, implicates embodied cognition: cognitive processes are rooted in physical interaction with the world (Wilson, 2002; Stevens and McKechnie, 2005). Embodiment emphasizes both physical exploration and knowledge (Kogan, 2002). Dancers understand the intention and action of others moving in the same space and use the body for problem-solving, demonstrating creativity by thinking with the body (Kirsh, 2011). Choreographic creativity uses both awareness of kinesthetic knowledge and experience in/through the body and explicit knowledge of the external world; cognition is situated (Risner, 2000; Kirsh, 2010, 2011). Thus creativity in dance is a process of using the body in novel ways in response to a task and the ability to successfully and fluidly link body positions into a developed sequence (Stevens et al., 2001; Stevens and McKechnie, 2005; Kirsh, 2011). These processes use memory, language and perception as well as space, time, motion and physical expression, with decreased emphasis on verbal and greater emphasis on nonverbal communication (Bläsing et al., 2010; Thomson and Jaque, 2017). Hagood (2001) writes that dance, in general, is "an extremely complex experience to attempt to measure" (p. 27). However, embodiment and process

are critical, which differs starkly from the pen and paper medium emphasized in time-limited psychology measurement traditions; thus studying creativity in dance would be wise to use dance in its natural movement based form.

One of the most widely advocated domain specific means of assessing situated creativity is the Consensual Assessment Technique (CAT; Amabile, 1982). The CAT is popular in psychology since it is unrelated to any specific creativity theory, meaning that its use is broad and relevant to any domain of creativity (Baer and McKool, 2009). In the CAT methodology, experts assess creativity using an implicit understanding within their specific domain (Amabile, 1982; Amabile and Pillemer, 2012). Similarly, assessors in dance use an implicit creativity definition to assess. For example, it is common to obtain mean scores from panels assessments during improvisation at an audition.

However, the CAT has some challenges. Namely, there are no clear guidelines for implementation, and many variations have been used to investigate specific domains. It is a process of obtaining evaluations from raters without using a formal tool or needing to provide explicit criteria against which creativity must be assessed. Conventionally, it is expected that raters should share some common understanding of the domain to support a consensus.

Although a large body of research has investigated audience responses to classical dance as a performance (See Calvo-Merino et al., 2005; Reason and Reynolds, 2010), there is a paucity of research into contemporary dance audiences which focuses on perceptions of creativity. Research has been undertaken to explore the associative and affective results of performance (e.g., Stevens and McKechnie, 2005), but no research has considered audience evaluations of creativity using the psychology of creativity methods such as the CAT. Research using the CAT supports that expert and non-expert creativity assessments of poems differed significantly different, with expert raters giving a higher rating than non-experts, thus is a suitable methodology for investigating choreographic creativity (Kaufman et al., 2008). Kokotsaki and Newton (2015) suggest a continuum of insider-outsider status that potential creativity assessors have, depending on their expertise and experience. Therefore, using a simple dichotomy of expert or non-expert may be too restrictive, particularly in dance where individuals gain experience through doing, making and watching.

The role of creativity has been the subject of considerable interest in psychology research but is yet to be explored in depth in dance within a scientific framework. Therefore, the purpose of this research was to establish an understanding of expertise on the attribution of creativity in contemporary dance choreography. We aimed to recruit a large sample of assessors to judge choreographic creativity of contemporary dance. Informed by the method of the CAT, we used a quantitative methodology to assess the impact of expertise in assessing creativity in contemporary dance to rate video clips of student choreographies (Amabile, 1996). Additionally, we collected measures of perceptions of technical ability, liking and ability to find meaning, as previous research has indicated that non-experts use these variables to assess creativity (e.g., Kozbelt, 2004; Glass and Stevens, 2005).

MATERIALS AND METHODS

Participants

Choreographers

Students ($n = 24$; male $n = 6$, female $n = 18$, mean age = 20.2 years; $SD = 1.6$ years) studying in the 1st year of a BA Contemporary Dance at Trinity Laban, a leading UK Dance Conservatoire, consented to participate in the research. Students entered onto the degree having been assessed for both technical and creative skill at audition (evaluated through a panel marked improvisation), thus had been selected onto the program for their creative potential. Their dance training consists of technique classes in Contemporary Dance (such as Graham and Cunningham) and Ballet, as well as Choreography classes focused on developing processes of exploratory non-stylistic ways of moving from within the body. Students take additional modules in performance and contextual studies. Students were all members of the same choreography class, taught by the same teacher, and had been randomly allocated to this teacher's class at the start of the academic year (from four possibilities).

Creativity Raters

We recruited creativity raters ($n = 1084$) from a variety of levels of expertise to the research. After data screening and cleaning, the final sample size was 850 raters (female $n = 682$, male $n = 158$, other $n = 10$). Participants ranged in age from 18 to 77 years ($M = 31.6$, $SD = 12.9$). We created dummy variables using the nine categories of experience and expertise seen in **Table 1**, whereby an individual who's answer was 'No' is coded as the reference category of '0', and an individual who's answer was 'Yes' to any degree of experience is coded as '1'. The employment categories were answered qualitatively and coded by the first author as 'No' or 'Yes'. An overview of rater experience and expertise in dance and creativity are shown in **Table 1**.

Measures

Video Stimuli

We obtained videos of a short solo choreography ($n = 23$; duration 172–194 s), which were created for the students' choreography module assessment. The choreography was danced by a classmate of the student, rather than the choreographer themselves. We filmed the choreographies in a mirrorless dance

studio in natural lighting to standardize the videos and remove confounding variables relating to production. We used a wide shot of the dance studio which replicated a head-on audience view. All dancers dressed in plain, dark colored practice clothes. An audio-visual expert removed the music and added a fade in and out at the start and end of each piece.

Creativity Ratings

Creativity was assessed using a seven-point Likert scale (*How creative did you think the piece was?*; 1. Not at all creative – 7. Very creative) informed by the method of the CAT (Amabile, 1983). In addition to the target question, participants answered three additional questions; *How much did you like the piece?* (1. Not at all – 7. Very Much); *How technically skilled did you think the dancer was?* (1. Not at all technically skilled – 7. Very technically skilled); *How able were you to find meaning in the piece?* (1. Not at all able to find meaning – 7. Very able to find meaning).

Procedure

We obtained institutional ethical approval. Following this, a choreography teacher provided initial consent to approach her first-year choreography students to provide choreographic material for creativity assessment in the research. The contemporary dance students consented at the end of a timetabled choreography class, 2 weeks before their choreography assessment. Each student's assessed work was a three-minute solo performed by a peer in the same class, so each student consented once for the inclusion of their choreography and a second time as a performer in a peer's work.

On the day of the assessment and filming for the research, each participant provided secondary verbal consent to confirm his or her inclusion. One participant was injured so did not undertake her performance, resulting in 23 videos. We embedded the clips into an online survey via a video hosting site. Snowball sampling was used to recruit creativity online raters through online platforms, social media and email groups. A variety of groupings were targeted, including those with experience in dance, those with experience in creative fields, and those who had no experience in dance and/or creativity. Participants completed comprehensive demographic questions to provide information about their background and training in dance, creativity and the arts. They then watched a randomly selected video, before completing the four assessment scales (creativity, liking, technique and meaning), which appeared in a random order. Each participant had the option to watch as many clips as they wished to, before completing the four scales at the end of each piece.

After 6 weeks, we had obtained sufficient data. Data were downloaded to Microsoft Excel and cleaned and screened, where participants with missing data or insufficient information were removed. We then transferred data into the Statistical Package for Social Sciences Version 23 (IBM Corp, 2016), and undertook preliminary analyses of variance and correlation. We conducted main analyses using the LAVAAN package (Rosseel, 2012) within R version 3.2 (R Core Team, 2015). A repeated measures linear mixed model was used to predict creativity score and determine the impact of experience and expertise at the nine levels. We used

TABLE 1 | Participant experience and expertise in dance and creativity.

Experience/expertise	No (N)	Yes (N)
Experience in child/adult dance classes	$n = 166$	$n = 684$
Experience in child/adult contemporary dance classes	$n = 444$	$n = 406$
Current/previous attendance at the dance institution	$n = 763$	$n = 87$
Experience in watching live contemporary dance	$n = 504$	$n = 346$
Experience in choreographing dance	$n = 262$	$n = 588$
Experience choreographing contemporary dance	$n = 605$	$n = 245$
Employed in any creative domain	$n = 383$	$n = 467$
Employed in an artistic, creative domain	$n = 483$	$n = 367$
Are you an expert in creativity?	$n = 714$	$n = 136$

a repeated measures mixed model as it is suitable for missing data, therefore allowing for the variation in the number of videos observed.

RESULTS

Descriptive Statistics

The numbers of videos viewed by each of the 850 creativity raters ranged from one to 21 videos ($M = 2.53$, $SD = 2.63$). In total, we obtained 2153 individual ratings with between 81 and 102 creativity ratings on each video ($M = 91.61$, $SD = 6.37$). Descriptive statistics of overall ratings from the 23 videos are shown in **Table 2**.

Preliminary Analyses

We undertook a series of one-way ANOVAS to determine a difference in the mean ratings of the videos. Creativity [$F(22,2130) = 6.85$, $p < 0.001$], likeability [$F(22,2130) = 5.90$, $p < 0.001$], meaning [$F(22,2130) = 4.77$, $p < 0.001$] and technique [$F(22,2130) = 11.44$, $p < 0.001$] all showed significant variation in scores between videos.

Next, Pearson's correlation analyses were conducted to obtain an understanding of the relationships between creativity, likeability, technique and meaning. **Table 3** shows significant moderate to strong positive correlations between all four variables, suggesting that people rate contemporary dance highly on creativity when it is also perceived as liked, well understood and well executed.

Repeated Measures Linear Mixed Model

A colleague of the authors' who was blind to the purpose of the research coded a random sample of 50 participants' qualitative employment responses 'Employed in any creative domain' and 'Employed in an artistic creative domain' to assess the reliability of the expertise and experience coding seen in **Table 1**. A positive inter-rater reliability (IRR) correlation = 0.83 was achieved.

TABLE 2 | Descriptive statistics of creativity, likeability, meaning and technique ratings.

	Min.	Max.	Mean	SD	Skewness	Kurtosis
Creativity	1.0	7.0	4.62	1.44	-0.45	-0.17
Likeability	1.0	7.0	4.13	1.61	-0.21	-0.67
Meaning	1.0	7.0	3.86	1.64	-0.17	-0.78
Technique	1.0	7.0	4.97	1.36	-0.55	-0.05

TABLE 3 | Pearson's correlation coefficients for creativity, likeability, meaning and technique ratings.

	Creativity	Likeability	Meaning
Creativity			
Likeability	0.71*		
Meaning	0.60*	0.67*	
Technique	0.62*	0.57*	0.45*

*Correlation is significant at the 0.05 level (2-tailed).

According to Cohen's Kappa statistic, an IRR of greater than 0.8 indicates a very good level of reliability between raters (McHugh, 2012).

We entered each of the experience or expertise categories in to the repeated measures linear mixed model as a fixed effect. Contemporary choreographic experience significantly predicted creativity, $F(1,2052.33) = 6.61$, $p < 0.001$, as did self-attributed creative expertise $F(1,2032.13) = 17.82$, $p < 0.001$, but none of other categories were significant. In those with contemporary choreographic experience, creativity was rated higher compared to the reference group [$b = 0.24$, $t(2067.48) = 2.71$, $p < 0.05$ (95% CI = -0.044 to 0.39)]. In those with self-attributed creative expertise, creativity was rated lower compared to the reference group [$b = -0.13$, $t(2032.13) = -4.44$, $p < 0.001$ (95% CI = -0.52 to -0.19)]. Next, video was entered as a random effect. Both the intercept ($b = -0.19$, $Wald Z = 32.56$, $p < 0.001$) and video were significant ($b = -0.12$, $Wald Z = 2.81$, $p < 0.05$), indicating that slopes were significantly different across the 23 videos.

DISCUSSION

We explored the role of experience and expertise in assessing choreographic creativity, using a novel online methodology that facilitated dance specific research. 850 assessors assessed creativity in 23 individual contemporary dance choreographies. Assessor experience and expertise were sampled from a continuum of expertise from those who had never taken a dance class to professional choreographers. The results demonstrate the impact of both dance specific experience and broader creative expertise in the assessment of choreographic creativity.

The results show that when an individual has experience in choreography, they rate creativity higher. That is, one needs experience in the choreographic process to judge a piece to be more creative. This supports the idea by Corazza (2016) that creativity is related to an ability to see the potential expression of a process. This is in line with the emphasis on the creative process in dance pedagogy (Butterworth, 2004; Farrer, 2014), yet suggests that this emphasis may be preventing those who do not have experience of choreography from identifying creativity. Our findings suggest that this level of expertise is essential in evaluator selection; experience in physically dancing or watching contemporary dance does not lead an individual to rate creativity higher. Instead, experience in knowing the process of making dance allows an individual to judge a piece as more creative.

These findings have implications with regards to accessibility of contemporary dance, in suggesting that training in dance *per se* does not necessarily facilitate an understanding of choreographic creativity, but that only those who learnt to make dance understand and rate higher. The level of expertise suggested by our findings regarding creativity is more specific than that which has been reported in the literature on dance performance, even beyond those studies involving fMRI recordings of audience responses (e.g., Calvo-Merino et al., 2005). Here, physical participation has led to significant differences in brain activity when watching dance. However, our findings indicate that experience of making or choreographing, beyond physical

participation in dancing, impacts creativity assessment (e.g., Calvo-Merino et al., 2005).

The results of the correlational analyses showed that creativity score is related to choreography that the evaluator likes, can find meaning in, and is performed by a dancer whom the evaluator perceives as technically skilled. Collectively, these correlational results indicate that an audience evaluates creativity in line with subjective elements which go beyond the criteria which underpin problem-solving tests such as the RAT (Mednick and Mednick, 1971) and TTCT (Torrance, 1974). Standard creativity tests previously used in dance, operationalise creativity by the ability to rapidly produce a large number of infrequent responses (e.g., Fink and Woschnjak, 2011). Two critical elements of creativity underpin most theoretical and research-based definitions) *originality* or *novelty* and b) *usefulness* or *appropriateness* (Stein, 1953; Barron, 1955; Amabile, 1983;). This dualistic criterion remains the most commonly accepted definition of creativity (Runco and Jaeger, 2012). Since creativity correlated highly with making meaning of the piece, one could argue that those who rated higher in the contemporary dance choreography subgroup had a clearer insight into the meaning of the work, because they had experience of the process and understood intention. Creativity in the arts may be assessed concerning intention at the moment of creation, with a proposition that it is intentionality rather than novelty which is vital (Kharkhurin, 2014; Weisberg, 2015). In turn, this supports previous authors who have discussed the lack of outsider dance audiences and the failure to understand contemporary dance (Van Dyke, 2010).

A second finding was that scores by those who self-assigned themselves as creative experts were lower than those who did not. This supports the value of the chosen method, and that asking judges to self-select whether they are an expert may be valuable when seeking to recruit judges. Experts will have had considerably greater exposure to creativity and therefore do not consider the work to be as creative; there is some interaction of expertise at this level, yet cause and effect cannot be established.

The implications for these findings are numerous when discussing the need for widening audience engagement in contemporary dance. These findings may imply a need for educating audiences about creative processes underpinning the dance product. Glass and Stevens (2005) note that '*Priming audience members about a particular work should assist them to engage with the work at a greater level of understanding*' (p.17). Educating an audience about the creative process might bridge the gap between the audience's understanding of creativity in dance and subsequent enjoyment of the work. This may be particularly true in an art form where the emphasis is on the process and the dancer's experience of making or creating a dance for the dancer's enjoyment (Lavender, 2009).

Importantly, the results of the analyses showed variation in the mean ratings of the videos, demonstrating that the snowball sampling method does not neutralize differences; that is, a varied audience collectively distinguish varying levels of creativity. Using a simple Likert scale for the CAT is therefore advocated as a simple yet effective measure of creativity. We recognize that there are numerous ways of implementing the CAT and the

present research was a considerable variation on the original. The use of this variation was beneficial since it is arguably the *only* available research methodology for creativity which is not inherently tied to a theory of creativity but facilitated a means of assessing dance specific creativity (Baer and McKool, 2009). The methodology assessed the manifestation of creativity through the body (Kirsh, 2010 and without pen and paper tests, while focusing on product also increased validity.

It is of note to consider the relationship between the choreographer and the dancer who is performing the work. Whilst our intention was to assess the choreographer's creativity, one could argue that the audience perception may also be related to the performer's creative interpretation of the movement, in the same way that it is related to their technical skill. Thus an additional facet in dance may be the dancer's ability to communicate and interpret the choreographic interpretation which is as important as the choreographer's creative skill at constructing the work (Smith-Autard, 2014).

The study is strengthened by the inclusion of 23 videos and a large sample of respondents, allowing a more substantial variation of scores to be given and to facilitate a broad audience, which is more reminiscent of real-life choreographic settings. Future research should endeavor to establish reliability amongst experts in dance specific creativity which is solely reliant on expert opinions, such as auditions. The present research was not intended to undertake IRR correlation analysis; however, IRR between experts has been highlighted as methodologically important (Kaufman et al., 2008; Haller et al., 2011). Furthermore, there is debate regarding the width of the Likert scale, with no consistent recommendations, aside from to include a neutral point. Thus, findings are not comparable across studies. However, in sum, although the method underpinning the CAT may be perceived to lack methodological stability, the breadth of application and validity has been demonstrated.

We had 87 (of 850) participants who currently/previously attended the institution, so we added 'current/previous attendance at the institution' as a predictor. This was not significant, thus did not impact on creativity ratings. Therefore the possibility of this as a confound was deemed to be minor, since only a small number of participants were potential classmates and this did not have a significant impact. In addition, although we did not ask whether the viewer knew the performer in the video, the video appeared randomly, so if they knew any performer, there was a 1 in 23 chance of them knowing the performer on video 1, 1 in 22 chance for video 2 and so forth. Since the average views were 2.5, the chances of knowing the performer were again relatively small.

The online methodology and use of snowballing enabled meaningful participant diversity, which was also sensitive to differences both in expertise and in evaluations of the videos. We recognize that snowballing can result in the loss of crucial information over participants, however, for the present research it facilitated a meaningful audience-like participant set. The use of such an online evaluation might facilitate repeated testing over time. Previous efforts to research dancers' creativity focused on domain-general measures and tended to be cross-sectional in nature; longitudinal research looking at the impact of the

environment or training on dancers' creativity has not yet been possible (e.g., Kalliopuskka, 1989; Stinson, 1993; Fink and Woschnjak, 2011). Although we note that there are limitations of online methodologies, such as being unable to establish reliability between evaluators (as is common in the original version), the results of the study support the viability of an online snowball sampling method to recruit both experts and non-experts. In particular, the effectiveness of adapting the CAT for research purpose is advocated.

The present online adaption has strength in its flexibility for use across many unique domains of creativity. Thus, by assuming neither domain generality nor specificity, it is a method which could be replicated using any creative performances or artifact across many arts such as music, or visual art, allow recruitment of both large samples of creative works and raters. In this variation, a methodological strength was that unknown to the raters, the individual performing the work was not the creator. Future research within the domain of dance should continue to use the CAT in its most original form, aiming to establish reliability between assessors in real life creative performance scenarios such as an audition, to understand selection methods, as well as evaluation of students in choreography and improvisation courses.

CONCLUSION

This research aimed to understand the role of expertise in assessing creativity in choreographic creativity. A secondary aim

was to use a large scale online methodology which went beyond the pen and paper problem-solving approaches which have predominated the literature. The use of choreographic videos allowed the expression of embodied creativity and recruitment of a large audience with varying degrees of expertise and experience in dance. The results showed that personal experience of the creative process increased ratings of creativity, while creative experts rated creativity lower. The use of online methodologies for assessing creativity is advocated across multiple domains of creativity.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Trinity Laban Ethics Guidelines and the BPS Code of Ethics. The protocol was approved by the Trinity Laban Conservatoire of Music and Dance Ethics Committee. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

LC led the work carried out on the paper, including the writing, research design, data collection, and analysis. JM supported the development and undertaking of statistical analyses. NLS assisted in developing the methodology. ER and JM were Ph.D. supervisors of the work.

REFERENCES

- Amabile, T. M. (1982). Social psychology of creativity: a consensual assessment technique. *J. Pers. Soc. Psychol.* 43, 997–1013. doi: 10.1037/0022-3514.43.5.997
- Amabile, T. M. (1983). The social psychology of creativity: a componential conceptualisation. *J. Pers. Soc. Psychol.* 45, 357–376. doi: 10.1037/0022-3514.45.2.357
- Amabile, T. M. (1996). *Creativity in Context: Update to the Social Psychology of Creativity*. Boulder, NV: Westview Press.
- Amabile, T. M., and Pillemer, J. (2012). Perspectives on the social psychology of creativity. *J. Creat. Behav.* 46, 3–15. doi: 10.1002/jocb.001
- Baer, J. (1998). The case for domain specificity of creativity. *Creat. Res. J.* 11, 173–177. doi: 10.1207/s15326934crj1102_7
- Baer, J., and McKool, S. S. (2009). "Assessing creativity using the consensual assessment technique," in *Handbook of Research on Assessment Technologies, Methods, and Applications in Higher Education*, ed. C. Schreiner (Lawrenceville, NJ: Rider University), 65–77.
- Barron, F. (1955). The disposition towards originality. *J. Abnorm. Soc. Psychol.* 51, 478–485. doi: 10.1037/h0048073
- Bläsing, B., Puttke, M., and Schack, T. (Eds). (2010). *The Neurocognition of Dance: Mind, Movement and Motor Skills*. New York, NY: Psychology Press.
- Burrows, J. (2010). *A Choreographer's Handbook*. New York, NY: Routledge.
- Butterworth, J. (2004). Teaching choreography in higher education: a process continuum model. *Res. Dance Educ.* 5, 45–67. doi: 10.1080/1464789042000190870
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., and Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cereb. Cortex* 15, 1243–1249. doi: 10.1093/cercor/bhi007
- Corazza, G. E. (2016). Potential originality and effectiveness: the dynamic definition of creativity. *Creat. Res. J.* 28, 258–267. doi: 10.1080/10400419.2016.1195627
- Csikszentmihalyi, M. (1999). "Implications of a systems perspective for the study of creativity," in *Handbook of Creativity*, ed. R. J. Sternberg (Cambridge: Cambridge University Press), 313–335.
- Csikszentmihalyi, M. (2014). *Society, Culture, and Person: A Systems View of Creativity*. Dordrecht: Springer.
- Farrer, R. (2014). The creative dancer. *Res. Dance Educ.* 15, 95–104. doi: 10.1080/14647893.2013.786035
- Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Pers. Soc. Psychol. Rev.* 2, 290–309. doi: 10.1207/s15327957pspr0204_5
- Fink, A., and Woschnjak, S. (2011). Creativity and personality in professional dancers. *Pers. Individ. Diff.* 51, 754–758. doi: 10.1016/j.paid.2011.06.024
- Glass, R., and Stevens, C. (2005). *Making Sense of Contemporary Dance: An Australian Investigation into Audience Interpretation and Enjoyment Levels*. Sydney, NSW: University of Sydney.
- Glăveanu, V. P. (2010). Paradigms in the study of creativity: introducing the perspective of cultural psychology. *New Ideas Psychol.* 28, 79–93. doi: 10.1016/j.newideapsych.2009.07.007
- Guilford, J. P. (1956). The structure of intellect. *Psychol. Bull.* 53, 267. doi: 10.1037/h0040755
- Guilford, J. P. (1950). Creativity. *Am. Psychol.* 5, 444–454.
- Hagood, T. K. (2001). Dance to read or dance to dance? *Arts Educ. Policy Rev.* 102, 27–29. doi: 10.1080/10632910109600014
- Haller, C. S., Courvoisier, D. S., and Cropley, D. H. (2011). Perhaps there is accounting for taste: evaluating the creativity of products. *Creat. Res. J.* 23, 99–109. doi: 10.1080/10400419.2011.571182
- H'Doubler, M. N. (1998). *Dance: A Creative Art Experience*. Madison, WI: University of Wisconsin press.
- Hong, S. W., and Lee, J. S. (2015). Nonexpert evaluations on architectural design creativity across cultures. *Creat. Res. J.* 27, 314–321. doi: 10.1080/10400419.2015.1087245

- Hu, W., and Adey, P. (2002). A scientific creativity test for secondary school students. *Int. J. Sci. Educ.* 24, 389–403. doi: 10.1080/09500690110098912
- Humphrey, D. (1959). *The Art of Making Dances*. New York, NY: Grove Press.
- IBM Corp (2016). *IBM SPSS Statistics: Version 23*.
- Julmi, C., and Scherm, E. (2015). The domain-specificity of creativity: insights from new phenomenology. *Creat. Res. J.* 27, 151–159. doi: 10.1080/10400419.2015.1030310
- Kalliopuska, M. (1989). Empathy, self-esteem and creativity among junior ballet dancers. *Percept. Mot. Skills* 69, 1227–1234. doi: 10.2466/pms.1989.69.3f.1227
- Kaufman, J. C., and Baer, J. (2004). Sure, I am creative—but not in mathematics!: Self-reported creativity in diverse domains. *Emp. Stud. Arts* 22, 143–155. doi: 10.2190/26HQ-VHE8-GTLN-BJMJ
- Kaufman, J. C., Baer, J., Cole, J. C., and Sexton, J. D. (2008). A comparison of expert and non-expert raters using the consensual assessment technique. *Creat. Res. J.* 20, 171–178. doi: 10.1080/10400410802059929
- Kaufman, J. C., and Sternberg, R. J. (2010). *The Cambridge Handbook of Creativity*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511763205
- Kharkhurin, A. V. (2014). Creativity. 4in1: four-criterion construct of creativity. *Creat. Res. J.* 26, 338–352. doi: 10.1080/10400419.2014.929424
- Kirsh, D. (2010). Thinking with external representations. *AI Soc.* 25, 441–454. doi: 10.1007/s00146-010-0272-8
- Kirsh, D. (2011). “Creative cognition in choreography,” in *Proceedings of the 2nd International Conference on Computational Creativity*, Mexico, 1–6.
- Kogan, N. (2002). Careers in the performing arts: a psychological perspective. *Commun. Res. J.* 14, 1–16. doi: 10.1207/S15326934CRJ1401_1
- Kokotsaki, D., and Newton, D. P. (2015). Recognising creativity in the music classroom. *Int. J. Music Educ.* 3, 491–508. doi: 10.1177/0255761415607081
- Kozbelt, A. (2004). Originality and technical skill as components of artistic quality. *Empirical studies of the arts* 22, 157–170. doi: 10.2190/NDR5-G09N-X7RE-34H7
- Lavender, L. (2009). Dialogical practices in teaching choreography. *Dance Chronicle* 32, 377–411. doi: 10.1080/01472520903276735
- Long, H., Plucker, J. A., Yuc, Q., Ying, D., and Kaufman, J. C. (2014). Research productivity and performance of journals in the creativity sciences: a bibliometric analysis. *Creat. Res. J.* 26, 353–360. doi: 10.1080/10400419.2014.929425
- Lubart, T. I. (2001). Models of the creative process: past, present and future. *Creat. Res. J.* 13, 295–308. doi: 10.1207/S15326934CRJ1334_07
- Łucznik, K. (2015). Between minds and bodies: some insights about creativity from dance improvisation. *Technoetic Arts* 13, 301–308. doi: 10.1386/tear.13.3.301_1
- May, J., Calvo-Merino, B., Delahunta, S., McGregor, W., Cusack, R., Owen, A. M., et al. (2011). Points in mental space: an interdisciplinary study of imagery in movement creation. *Dance Res.* 29(Suppl.), 404–432. doi: 10.3366/drs.2011.0026
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochem. Med.* 22, 276–282. doi: 10.11613/BM.2012.031
- Mednick, S. A., and Mednick, M. (1971). *Remote Associates Test: Examiner's Manual*. Boston, MA: Houghton Mifflin.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reason, M., and Reynolds, D. (2010). Kinesthesia, empathy, and related pleasures: an inquiry into audience experiences of watching dance. *Dance Res. J.* 42, 49–75. doi: 10.1017/S0149767700001030
- Risner, D. (2000). Making dance, making sense: epistemology and Choreography. *Res. Dance Educ.* 1, 155–172. doi: 10.1080/713694259
- Rosseel, Y. (2012). Lavan: an R package for structural equation modelling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02
- Runco, M. A. (2014). *Creativity: Theories and themes: Research, development, and practice*. San Diego, CA: Elsevier.
- Runco, M. A., and Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creat. Res. J.* 24, 66–75. doi: 10.1080/10400419.2012.652929
- Runco, M. A., and Jaeger, G. J. (2012). The standard definition of creativity. *Creat. Res. J.* 24, 92–96. doi: 10.1080/10400419.2012.650092
- Sawyer, R. K. (2014). *Group creativity: Music, Theatre, Collaboration*. New York, NY: Psychology Press.
- Simonton, D. K. (1999). *Origins of Genius: Darwinian Perspectives on Creativity*. Oxford: Oxford University Press.
- Simonton, D. K. (2015). On praising convergent thinking: creativity as blind variation and selective retention. *Creat. Res. J.* 27, 262–270. doi: 10.1080/10400419.2015.1063877
- Smith-Autard, J. M. (2014). *Dance Composition: A Practical Guide to Creative Success in Dance Making*. New York, NY: Routledge. doi: 10.4324/9781315060033
- Stein, M. I. (1953). Creativity and culture. *J. Psychol.* 36, 311–322. doi: 10.1080/00223980.1953.9712897
- Stevens, C., Glass, R., Schubert, E., Chen, J., and Winskel, H. (2007). “Methods for measuring audience reactions,” in *Proceedings of the inaugural International Conference on Music Communication Science*, Sydney, NSW, 155.
- Stevens, C., Malloch, S., and McKechnie, S. (2001). Moving mind: the cognitive psychology of contemporary dance. *Brolga* 12, 7–14.
- Stevens, C., and McKechnie, S. (2005). Thinking in action: thought made visible in contemporary dance. *Cogn. Process.* 6, 243–252. doi: 10.1007/s10339-005-0014-x
- Stevens, C. J., Schubert, E., Morris, R. H., Frear, M., Chen, J., Healey, S., et al. (2009). Cognition and the temporal arts: investigating audience response to dance using PDAs that record continuous data during live performance. *Int. J. Hum. Comput. Stud.* 67, 800–813. doi: 10.1016/j.ijhcs.2009.06.001
- Stinson, S. W. (1993). Testing creativity of dance students in the People's Republic of China. *Dance Res. J.* 25, 65–68. doi: 10.1017/S0149767700008056
- Strauss, M. R., and Nadel, M. (2012). *Looking at Contemporary Dance: A Guide for the Internet Age*. Princeton, NJ: Princeton Book Company.
- Thomson, P., and Jaque, S. V. (2017). *Creativity and the Performing Artist*. San Diego, CA: Elsevier.
- Torrance, E. P. (1974). *Norms-technical Manual: Torrance Tests of Creative Thinking*. Lexington, MA: Ginn & Company.
- Van Dyke, J. (2010). Vanishing: dance audiences in the postmodern age. *Dance Chronicle* 33, 208–230. doi: 10.1080/01472526.2010.485902
- Wallas, G. (1926). *The Art of Thought*. London: CAPE.
- Weisberg, R. W. (2015). On the usefulness of “value” in the definition of creativity. *Creat. Res. J.* 27, 111–124. doi: 10.1080/10400419.2015.1030320
- Williams, R., Runco, M. A., and Berlow, E. (2016). Mapping the themes, impact, and cohesion of creativity research over the last 25 years. *Creat. Res. J.* 28, 385–394. doi: 10.1080/10400419.2016.1230358
- Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi: 10.3758/BF03196322

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 Clements, Redding, Lefebvre Sell and May. 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.



The Improvisational State of Mind: A Multidisciplinary Study of an Improvisatory Approach to Classical Music Repertoire Performance

David Dolan¹, Henrik J. Jensen^{2,3}, Pedro A. M. Mediano⁴, Miguel Molina-Solana^{4,5}, Hardik Rajpal², Fernando Rosas^{2,6} and John A. Sloboda^{1*}

¹ Guildhall School of Music and Drama, London, United Kingdom, ² Department of Mathematics, Centre of Complexity Science, Imperial College London, London, United Kingdom, ³ Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan, ⁴ Department of Computing, Imperial College London, London, United Kingdom, ⁵ Data Science Institute, Imperial College London, London, United Kingdom, ⁶ Department of Electrical and Electronic Engineering, Imperial College London, London, United Kingdom

OPEN ACCESS

Edited by:

William Forde Thompson,
Macquarie University, Australia

Reviewed by:

Eleonora Concina,
Università degli Studi di Padova, Italy
Glenna Batson,
Wake Forest University, United States

*Correspondence:

John A. Sloboda
john.sloboda@gsmd.ac.uk

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 30 April 2018

Accepted: 12 July 2018

Published: 25 September 2018

Citation:

Dolan D, Jensen HJ, Mediano PAM,
Molina-Solana M, Rajpal H, Rosas F
and Sloboda JA (2018) The
Improvisational State of Mind: A
Multidisciplinary Study of an
Improvisatory Approach to Classical
Music Repertoire Performance.
Front. Psychol. 9:1341.
doi: 10.3389/fpsyg.2018.01341

The recent re-introduction of improvisation as a professional practice within classical music, however cautious and still rare, allows direct and detailed contemporary comparison between improvised and “standard” approaches to performances of the same composition, comparisons which hitherto could only be inferred from impressionistic historical accounts. This study takes an interdisciplinary multi-method approach to discovering the contrasting nature and effects of prepared and improvised approaches during live chamber-music concert performances of a movement from Franz Schubert’s “Shepherd on the Rock,” given by a professional trio consisting of voice, flute, and piano, in the presence of an invited audience of 22 adults with varying levels of musical experience and training. The improvised performances were found to differ systematically from prepared performances in their timing, dynamic, and timbral features as well as in the degree of risk-taking and “mind reading” between performers, which included moments of spontaneously exchanging extemporized notes. Post-performance critical reflection by the performers characterized distinct mental states underlying the two modes of performance. The amount of overall body movements was reduced in the improvised performances, which showed less unco-ordinated movements between performers when compared to the prepared performance. Audience members, who were told only that the two performances would be different, but not how, rated the improvised version as more emotionally compelling and musically convincing than the prepared version. The size of this effect was not affected by whether or not the audience could see the performers, or by levels of musical training. EEG measurements from 19 scalp locations showed higher levels of Lempel-Ziv complexity (associated with awareness and alertness) in the improvised version in both performers and audience. Results are discussed in terms of their potential support for an “improvisatory state of mind” which may have aspects of flow (as characterized by Csikszentmihalyi, 1997) and primary states

(as characterized by the Entropic Brain Hypothesis of Carhart-Harris et al., 2014). In a group setting, such as a live concert, our evidence suggests that this state of mind is communicable between performers and audience thus contributing to a heightened quality of shared experience.

Keywords: improvisation, classical performance, musical communication, neural complexity, motion analysis, state of mind, classical improvisation, flow

INTRODUCTION

Motivation

Although classical music performance is recognized as a creative practice, its parameters have been restricted by a longstanding ethos of “faithfulness to the composer’s score” which limit the bounds of acceptable deviation (Leech-Wilkinson, 2016). This ethos has dominated classical music performance since the late nineteenth century. However, historical research has revealed that Western art-music composers from Bach, through Mozart and Beethoven and onwards into the romantic era expected and encouraged performers to creatively depart from the score in a far more radical way than is common today, including the insertion of new notes (Eigeldinger, 1986; Hamilton, 2008).

In these earlier times improvisation was not only encouraged, but it was believed by many to be an essential component of complete musicianship and mastery. For instance, improviser and composer Johann Nepomuk Hummel (1778–1837) recommended “free improvisation in general and every respectable form to all those for whom [music] is not merely a matter of entertainment and practical ability, but rather principally one of inspiration and meaning in their art” (quoted in Goertzen, 1996, p. 305)¹. Hummel stated in 1828 that this matter was urgent, and cautioned, “Even if a person plays with inspiration but also from a written score, he or she will be much less nourished, broadened, and educated than through the frequent immersion in free fantasy practiced in the full awareness of certain guidelines and directions, even if this improvisation is only moderately successful” (Goertzen, 1996).

In recent years there has been a renaissance and awakening of interest in practicing, teaching, learning, and researching Western classical music improvisation (e.g., Berkowitz, 2010). For example, while in most high profile international competitions improvising repeats, preludes, fermata points or cadenzas is still considered by competitors to be an unwise risk, the Bach international piano competition in Leipzig (under the artistic direction of Robert Levin) encourages it explicitly, by saying in the instructions to competitors that extemporized repeats are welcome and encouraged (<http://www.bachwettbewerb-leipzig.de/en/bach-competition/competition-programme-2018>).

Improvisation is beginning to find its way into the pedagogical curriculum for music (Azzara and Snell, 2016). However, this is still sufficiently uncommon for Shehan Campbell et al. (2014) to be able to conclude “That the majority of music students graduate with little to no experience, let alone significant grounding, in

the essential creative processes of improvisation and composition represents one of the most startling shortcomings in all of arts education”.

The re-insertion of this “improvisatory approach” into classical music professional practice is sufficiently new that the contemporary practitioners of this approach have predominantly been schooled in the mainstream approach of score faithfulness, and switch between the two approaches in their artistry, thus affording researchers the possibility of comparing the nature and effects of improvised performances with “conventional” performances of the same pieces by the same performers.

It is this unique juncture in artistic history which has motivated and enabled us to investigate exactly what it is that differentiates the improvisatory approach to performance from the conventionally prepared one, in its nature, its cognitive and neural underpinnings, and its effects.

Background

Most of the recent scientific investigations into musical improvisation have centered on jazz. These studies have analyzed improvisation using tools from neuroscience (Donnay et al., 2014; Pinho et al., 2014; Lopata et al., 2017), musicology (Norgaard, 2011, 2014) and psychology (Tervaniemi et al., 2016; Love, 2017). Although these research efforts are relevant to broaden our understanding of improvisation in music, it is not straightforward how to isolate the effect of improvisation as there is no natural baseline to compare with. Improvisation is a fundamental and omnipresent ingredient in Jazz music and therefore is to be expected that Jazz musicians and listeners will have a preference for it. In contrast, in classical music the default choice for the last 100 years is to perform without improvisational elements.

The distinctive feature of classical music improvisation (at least in the present day) is the existence of a strong canonical form (usually represented by a written score and well known within the community of listeners) from which improvisation is a deliberate deviation. Faithfulness to the canonical score is also a valid artistic response, whereas within other artistic forms, such as Jazz, the faithful rendition of a “cover” melody would be considered of little artistic interest. For a more detailed discussion of the nature of classical improvisation see Dolan et al. (2013, pp. 1–6). Although very few existing studies examine improvisation in the context of Western classical art-music, a notable exception is Després et al. (2017), who explore strategies applied by five internationally recognized classical music solo improvisers by means of analyzing semi-structured retrospective interviews. However, this study did not gather any data from

¹ We thank Robert Levin for referring us to this document.

actual performances and therefore sheds only indirect light on performance characteristics and audience response.

Improvisation is a listener-directed art, and so it is critical to our understanding of it to know what effect it has on listeners/audiences. There are many anecdotal and historical accounts of the power and impact of improvisatory performances of classical music, such as the report of the “tumultuous applause” that greeted a 30 min improvisation by Mozart in Prague in 1787 (Johann Nepomuk Stiepanek, reported in Abert 2007, p. 827). However, very few studies have attempted to investigate the impact on traditional concert audiences of listening to live performances of classical music that vary in their expressive intent. Some studies have measured the subjective responses of audience members via questionnaires and/or interviews (Pitts, 2005; Thompson, 2006, 2007; Pitts and Spencer, 2007; Dobson, 2008), but none of them directly addresses responses to the improvised or spontaneous elements of the performance. Also, the substantial neuroscience literature on the relationship between music and language (see Hutka et al., 2013 and references therein) focuses on sensory and semantic processing of individuals, and does not address the interaction between performers and listeners. Studies measuring brain activity of individuals listening to improvised Western classical music hardly exist. The only available data to date come from a pilot study reported by Dolan et al. (2013) (further analyzed in Wan et al., 2014), which studied the effect of conventional and improvised live performances of pieces from the classical repertoire on both musicians and listeners. The results showed significant differences in performance features, subjective experience and brain activity between prepared and improvised performance, providing initial evidence that improvised performances of the classical repertoire can heighten musical effectiveness and audience response.

Understanding the Improvisatory Approach as a State of Mind

In this study we explore and elaborate the notion that improvisational activity induces a particular state of mind in performers and audience different from that habitually present in prepared performances. By “state of mind” we refer to a distinct mental and neural configuration which may be maintained for a period of time, and which involves specific cognitive and affective components. We seek to shed light on how might such a state be best characterized, how it relates to other states of mind, and how and in what ways such a state is communicable or transferable to listeners. We consider two separate but related lines of prior empirical enquiry as of particular relevance.

One is the body of investigation into the multidimensional phenomenon known as *Flow*, as introduced into Psychology by Csikszentmihalyi (1975). Originally described as “the holistic sensation that people feel when they act with total involvement,” this state of mind is characterized by full engagement, sensation of creativity combined with enhanced well-being, effortless control and concentration, a sense of having clear goals and full presence in one’s performance together with a reduced awareness of the time passing (Chirico et al., 2015). Moreover, *flow* is to be distinguished from creativity, the latter meaning the creation of

novelty while the former refers to an effortless yet highly focused state of consciousness (Csikszentmihalyi, 1996).

There exists a close relationship between the state of flow and music experience. In fact, it has been claimed that music is the activity in which it is easiest to reach an experience of flow (Csikszentmihalyi, 1997; Lowis, 2002). Chirico et al. (2015) review recent investigations into the relationship between music and flow, covering musical performance, composition and listening. Improvisation as a source of flow has been neglected, although Després et al. (2017), suggest that Berkowitz (2010) characterization of a “witness” state of mind in the solo classical improvisation of Robert Levin and Malcolm Bilson may hint at elements of flow. In the “creator” state, a musician develops the improvisation consciously and deliberately, using declarative knowledge. In the “witness” state, the improviser is more akin to a spectator of his or her own unfolding improvisation which emerges through implicit procedural knowledge. However, in both Berkowitz (2010) and Després et al. (2017) investigations of solo classical improvisation, the data came from extended retrospective interviews separated from any specific performance, and thus not optimal for uncovering evidence of flow states which, by definition, are “in the moment.” In addition there was no consideration in any prior studies of how such states may be shared between musicians in group improvisation or communicated to listeners. There is thus much still to discover about the way that different levels of conscious awareness guide the real-time decision making process.

A second line of enquiry comes from work into the “entropic brain hypothesis” (Carhart-Harris et al., 2014). Combining recent neuroimaging findings with psychoanalytic concepts, the EBH distinguishes between two different styles of human cognition: *secondary states* that are characteristic of the experience of contemporary adult humans, and *primary states* to which the mind regresses under specific conditions, e.g., in response to severe stress, psychedelic drugs or in REM sleep. Physiologically, primary states are characterized by an elevated entropy in various brain function that is manifested in e.g., fMRI or EEG measurements with high signal complexity, which correlates with diversity and richness of experiential content. Conversely, entropy is suppressed in secondary states generating measurements with lower signal complexity and hence more regular and stable cognitive processes, hence enabling metacognitive functions including reality-testing and self-awareness.

The EBH further hypothesize that primary states are evolutionarily older than secondary states:

“... the mind has evolved (via secondary consciousness upheld by the ego) to process the environment as precisely as possible by finessing its representations of the world so that surprise and uncertainty (i.e., entropy) are minimized. ... In contrast, in primary states, cognition is less meticulous in its sampling of the external world and is instead easily biased by emotion, e.g., wishes and anxieties.” (Carhart-Harris et al., 2014)

However, although primary consciousness may be a sub-optimal mode of cognition, it seems to be more than a mere psychological atavism. Plenty of reports show how events involving primary

states can bring deep experiences and have profound therapeutic effects (Griffiths et al., 2008; Carhart-Harris and Nutt, 2010; MacLean et al., 2011). In effect, the high entropy of primary states seems to allow overcoming the inability to think and behave in a flexible manner, narrow-mindedness and aggressive self-critical attitudes.

Although the EBH was developed to provide a theoretical basis for therapeutic uses of psychedelic drugs, it is natural to ask if it is applicable to the domain of musical experience, and in particular musical improvisation. Is the improvisational state of mind a primary state? Could one find traces of primary states in musicians and audience during improvisational activities?

Scope of the Present Study

The present study aimed to answer these questions, building on Dolan et al. (2013), and addressing a number of key shortcomings and limitations.

A first limitation of Dolan et al. (2013) study was that it employed a traditional EEG analysis to track the activation of various cortical areas related to alpha and beta frequency bands. In contrast, the present study focuses on the *Lempel-Ziv complexity* (LZ) of the EEG signals, which is the preferred method for studying brain entropy and signal complexity within the EBH framework (Carhart-Harris, 2018). The method was introduced by Abraham Lempel and Jacob Ziv to study the complexity of binary sequences (Ziv, 1978), and was later extended for EEG signals to study epilepsy (Radhakrishnan and Gangadhar, 1998) and depth of anesthesia (Zhang et al., 2001). When characterizing states of mind, LZ is higher in subjects during wakeful rest than in subjects during sleep or general anaesthesia (Casali et al., 2013; Schartner et al., 2015). LZ is also higher than normal when the brain is under the effect of psychedelic substances (Schartner et al., 2017). Even at the individual level, LZ is correlated with a more vivid imagination and ego dissolution (Schartner et al., 2017). Also, the brain's response to a given stimulus scores higher LZ when the stimulus is more meaningful to the viewer (Boly et al., 2015). In summary, there is strong evidence in the literature that suggests that LZ is a reliable indicator of awareness and alertness.

A second limitation of the pilot study concerned the composition of the audience, which was primarily drawn from highly-trained students and staff of a conservatoire. It is possible that the significant effects of improvisation could result from a sophisticated level of musical training and awareness, and would not be generalizable to a broader public. In order to better characterize the impact of improvisation on the listening population, an audience containing a wider range of musical knowledge and experience is needed.

Thirdly, informal observations by Dolan et al. (2013) suggested that musicians engaged in larger bodily gestures during the improvised performance than during conventional performances. It is possible that some of the audience effects observed were not due to the differences in sound parameters as such, but the visual aspects of the performance. To explicitly assess the differential effects of sound and vision on audience response, formal measurement of performer movement would be needed, as well as comparing responses of audience members of

those who could hear but not see the performances, with those who could both see and hear.

Fourthly, the pilot study examined performance data from only two composers, the baroque composer Telemann, and the post-romantic/impressionist composer Ravel. Analysis of the performance related parameters revealed that although the performers performed both works with style and period-specific approach to tone and articulations (in both performance modes), they used similar performing strategies when applying improvisational approach to performing both Telemann and Ravel's works. During the improvised performances more attention was given to longer-term gestures, phrasing was more coherent structurally while at the same time inserting spontaneous but shared extemporized passages. This might be seen as an unexpected result, since improvisation, because of its unplanned nature, is often presumed to be unstructured and less coherent than non-improvised performance. The generality of these characteristics would be better established by investigating their occurrence in other classical styles, such as the early romantic period typified by a composer such as Franz Schubert.

Fifthly, while gathering verbal feedback from the audience, and brain measurements from both performers and audiences, the Dolan et al. (2013) study did not formally capture the insights and impressions of the performers themselves. For a fuller understanding of the parameters of an improvisatory state of mind, objective measures (of performance parameters and brain activities) should be compared with the subjective experience of the players.

Research Questions and Paper Structure

The primary questions which motivates the current study are

1. Does the improvisational act induce a different state of mind in performers, and is it transferable to listeners? Following from that,
2. if such state exists, how can we best characterize it?

At a more detailed level, further elaborating question 2 in respect of the key concepts of flow and the EBH:

3. Do performers' subjective accounts of their improvisatory experiences contain elements indicative of a flow experience?
4. Are there quantitative signatures of a shift from a secondary toward a primary state of cognition when comparing the brain activity during the prepared and improvised performances? In particular, can one find significant differences in terms of the LZ complexity of the EEG signals of musicians and audience?

Finally, as control questions aimed at resolving the limitations of earlier work discussed in section Scope of the Present Study above:

5. Do the body movements of musicians as visually experienced by audience members affect the magnitude of their response to the improvised performances? This will help clarify which medium is the basis of the communication of the contrasting states.
6. Does the level of musical training or knowledge of audience members affect their response to the improvised

performances? Can the effects only be manifested between trained people?

7. Do the objective performance characteristics that distinguish improvised performances of Telemann and Ravel extend to the music of a different period exemplified by Schubert? In particular, is there evidence of a greater degree of coherence and longer-term phrasing in the improvised version of Schubert?

Posing these questions has led us to the use of a combination of different methodologies in an interdisciplinary approach to capturing and analyzing multiple aspects of concert performances of items from the classical chamber ensemble repertoire. The design of the study allows us to measure the following features of conventional and improvised performances (given here in the order in which they are treated in the results section):

- (A) sonic and performance related parameters characteristics of the performances (notes played, and the timing, dynamics, and timbral qualities of those notes).
- (B) Post-performance assessments by the musicians themselves;
- (C) Continuous body motion tracking;
- (D) Post-performance audience ratings;
- (E) Real-time continuous monitoring of brain activity (EEG) of performers and audience members;

This order or presentation represents a progression from examining aspects of the performances and the performer experience, to examining the audience experience, and finally the co-ordination between performers and audience.

METHODS

Participants

Musical performers consisted of a professional trio—Kate Smith (voice), Rosie Bowker (flute) and Thibault Charrin (piano)—expert in classical improvisation, recruited and mentored by the 1st author. In particular their improvisatory practice was deeply informed by the performance practice developed over a lengthy period in the context of an advanced pedagogical center headed by the 1st author. The performers, although now independent professional practitioners had experienced extensive tuition and professional development in that context.

The invited audience comprised 22 adults, mainly postgraduate students and staff from the two UK academic institutions involved in the study. They contained individuals with a wide range of experience with, and training in, classical music. This was ensured by asking potential audience members to complete a pre-screening questionnaire, with questions about musical experience (for details, see results of questionnaire data).

Informed consent was obtained through a letter of invitation to all participants outlining what would take place in the experiment and asking them to confirm their acceptance of the invitation. Once accepted, a small subset of the audience were invited in writing to participate in the EEG study. Of the initial four audience members invited, one declined, and was

replaced by a fifth who accepted. Performers gave explicit written permission for their identity to be revealed.

General Procedure

The experiment took the form of a live chamber music concert on 21 March 2017. It took place in the Data Observatory at the Data Science Institute, Imperial College London (institution of the 4th author) with the aim of using its motion capture facilities, in the presence of an invited audience, all of whom had agreed in advance to be participants in the experiment. A Yamaha C-7 grand piano was hired to ensure the closest approximation to a fully professional concert. The seating was arranged such that half the audience could only hear but not see the performers. The size of the audience was the maximum feasible given the available space in the laboratory, in addition to the performers, the research team, and the scientific and musical equipment in place.

During the experiment each piece was performed twice: once in what the performers themselves chose to describe through their shared professional understanding as a “strict” mode (corresponding to a prepared interpretation), and once in what they described as a “let-go” mode (corresponding to the improvisatory approach in which they had been mentored). In the strict / prepared mode the players focused mainly on controlling technical precision, timing co-ordination, accuracy of the score’s details, avoiding risks, while at the same time creating the most convincing and expressive performance possible. In contrast, during the let-go / improvised performance the players were asked to play freely, as they would do for friends, expressing themselves spontaneously and not putting an imperative focus on “no wrong notes.” Note that the let-go performance still requires thorough knowledge of the written work, its harmonic and stylistic language and at the same time the ability to deviate from the written text in an unplanned coordination with the other ensemble partners. Moreover, the musicians were not operating according to any explicitly articulated set of rules for guiding these improvised deviations from the score. The order of the prepared and improvised performances was randomly varied from item to item, and this order was known only to the performers, who decided the order on the spur of the moment (i.e., audience members were unaware of which version was played each time).

The audience was briefed by one of the researchers that they were about to hear a sequence of pairs of trio performances that would involve some elements of improvisation. All members of the audience were asked to provide verbal responses via a questionnaire which was distributed prior to the start of the performance. After each performance members of the audience were given a short time to rate it for the degree to which they detected or experienced five qualities: improvisatory in character, innovative in approach, emotionally engaging, musically convincing, and risk-taking. These questions were identical to the ones used in the Dolan et al. (2013) study. Responses were made using a six-point Likert scale, ranging from “not at all/none” to “totally/completely.”

The continuous movements (3-dimensional positions of up to 20 joints) of the three performers were captured by means

of an existing motion tracking system formed by five Microsoft Kinects devices distributed circularly around the performers (further technical specifications are given in the results section Continuous Body Motion Tracking below).

EEG brain activity of four audience members as well as the three performers were captured with seven high-performance EEG recorders using 19 electrodes for each person (further technical specifications are given in the results section dedicated to EEG data analysis below). Two of these audience members could both see and hear the performances, the other two could hear but not see them. Within each pair, one participant had a high degree of training in classical music, the other a low degree.

High quality audio and video recordings of the performances were made by means of two HD videocameras located in different positions.

DATA ANALYSIS AND RESULTS

In this paper we confine our analysis to data from two performances of the opening Andantino section of Franz Schubert's "*Der Hirt Auf Dem Felsen*" (*The Shepherd on the Rock*) Op 129. Within the pieces measured during our experiment, drawn from the existing repertoire of the performers, this was the piece which the performers judged best realized their differential intentions for the two performance modes, and provided sufficient data for an intensive analysis.

The analysis proceeds from sonic and musical features of the performance, as experienced and characterized by the musicians involved, through the visual features of those performances captured by movement, leading to the explicit audience response to these performances, and concluding with the neurophysiological data examining relationships across and between performers and audience at a level beneath the conscious and explicit.

Sonic and Performance Related Parameters Characteristics of the Performances

This section presents an analysis of the performance-related parameters of the prepared and the improvised versions. This analysis was undertaken by the 1st author with the aid of repeated critical listening (jointly with the performers) and *Sonic Visualizer software*² which provided a visual trace of key physical characteristics of the performances.

Below we first summarize some overall characteristics of the performances, and then present a more detailed analysis of three particular—yet characteristic—moments where the musicians spontaneously took enhanced risks in the improvised version—by deviating from the score's instructions in terms of timing, dynamics, and timbre, actual extemporized notes, or a combination of all three. The audio/video clips of each moment in the two performances are added as Supplementary

Files (**Videos 1–6**) respectively where the first file of each pair is extracted from the prepared performance and second is the improvised).

In what follows in this analysis, objective measures (duration, intensity, frequency) are interpreted in the light of inter-subjective judgment of the first author and performing musicians. Thus, all evaluative remarks (terms such as "better") reflect the joint musical judgement of the individuals concerned.

General Observations

When comparing the prepared and improvised performances we found significant differences in six features, the first four of which pertain to physically measurable sonic and temporal characteristics of the performances, and the last two of which pertain to structural features of the performance.

Timbre

In the improvised version there is a wider range of timbre changes both individually and in the group orchestration (see Example 1 below).

Speed (tempo/duration)

The improvised performance is objectively slower (average crotchet/quarter note = 88 bpm) than the prepared one (average crotchet/quarter note = 92 bpm). However, the critical listening confirmed that despite the slower tempo—in absolute terms—of the improvised version, it gave the subjective impression of being faster and more "forward going."

Dynamics

In the improvised version the dynamic diversity is larger compared with the prepared version. For example, the intensity in the prepared version of the start of the performance (bars 7–9) varies between -17.75 and -14.40 dB, whereas in the improvised version it varies between -31.14 and -14.77 dB (a range 14 dB bigger) (see Examples 1 and 2 below).

Pulse, meter and metrical division

The improvised performance contains more longer-term phrasing gestures. These are better coordinated between performers and more in line with Schubert's written instructions (dynamics, timing and expression) compared with the prepared mode. For instance, there is one phrasing slur mark in Schubert's score, running from the last beat of bar 19 to the first beat of bar 21. Expert critical listening, supported by the Sonic Visualizer data, confirms that the improvised performance follows this instruction more closely than the prepared version. In the improvised version there are smoother timing and dynamic transitions from bar 19 to 20, and 20 to 21, whereas in the prepared version there are discontinuities which emphasize individual crochets and the start and end of each bar unit, thus breaking the indicated phrasing.

²The Sonic Visualizer was developed in Queen Mary University of London as a part of the CHARM project. We thank the double-bass player and researcher, Mark Gilenson (Schola Cantorum, Basel), for his assistance with the Sonic Visualizer analysis.

Moreover, whole-bar beats and hyper-measures³ of two bars are clearly heard (and seen) in the improvised version, while hardly existing in the prepared version. This might account for the impression of a more forward going musical movement, felt during the critical listening sessions, while the prepared version is at times more fragmented. An example of this is discussed in more detail below (Example 2).

Risk taking

In the prepared version, the musicians perform the written instructions literally, and thus make it more predictable (easier to anticipate what happens next). In the improvised version they spontaneously deviate from the text by means of timing, extended dynamics and timbre as well as extemporized notes, making the performance less “safe” to manage. And yet, repeated critical listening concluded that it is in the latter version where the performers were better coordinated in key moments (end of phrases, moments of harmonic resolution and significant harmonic changes).

“Mind-reading” during shared extemporized gestures

By “mind reading” we mean moments where one musician deviates from the score by extemporizing notes and another extemporizes in response instantly, creating together an unplanned, yet coherent joint musical gesture that reaches a final goal point together. Such moments only occur during the improvised performance (compare **Videos 5** and **6**), suggesting heightened listening. This is also confirmed on in the musicians’ reports.

Detailed Analysis of Specific Representative Examples

Below we present an analysis of three indicative examples, illustrating in more detail the artistic differences between the prepared and improvised performances. There are more similar examples throughout the performance which space precludes mentioning, but which will form the basis of a more detailed musicologically oriented publication (in preparation).

Example 1

This tiny 3-note long flute solo playing (see the corresponding score in **Figure 1**, bars 7–8) is a microcosm of the improvisational approach which permeates the entire performance and illustrates to a greater or lesser extent all 6 features outlined in the “general observations section.” We will take them one by one and show how they are manifested in this segment.

Timbre. We chose this point for illustrating the timbral element because it is the only moment where timbre is clearly analysable by the sonic visualizer software, as only one instrument is playing. The two spectrographs in **Figure 2** visually illustrate timbral characteristics to be heard in the audio clips of the two performances. In the improvised performance there is a gradual

evolution of the timbre during the first note played (reflected in the harmonics appearing gradually), while in the prepared version the first three harmonics appear more strongly together from the outset. In the improvised version the fundamental frequency as well as the lower harmonics are stronger (manifested by the thicker and more emphasized colors of these first four spectrograph lines in the improvised version, as seen in both spectrographs of **Figure 2**). The higher harmonics are relatively less present in the improvised version, comparing with the prepared version. (Peak of the harmonics in the prepared version is 5,380 Hz, in the improvised one (4,780 Hz). This contributes to the improvised version having a softer timbre (less emphasized higher harmonics) on the flute’s *e flat* and *f*.

The tone quality in the prepared version is as excellent as it is in the improvised version (with hardly any use of vibrato) stable and in full control, suggesting a choice rather than a “better” performance.

Tempo/duration. There is no clear tempo at the beginning of the improvised version, which creates an “out of time” effect in the solo flute’s entry. It is achieved by the significantly longer duration of the opening *d* (comparing with the prepared version) 4.15 s vs. 2.8 s respectively, fluctuations in the speed of vibrato, and a dynamic wave mentioned below.

Dynamics. Unlike the prepared version, in the improvised version there is an extreme dynamic range, with an unexpected additional dynamic “wave” of down and up again—this time with a narrow vibrato toward the end of the long *d*, continuing into the *e-flat* followed by the *f* without separate articulations. In the prepared version the flutist applies a milder, consistent crescendo, (without the dynamic “wave” at the end of the long *d* note).

Pulse and Meter. Together with the fact that here is no clear beat in the opening of the improvised version, the above points mean that the gesture *e flat* \Rightarrow *f* is performed more as a prolongation of the *d* than a separate rhythmical event leading to a different bar. By doing so, bars 7 and 8 become one hyper-measure of two bars, with a first part (bar 7), being free, out of tempo and “out of time,” fulfilling Schubert’s fermata instruction to the fullest. In the prepared version, there is a clear distinction made between bar 7 and bar 8, through the more metronomic use of accents. The distinction is further confirmed by the pianist entering in bar 8 with even quaver beats (unlike the improvised version where the pianist’s meter is clearly one beat per whole bar, with one gesture every two bars). The result is a more subdivided rhythmic approach in the prepared version.

Risk taking. There are two risks the flutist takes within the first few seconds of the improvised performance. The first is her choice to open with a gradual evolving of the opening note’s tone color mentioned above. This is a harder choice than the conventional way of approaching a tone’s outset, with a higher level of risk-taking (the risk of losing the tone all together). The other risk relates to the previous point mentioned above about creating one hyper-measure of two bars (rather than relating to individual crochet beats). By so doing the flutist is taking the risk of not meeting the pianist in time for the next bar, as she “gives

³The term Hyper-measure, attributed to Cone (1968), refers to groups of bars, where bars act as beats, leading to a larger-scale basic rhythmic gesture. A generation before Edward Cone, the highly influential pianist and teacher Arthur Schnabel, used this concept and terminology in his teaching. (Cone studied with Arthur Schnabel’s son, Ulrike Schnabel).

October 1828.

1 Andantino.

Clarinette in B.

Singstimme.

Pianoforte.

pp > pp p
lange Haltung

8

FIGURE 1 | Score of Schubert's Shepherd on the Rocks, Op 129, Bars 7–8. The very opening of the performance. [Associated video/audio clips are 1 [prepared bar 7 - bar 8 beat 1], 2 [improvised bar 7 - bar 8 beat1], 3 [prepared bars 8–12] 4 [improvised bars 8–12].

up” the markers of the crotchet beats to which the pianist can relate when preparing for joining the flutist in bar 8.

Mind reading. Despite this flutist's risky choice, they end up finding each other absolutely on time, which may suggest heightened listening and “musical mind reading” as defined above. These timing and loudness variations do not appear in the score, and are the flutist's personal spontaneous interpretation. The free rhythmical approach that the pianist takes from the start of his entry in bar 8 (as can be heard in video/audio clip 4, in contrast with 3) may be his spontaneous response to the rhythmical freedom applied by the flutist the bar just before.

This example clearly illustrates the varieties of means of implementing an improvisational approach that do not require the extemporization of new notes, but variations in the performance parameters of composer-notated elements. This is the only example where we were able to analyse timbre in a formal way, however there are multiple examples of some of the other features. The next two examples are chosen to illustrate respectively meter and dynamics (Figures 4, 5), and extemporized notes, risk-taking and mind reading (Figure 3).

Example 2

In this example we concentrate on tempi/durations, dynamics, pulse and meter and the inter-relations between them. We chose to look into performance related parameters in bars 8–9 (see Figure 1), as we concentrate on a specific, and early, example of a difference between prepared and improvised approach which recurs throughout the performances.

Tempo/duration. In the prepared version there is a greater evenness of quaver and crotchet beats comparing with the improvised rendition. The range of tempo-changes (gap between slowest and fastest) in the prepared version is slightly narrower. Also, these changes are more frequent (up-down-up-down), compared with the improvised version where there are less tempi fluctuations (just one down-up wave).

Dynamics, pulse, meter, and phrasing—ingredients of musical flow. In the prepared version, the frequent peaks in the loudness profile are indicative of micro-accents on each quaver beat, while in the improvised version we notice a larger and smoother wave shape, signifying the avoidance of these frequent, regular accents.

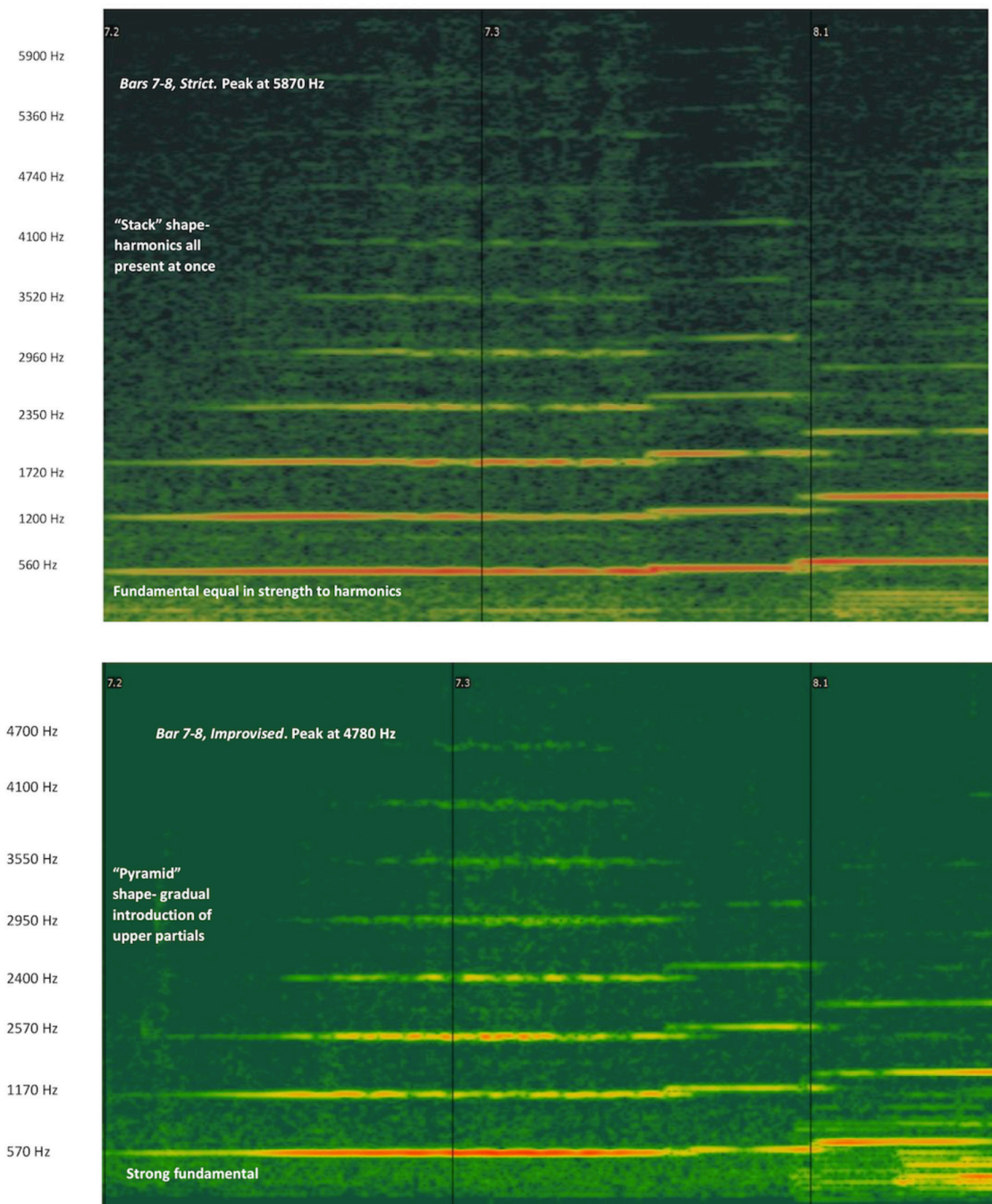


FIGURE 2 | Flute harmonics up to 5,400 Hz (fundamental frequency at the bottom plus the first 10 overtones above). Prepared performance in the top figure, improvised performance in the bottom one. 7.2 and 7.3 indicate the 2nd and 3rd beats of bar 7 respectively.

This can be heard in the audio/video clips number 3 and 4 respectively and observed by the number of peaks in the curve of intensity (20 peaks in the intensity curve of the prepared version, vs. 10 in the improvised version for this segment). The overall shape of the loudness curve in the improvised performance indicates waves of dynamics in accordance with the two bars hyper-measures, resulting in a less fragmented

and more flowing musical movement. The occurrence of whole-bar gestures and hyper-measures of 2-bars through large parts of the improvised performance was also identified by the musicians during the critical listening sessions, in contrast with the notion of 3 beats per bar that the musicians identified in the prepared performances (see section Post-performance Assessments below).

FIGURE 3 | Score of Schubert's Shepherd on the Rocks, Op 129, bars 164-177. The notation added in red by the first author after the concert indicates the musicians' extemporizations in the improvised version. Related video/audio clips are 5 [prepared] and 6 [improvised].

It is also noticeable that there is a relationship between the tempi and the loudness curves: they increase and decrease together across whole-bars units of time. Such a feature is seen as adding to the overall higher level of coherence and forward movement experienced in the improvised version, comparing with the prepared one. This is even though, according to their own reports, the musicians were much less aware of metronomic and metric (tactus) control during the improvised version, in contrast with the prepared version. Yet the actual result suggests the opposite. This dissociation between performance decisions and conscious awareness of these decisions is one of the characteristics of a state of flow (as pointed out by Després et al., 2017).

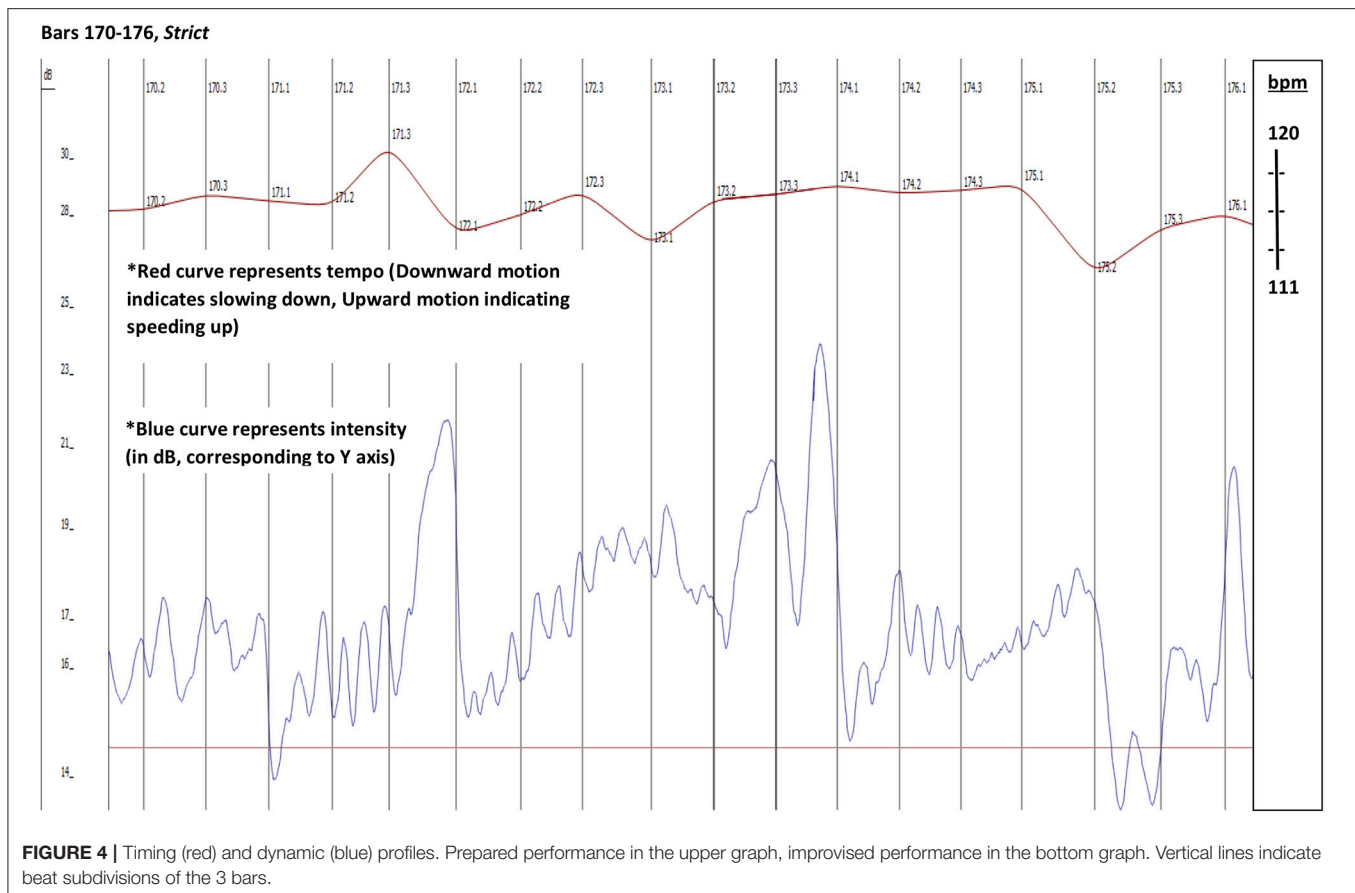
Example 3

This example, whose score is presented in **Figure 3**, is an illustration of the way in which sonic & temporal characteristics of the performances, contribute and support structural features of the performance. It also illustrates the use of extemporized notes which were not present in the original score.

The singer connects bar 171 to bar 172 with an improvised upbeat "d" to the following *e flat*. A bar later (173-174) the flautist extemporizes an upbeat passage to her *e flat* with all three

Schubert's notes of this motive: $c=>d=>e \text{ flat}$. Unlike the singer, the flautist starts her extemporized gesture *before* her entry is due in the score, and thus takes a significantly greater risk of losing her partners. Listening to bar 173 reveals the mechanism that made this possible—the pianist provides the flutist with the additional time needed to fit her extemporized responding gesture off the beat, by spontaneously slowing the tempo down, as well as playing a significant diminuendo, and thus making the *rallentando* more coherent (see the different tempo curve in **Figures 4, 5** from bar 173.3 to 174.1). Following the singer's extemporized upbeat gesture at the end of bar 171, the trio maintains a noticeably slower tempo throughout bar 172.

Bars 172–175 (including) have in the base (the pianist's left hand) one minim long *d* (musically described as a "pedal") in each of these four bars. The resulting sound effect (as exemplified in the audio clips) is of bars 172 and 173 of the improvised version being one rhythmic gesture (hyper-measure), with bars 174 and 175 being another, creating two longer gestures of two bars, where every bar is one beat: the first emphasized and the second released (compare the different intensity curve in **Figures 4 and 5**, between bars 171 and 175). This larger scale gesture is another factor that enabled the singer and the flutist to have the extra time they needed to accomplish this extemporized



dialogue over the pianist's pedal. Indeed, this is the only moment in this section where the composer stops the movement of the baseline and the harmonic progression. One may speculate that the extemporized enhancement performed by the singer and the flutist, with the crucial support of the pianist, amplifies and makes more explicit the composer's intention at this point.

Even if the members of the trio would have decided to try to, there wasn't enough time to plan the details of such a complex chain of events involving all three performers abandoning the conventional route of following the score's instructions. Listening to the recording after the performance, they were surprised to discover this moment in the improvised performance version, which suggests it was done without full awareness of the details.

No deviation from the score occurs in the prepared version, where in the same passage there is a mild increase of tempo during the first two beats of bar 172 (contrary to the decrease of tempo in the improvised performance).

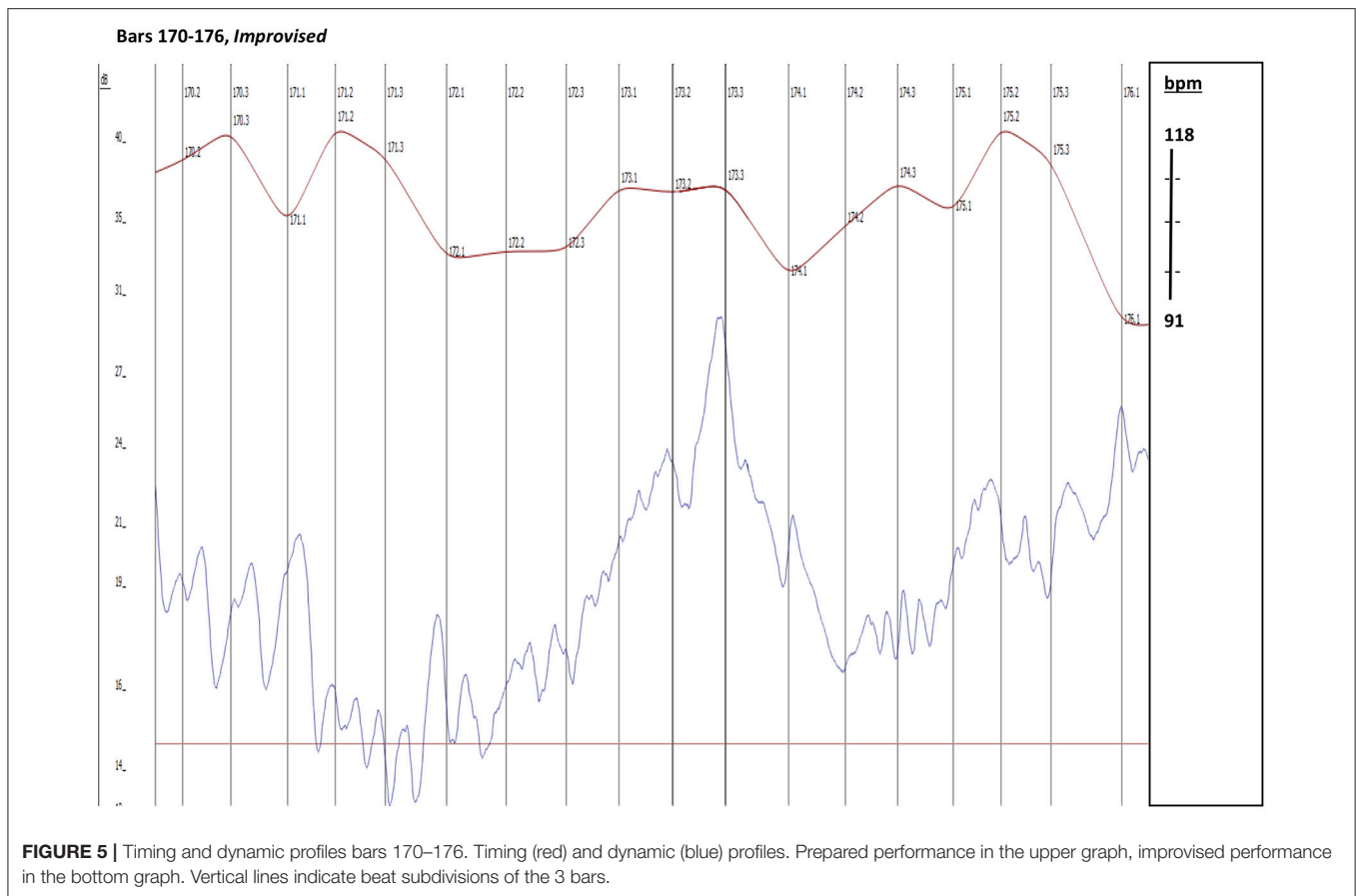
The six described types of difference between the improvised and prepared versions of this Schubert movement, are closely similar to those found in Dolan et al. (2013) through analysis of performances of works by Telemann and Ravel, even though the compositional periods and languages (and the actual musicians undertaking the performance) were different. This lends support to the notion that the improvisatory state of mind enables a particular constellation of performance features which can be applied to music of varying styles; these features include the use of larger phrasing units, a greater range of dynamic and timbral changes, less emphatic metrical divisions, and extemporized gestures, spontaneously split and shared between partners, with the risk taking it represents (cf p. 32–33 of Dolan et al., 2013).

Post-performance Assessments

The performers were invited to reflect on their performances and the performance process a few days after the concert-experiment and again 20 weeks later. Rosie Bowker (flute) made a written account summarizing these responses, which can be read in full in **Appendix 1**.

The reflections followed critical listening sessions involving the three performers, facilitated by the first author. In the first reflections, the memory of the subjective experience was relatively fresh and present. The experience of watching and

⁴The score attached is the performers' working draft. What is marked here as bar 1 is in fact an upbeat, meaning that bar 1 is actually the bar marked as 2 in the musicians' working score. Since the musicians worked with this score, we will refer to their markings. The Sonic Visualizer graphs refer to the musicians' working score bar numbers. In this performance the part written for clarinet in B was played by a flute in C. This means that all flute pitches referred to in the text are one whole tone lower than what is notated, and the performance began at bar 7.



listening enabled the musicians to re-live the experience and retrieve some of the subjective experiences of the performance. Twenty weeks later, the memory of the experience was more remote. Therefore, the critical listening was more focused on the musicians' considered assessment of the features present in the audio-visual recordings, as well as reflections on the nature of the contrasting mind-sets in the two types of performance.

The question discussed by the performers during the first series of the critical listening sessions was: "How would you describe the differences you felt as performers, before and while performing, between the two mindsets?"

In response, the musicians reported that the prepared version had to do with "... greater feeling of mental and physical control... and being more precise about counting and note values... Overall the increased control resulted in a performance in which we played more consistently together within each bar because we were playing more in time, metronomically speaking." This corresponds with our findings about more emphasis in the prepared versions on shorter-term beats of quavers and crochets evenly emphasize (rather than whole bars or hyper-measures of two bars).

In the improvised version, where our analysis found larger beat and freer and longer-term phrasing, the musicians reported—"... the freedom of the 'let go' mindset allowed me to create a wider range of colors and dynamics..." This is confirmed by the analysis of the performances in section Sonic

and Performance Related Parameters Characteristics of the Performances.

Twenty weeks later the author invited the musician to a second series of critical listening sessions, asking the following question: "Please, could you share your thoughts about the performances and how you feel about them when you listen to the performances now, 20 weeks later?"

In response, the musicians confirmed their perception of "... a greater range and variety of timbre, dynamics and colors".

Further comments about the two modes of performance were in terms of performance attitude, artistic outcome, and well-being. One important feature was the sense of connection between the players. In the prepared/strict version performers got the experience of: "... listening to individual performers one at a time and reported having very little sense of connection between the performers". In the improvised version—"When listening back to the 'let go' performance all of us responded to the video by saying that the performers were more integrated — there was a greater sense of connection and the ensemble work was more convincing."

The musicians noted the sense of trust that was manifest in the improvised performances, e.g., "Trust in my own musical instincts and the capability to complete the task..." They asserted that "Trust between performers is imperative for being able to apply an improvisational state of mind..."

A final feature related to experienced well-being/anxiety, e.g., *“If the trust isn’t there between performers it becomes increasingly difficult to stay in the ‘let go’ mindset and much easier to revert to the ‘strict’, controlled and anxious mindset.”* Another related comment was *“the ‘strict’ mindset also resulted in Thibault and I reporting more self-conscious performances, increased levels of performance anxiety and more internal critical chatter”*. These statements are very consistent with the reported experiences in states of flow, and suggest that these states are conducive (possibly even necessary) to the kinds of performance characteristics observed.

Continuous Body Motion Tracking

Methods

We utilized Microsoft Kinect v2, a commercial motion tracking device, providing computer vision based motion sensing via mature APIs (Zhang, 2012). This version can provide data of up to 25 joints per body, with an improved tracking accuracy due to an enhanced depth sensor. By means of a scalable data fusion system, we could concurrently gather information from 5 Kinects sensors, improving the data resolution and overcoming some limitations such as occlusion when several bodies are together in the space.

We judge that wearables and wearing markers are generally more accurate than purely computer vision systems. However, the former are generally more cumbersome and might affect the performance. Although most research in this area has been focused on the fine-grained movements of fingers, wrists, or lips (Grosshauser et al., 2015; MacRitchie and McPherson, 2015), our research goals nevertheless focused on the broader head and body movements and in their comparison with performing styles. Putting that together with the need of a non-intrusive setup, makes the Kinect setup the most appropriate and cost-effective solution.

Data

The data collected through this system regarding motion consisted of a multivariate time series for each one of the detected bodies. Each multivariate time series is composed of 25 variables corresponding to the 3D positions of 25 joints that Kinects v2 can detect.

The recorded data was, unfortunately, heavily affected by noise due to imprecisions of the Kinect tracking mechanism. Also, due to the Kinect aligning system, the data points were sampled at irregular times, having no fixed sampling frequency. In order to reduce the impact of these impairments, the data was pre-processed as follows:

1. All data-points occurring within time bins of 250 ms were grouped together, and their median was taken as a representative. This generate time series with a regular sampling frequency of 4 Hz.
2. Positions were transformed into velocities, and all velocities above a given threshold were rejected as artifacts.
3. The joint velocities were grouped together (adding the magnitude of the 3-dimensional vectors) in 3 groups: head, upper body and lower body.

For the results described in the rest of this section, we have solely used the movements of the singer and the flutist. We chose them because they were the only two individuals with freedom to move their feet and move around, in contrast with the pianist and audience members who remained seated.

Statistics

The mean power spectrum and coherence between signals was computed using the well-known Welch method, using Fourier windows of 16 samples. The spectrum of velocities was divided in slow movements (below 0.75 Hz), medium movements (0.75–1.25 Hz) and fast movements (above 1.25 Hz). Also, linear regressions of the movements of one musician given the other's movements were computed over sections of 100 samples. Statistical significance is calculated using unpaired *t*-tests, and effect sizes are measured with Cohen's *d*.

Results

We investigated the variations in the amount of movement in each musician between the prepared and improvised renditions, as given by the mean value of the total velocity of each of the three body segments. We found a consistent increase in movement in the prepared version, being significant for the fast movements of the head and lower body of the singer and all comparable movements of the flutist (see **Figure 6**).

When studying the covariance and Pearson correlation coefficient between velocities of the body segments of flutist and singer we found no significant differences. However, when decomposing the covariance in its spectral components, we found that the correlation between the fast component of motion is markedly different during the improvised and prepared performance modes (see **Figure 7**). In particular, fast movements tend to be less correlated in the prepared (strict) than in the improvised (let-go) versions. Note that the coherence is a normalized quantity, and hence is not affected by changes in the total amounts of movement, making this finding independent of the previous one. Moreover, these two findings together imply that when shifting to the improvised performance the musicians' movements are reduced, and an important part of this reduction takes place over fast uncorrelated movements.

Finally, by comparing the residuals obtained after running a linear regression over the movements given the movements of the other musician, we found that on average all the residuals are larger in the prepared version, this difference is significant for the head and lower body movements of the flutist (see **Figure 8**). The consistency of this result supports our previous explanation, providing additional evidence toward the idea that an important cause of the additional movement found in the prepared version is due to movement that is not coordinated between musicians.

Post-performance Audience Ratings

Levels of musical engagement/training were assessed through seven scaled items adapted from the Goldsmith's Musical Sophistication Index (Müllensiefen et al., 2014). These assessed, number of musical instruments played (including voice), amount of practice on these instruments, amount of formal training in music performance and music theory, and amount of listening

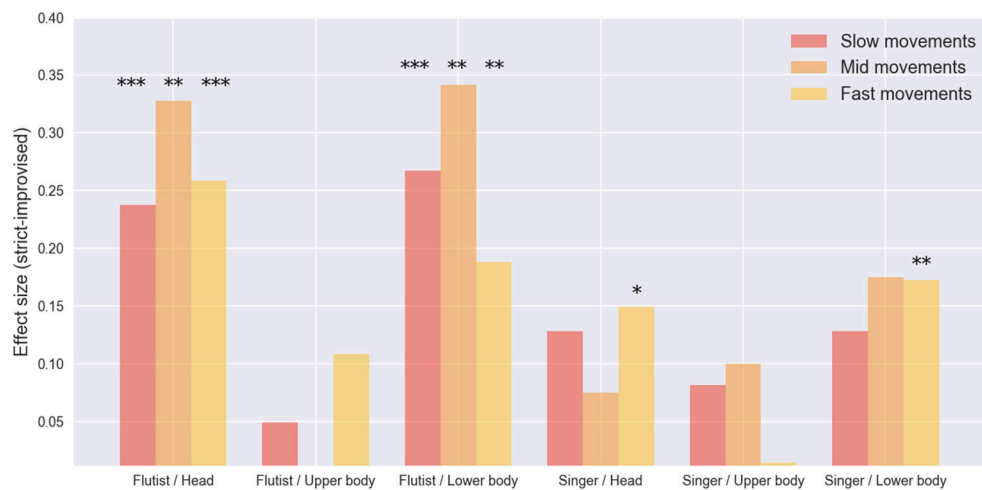


FIGURE 6 | Effect size for movement differences between prepared (strict) and improvised (let-go) performances for the flautist and singer. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

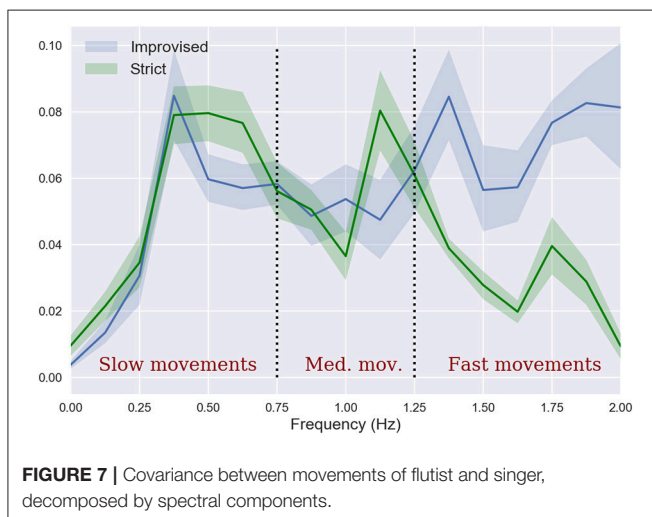


FIGURE 7 | Covariance between movements of flutist and singer, decomposed by spectral components.

to music (both recorded and live). A composite measure of engagement was obtained by adding these 7 scores together. These scores ranged from 6 to 33 with a mean of 22. Participants scoring 22 or less ($n = 10$) were assigned to the “lower engagement” group, those scoring 23 or more to the “higher engagement” group ($n = 12$).

Two-way ANOVAs were undertaken for each of the five post-performance ratings, with performance type (prepared or improvised) as a within-subjects factor, and level of musical engagement as a between-subjects factor. There was a significant main effect of performance for “emotionally compelling” [with the mean rating for the improvised performance being 3.8, as compared to the prepared performance at 2.6 ($F_{(1,20)} = 13.6$, $p < 0.001$, Eta squared = 0.259)]. There was also a significant main effect of performance for “musically convincing” [with the mean rating for the improvised performance being 4.1, as compared to the prepared

performance at 3.2 ($F_{(1,20)} = 7.4$, $p = 0.01$, Eta squared = 0.320)]. There were no significant main effects or interactions involving the engagement variable, thus indicating that musical experience/training was not a significant influence on audience judgment.

Familiarity with the music of Franz Schubert was assessed by a single 4-point scale question, ranging from “not at all familiar/don’t know” to “I know his music very well (i.e., possess recordings/have studied it).” 13 participants were assigned to the high-familiarity group (scoring 3 or 4), and 9 participants to the low-familiarity group (scoring 1 or 2).

Two way ANOVAs were undertaken for each of the five post-performance ratings, with performance type (prepared or improvised) as a within-subjects factor, and familiarity with Schubert as a between-subjects factor.

Table 1 shows the mean ratings in each condition. In addition to significant main effects of performance on “emotionally compelling” and “musically convincing” there was also a significant main effect of familiarity with Schubert on the “musically convincing” rating. Audience members who were familiar with Schubert rated the performances as less musically convincing (mean = 3.3) than those unfamiliar with Schubert [mean = 4.1, $F_{(1,20)} = 6.8$, $p < 0.02$, Eta squared = 0.088]. There was also a significant interaction. For the dimensions of “emotionally compelling” [$F_{(1,20)} = 10.0$, $p < 0.005$, Eta squared = 0.054] audience members familiar with Schubert showed a significantly greater difference in mean rating between the two versions (prepared = 2.6, improvised = 4.0) than those unfamiliar with Schubert (prepared = 4.0, improvised = 4.2).

Finally, two way ANOVAs were undertaken for each of the five post-performance ratings, with performance type (prepared or improvised) as a within-subjects factor, and with the presence or absence of sight of the performers as a between-subjects factor. In no case was there a significant effect of sight, either as a main effect or in interaction.

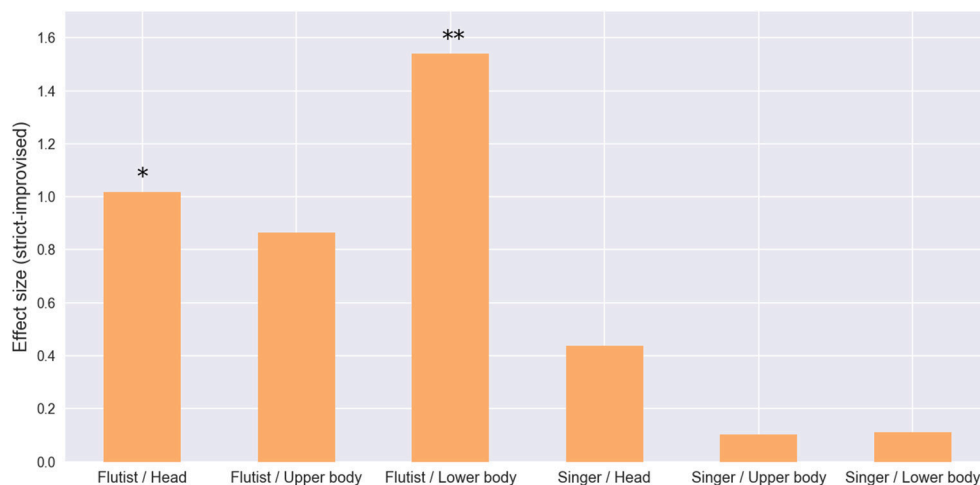


FIGURE 8 | Residuals from linear regressions of flutist and singer given the movements of the other musician. * $p < 0.05$, ** $p < 0.01$.

TABLE 1 | Mean audience ratings (max = 5) on five assessment scales, according to performance mode and familiarity with the music of Franz Schubert.

Measure	Low familiarity with Schubert		High familiarity with Schubert	
	Prepared	Improvised	Prepared	Improvised
Improvisatory	2.0	2.5	1.4	2.1
Innovative	2.2	2.9	1.5	2.1
Emotional	3.4	3.7	1.9	4.0
Musical	4.0	4.2	2.6	4.0
Risk-taking	1.8	2.6	1.6	2.3

In sum, two of the post-performance rating scales (“emotionally compelling” and “Musically convincing”) were sensitive to the differences between the prepared and improvised version, with the improvised version rated higher than the prepared version. This effect did not depend on whether the audience members could see the performers, nor was it affected by the level of musical training of audience members. Familiarity with the music of the composer did, however, impact on the results. Those familiar with Schubert judged the improvised version more emotionally compelling when compared to the prepared version, than did those unfamiliar with Schubert.

Real-Time Continuous Monitoring of Brain Activity (EEG) of Performers and Audience Members

Methods

Data acquisition

Raw EEG signals of the three performers and four audience members were measured using CE-certified devices (NCLogics AG, Munich, Germany). For each participant, 19 Ag/AgCl electrodes were placed on the following locations (all according to the 10–20 electrode position system; Klem et al., 1999): Fp1,

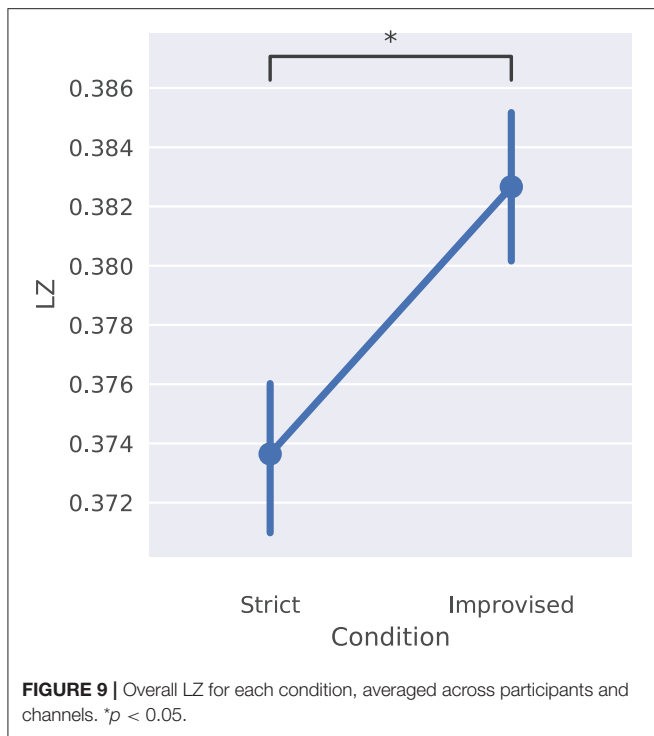
Fp2, F3, F4, F7, F8, C3, C4, T7, T8, P3, P4, P7, P8, O1, O2, Fz, Cz, Pz. The reference electrode was placed behind Cz and the ground electrode on the forehead. All locations were cleaned with abrasive gel and conductive gel was used to ensure low skin impedance. EEG data were collected at 250 Hz, and bandpass filtered between 2 and 40 Hz. All devices were synchronized via a local Wifi network. Start and ending of each measurement were remotely controlled and synchronized. Times series EEG data were stored and exported for further analysis. Bad channels and bad epochs were visually identified and removed from the analysis.

Signal complexity

The method for calculating the LZ consists of two steps. First, the amplitude of a given signal X of length T is digitalized, calculating its median value and turning each data point that is above it to “1”s and each point below it to “0”s. Then, the resulting binary sequence is scanned sequentially, looking for distinctive structures that are used to form a “dictionary of patterns.” Finally, the signal complexity is determined by the number of patterns that compose the dictionary, denoted by $c(X)$. Note that regular signals can be characterized by a small number of patterns and hence have low LZ complexity, while irregular signals with no characteristic patterns requires long dictionaries and hence have large LZ complexity. Moreover, the quantity

$$\frac{c(X)\log(T)}{T}$$

is an efficient estimator of the entropy rate of X (Ziv, 1978), which has various interpretations within information theory (Cover and Thomas, 2012) and thermodynamics (Mézard and Montanari, 2009). This makes this normalized LZ a principled, data-efficient and timescale-independent estimator of the diversity of the underlying neural process. In the rest of the manuscript we refer to the quantity in the formula above generically as LZ.



Statistics

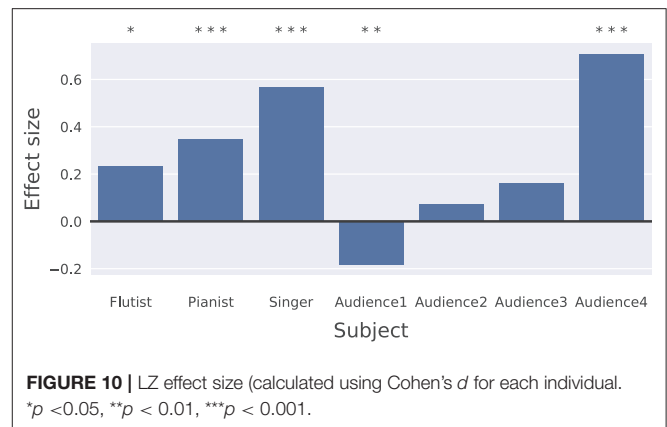
The neural signal was split in segments of 2 s, which provides enough data points to have an accurate estimation of LZ while being short enough to keep safe the stationarity of the data. The values of each segment were then binarized using the corresponding median value as a threshold. The LZ was finally calculated for each temporal segment of each electrode, and then averaged across time and electrodes to obtain one LZ value per subject per condition. Due to our small sample sizes, statistical significance is determined with t -tests (paired when possible, and unpaired elsewhere) and effect sizes are measured with Cohen's d .

Results

Increased complexity in the improvised version

Based on the properties of LZ outlined above, we investigated the complexity of the measured EEG signals of the three performers and four audience members in both conditions, under a working hypothesis that LZ is higher during the improvised than during the prepared condition. Our main result is that LZ increases in the improvised condition with respect to the prepared condition by a difference of 0.009 (95% CI: 0.001–0.016, $n = 7$, $p = 0.031$), shown in **Figure 9**. Significance was calculated using a two-sample (i.e., paired) t -test. **Figure 10** contains the effect sizes (Cohen's d) for each participant with subject-level significance calculated using a Mann-Whitney U -test.

The small p -value for the group-level test is caused by the fact that the observed LZ increase is very consistent across subjects, with 6 of the 7 participants showing changes in the same (positive) direction. While results among the audience are mixed, all three musicians show substantial increases in LZ during the improvised performance, and this effect is most significant in the singer and the pianist (see **Figure 10**).



Complexity increase comes from the right brain hemisphere

Following up on our main result, and in agreement with accepted neuroscientific theories, we find that the LZ increase is mainly localized in the right hemisphere (average difference in LZ increase between right and left hemisphere: 0.01, 95% CI: 0.004–0.016, $p = 0.003$). The right hemisphere is conventionally associated with cognitive processes like creativity and divergent thinking, which indicates that musicians were more engaged in a creative process during the improvised performance, and were less likely to enter the logic-driven and rule-following states usually associated with the left hemisphere. **Figure 11** shows the average difference in LZ increase and **Figure 12** its spatial distribution.

Changes in EEG power spectrum

We also calculated the average power located in each frequency band of the EEG signals of musicians and audience in the two conditions. We found that during the prepared performance there is more power located in low frequencies (delta, theta and alpha bands), while high frequencies (beta and gamma bands) are more active during the improvised mode. Interestingly, a similar phenomenon has been found when comparing EEG data from sleep conditions: high frequencies exhibit relatively more power during REM sleep and low frequencies are relatively more active during unconscious, dreamless sleep (Achermann et al., 2016). This suggests a relationship between this “crossed spectrum” (as shown in **Figure 13**) and various degrees of awareness, providing additional evidence to support the hypothesis that musicians and audience are more aware during the improvised performance than in the prepared version.

Discussion

This study confirmed distinct differences between prepared and improvised approaches to performances of the same piece of music. These differences were revealed through complementary analyses of (a) objective characteristics of the sound recordings, (b) musicians' self-report, (c) musicians' movements during the performances, (d) listener ratings, and (e) EEG measurements on both performers and listeners.

We take each of the detailed research questions in turn and briefly discuss what light our research has shed on each of them.

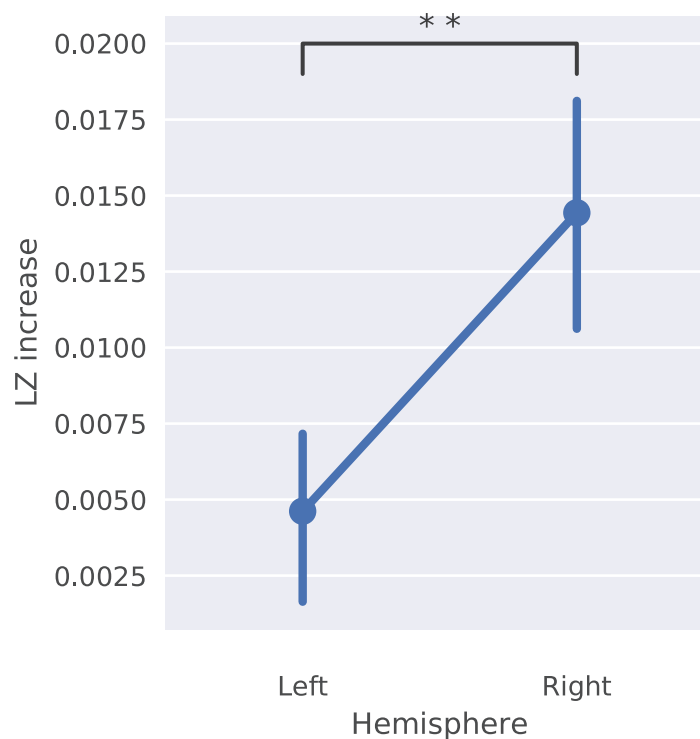


FIGURE 11 | Average LZ increase between conditions observed in all channels in the left and right brain hemisphere. $**p < 0.01$.

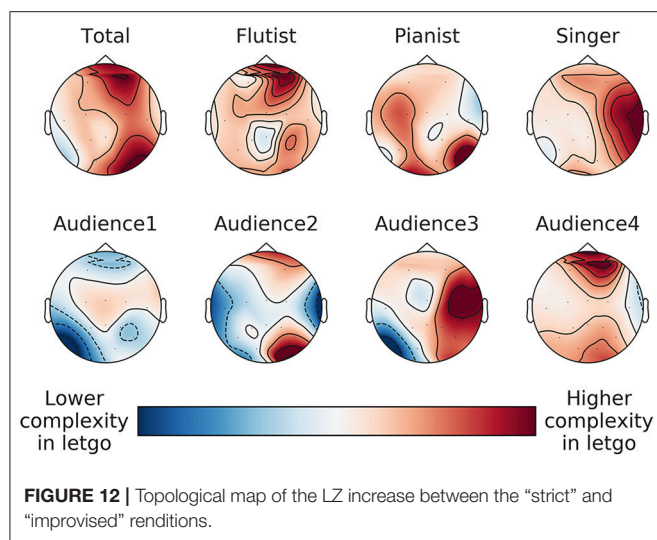


FIGURE 12 | Topological map of the LZ increase between the "strict" and "improvised" renditions.

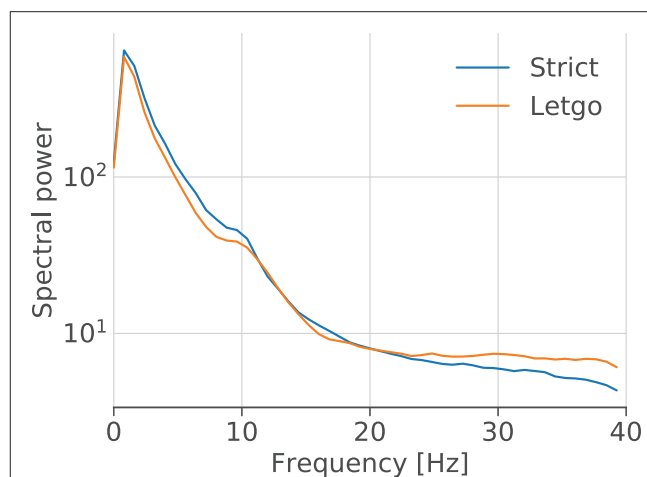


FIGURE 13 | Aggregated power spectrum. Power spectrum averaged across all subjects, calculated using Welch's method.

Do Performers' Subjective Accounts of Their Improvisatory Experiences Contain Elements Indicative of a Flow Experience?

The fact that the musicians reported surprise at discovering what they had done, suggest that to some degree, their actions were driven by intuition, and accessing knowledge in a non-conscious-analytical way, rather than conscious planned decision. Moreover, during the improvised rendition the performers took a significant number of risky choices and yet the

results sound more coherent, while the musicians experienced less anxiety and effort, and more pleasure.

Are There Quantitative Signatures of a Shift From a Secondary Toward a Primary State of Cognition When Comparing the Brain Activity During the Prepared and Improvised Performances? In Particular, can one Find Significant Differences in Terms of the LZ Complexity of the EEG Signals of Musicians and Audience?

While the literature about states of flow is mainly based in psychology, discussion of the EBH are mainly rooted in neuroscience. We link these two previously disconnected literatures by raising the tentative idea that *all states of flow are primary states* (but not vice-versa). In other words, all the descriptions associated with feelings of flow are consistent with the characteristics of primary states of cognition, while it is clear that not all primary states involve flow.

Currently, mainly because of their epistemological origins, the presence of states of flow and primary states are generally established by different but complementary methods: primary states are related to high entropy in brain functions to be found in quantitative properties of neural measurements, while states of flow are found by subjective reports. Some effort toward finding biomarkers of states of flow have been reported in a study undertaken with theatre artists (Noy et al., 2015). That study presented kinematic (CC motion) and physiological evidence (heart rate and subjective ratings) consistent with the subjective reports of the artists. Our study reveals that LZ complexity is one such potential marker. However, further experimental evidence will be required to fully corroborate this claim.

In this multidisciplinary study, using standard methods from computational neuroscience and psychology we provided evidence that the improvisatory state of mind in musicians can be conceived of as both a primary state and a state of flow, as would need to be the case if all states of flow are primary states.

The identification of the improvisatory state as being a primary state is supported by the higher level of LZ complexity found in the EEG signals recorded during the improvised performance. Moreover, the LZ increase was mainly localized in the right hemisphere, suggesting more engagement in a creative process during the improvised performance. The LZ effects were further supported by the profile of the power spectrum found in the EEG signals of the prepared and improvised performances, which resemble the transition between sleep and awake states as reported in the literature.

Characterizing the improvisatory state of mind as involving elements of a state of flow is supported by the musical analysis, which reveals features of the improvised performance such as longer-term phrasing gestures, and “mind-reading” in the passing of improvised gestures from one to the other. This is supported by the audience ratings, which found the improvised performance more emotionally compelling and musically convincing. An additional element that supports the state of flow in the improvised performance is the existence of longer and more flowing musical gestures, which are suggested by both the musical analysis and the reduced amount of uncorrelated fast movements in the motion analysis. The features of the performance found in these performances of Schubert are similar in nature to the features discovered in an earlier study when different musicians performed music by Telemann and Ravel, thus suggesting that these are quite general, high level, features of the improvisatory approach (beyond particular stylistic devices of different historical periods, or specific performers).

A significant question about the improvisatory state of mind, not previously addressed in the literature, is whether it is

transferable from musicians to audience members. The results obtained from the EEG measurements and the psychological questionnaires both suggest this transfer is possible, although the fact that only three out of four audience members showed the LZ effect demonstrates that other factors not measured here (e.g., focus of attention) may intervene. Interestingly, our results suggest that this transfer is not affected by visual aspects of the performance, as the most heavily affected audience member was actually blindfolded and hence only listening to the performance. Moreover, the fact that there was less movement displayed by the musicians during the improvised than during the prepared version, and the fact that the correlated movement do not increase significantly, suggests that the causes of the change in brain activity of the audience is not due to the musicians’ movements. Moreover, musical training seems not to affect the transfer of the improvised state of mind to an audience, since the effects shown both by questionnaire and also by EEG measurement were present in people with both higher and lower levels of musical training. This is encouraging, as it suggests that this experience is open to a broad range of people, not just those schooled in formal elements of musical language. This may suggest that the phenomenon is driven in part by underlying universal elements of expression (Cohen and Inbar, 2002; Godoy and Jorgensen, 2012). Further support for the relevance of reference to universal elements of expression is the re-appearance of similar gestures of musical expression by different musicians, performing different musical styles to different audiences in two different studies, when the performance consisted of improvised approach. It is therefore tempting to say that the improvisatory state of mind is a specific state of flow, which is in turn a specific kind of primary state. This would require however to find a specific difference that distinguishes the improvised state of mind from other states of flow. Three ingredients seem to be particularly distinctive of group musical improvisation: real-time creativity, shared risk-taking, and a feeling of enhanced listening/togetherness. This latter phenomenon has been explored in the context of movement interaction (Noy et al., 2015), and also in collective musical performance (Müller and Lindenberger, 2011). Some recent studies have also reported inter-brain synchronization between musicians that are performing together (Sänger et al., 2012, 2013; Müller et al., 2013). The statistical framework used in Dumas et al. (2010) is appropriate for such explorations. However, the experimental protocol of the current study was not suitable for exploring this issue in the current data. The approach taken here does however offer the prospect of discovering further commonalities across improvising performers, and between performers and audience.

Do the Body Movements of Musicians as Visually Experienced by Audience Members Affect the Magnitude of Their Response to the Improvised Performances?

Musicians moved significantly less during the improvised performance in comparison to the prepared performance. Since both EEG complexity and audience ratings increased for the improvised performance, these increases could not be attributed to more body movement. This is confirmed by the

comparison between those audience members both seeing and hearing the performance, and those only hearing it. Seeing the performers made no significant difference to the response. A plausible explanation for the lower level of movement in the improvised performances is that such movements are linked to prominent metrical beats. The analysis of the performances' sonic characteristics has shown that the improvised performances emphasize longer beats ("hyper-measures") and de-emphasize individual, shorter-term beats.

Does the Level of Musical Training or Knowledge of Audience Members Affect Their Response to the Improvised Performances?

Some post performance rating scales were sensitive to the differences between the prepared and improvised version, with the improvised version rated higher than the prepared version. This effect did not depend on the level of musical training of audience members. However, those familiar with Schubert judged the improvised version to be more emotionally compelling than did those unfamiliar with Schubert. Arguably this may be a response to "novelty," as evidence exists that musical emotionality is linked to the level of unexpectedness of what is experienced (e.g., Steinbeis et al., 2006). For those unfamiliar with Schubert, both performances would be relatively novel. For those familiar with Schubert, the improvised version would be experienced as more novel than the prepared version.

Do the Objective Performance Characteristics That Distinguish Improvised Performances of Telemann and Ravel Extend to the Music of a Different Period Exemplified by Schubert?

The analysis of improvised performance characteristics shows a significant convergence across three separate classical periods, in a common more free use of timbral variations, and longer temporal and dynamic units, which de-emphasize individual beats and bars, as well as showing more "mind reading" and risk-taking between performers. This gives us some confidence that we are tapping quite general features of the improvisatory approach which at least to some extent transcend genres and periods, and may therefore reflect more universal features of human behavior, consistent with the postulated existence of a biologically universal primary state which is to some extent driving behavior during the application of an improvisatory approach.

In addition, there is a strong suggestion both from the audience responses (of "more musically convincing") and also from the critical listening of the musicians, that the improvised performances were not only more impactful, but had a higher artistic quality.

CONCLUDING REMARKS

The research we have presented indicates that improvisation is related to a special state of mind, both amongst the performers and their listeners. The creation of music and its appreciation is a highly multifaceted phenomenon, and therefore developing insight about its nature necessitates research that combines assessments of physiological, psychological and

interpersonal communication. We believe that an improved integrated understanding of psychological and neuroscientific aspects of improvisation is of fundamental importance.

The current increase in the number of mental health cases that our society is experiencing may be related to a lack of ability to apply an improvisatory attitude during a daily life that becomes ever more unpredictable. To study how classical musicians are able, at will, to switch between improvised and non-improvised performance modes presents a unique opportunity, in which a careful comparison between these two ways of behaving can be carried out. What we noticed may suggest that, unlike the prepared performances, in improvisatory state of mind the musicians aim spontaneously toward the macro-structure, while the "local" tasks are performed more successfully, with less effort and anxiety, and in full accordance with the definition of a flow state presented in Csikszentmihalyi (1975, 1997).

It would be interesting in future research to develop measurement techniques that are minimally intrusive though still allow recording of both individual and collective brain, body and psychological responses during concerts. The closer the research can get to a real-life concert situation, the more relevant the findings become, as the corresponding objective and subjective findings might better reflect fundamental elements of human experience. These insights might contribute to deepen our understanding of the musical experience, which in turn can help to improve artistic and pedagogical praxis. Moreover, we hope that our findings can motivate further investigations on the effects of improvisation in well-being, potentially relevant to the links between performing arts and therapy.

ETHICS STATEMENT

The experiments reported in our manuscript were part of the protocol approved by the Ethics Committee of Guildhall School of Music and Drama. A separate ethics approval for the research reported in our manuscript was not required as per Imperial College Research Ethics Committee's guidelines as well as national regulations.

AUTHOR CONTRIBUTIONS

DD designed the concert experiment (the performance followed DD's approach to classical improvisation and its applications on performance), analyzed the performances and contributed to the writing process. HJ took the overall lead on the design and analysis of the neuroscience component of the study, and contributed to writing process. PM-M contributed to the pre-processing and analysis of EEG data. MM-S contributed to the setup of the experiment and its logistics; gathered, managed and processed the movement data. HR contributed to the analysis of EEG data. FR contributed to the analysis of the EEG data and the movement data, to the interdisciplinary analysis, and to the overall writing process. JS designed and analyzed the audience questionnaire, contributed to overall project design and management, and led the drafting process.

FUNDING

MM-S has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 743623. FR was supported by the European Union's H2020 research and innovation programme, under the Marie Skłodowska-Curie grant agreement No. 702981.

ACKNOWLEDGMENTS

Björn Crütz and Paul Perreijin, NCLogics AG, Munich; Rosie Bowker, Flautist, London; Guillaume Dumas, Institut Pasteur, Paris; Markus Muller, Centro de Investigación en Ciencias, Cuernavaca; Mark Gilensman, double-bass player and researcher,

Basel. Lu Chao, Data Science Institute. Open access fees were kindly covered by Imperial College Library.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01341/full#supplementary-material>

Video 1 | Bars 7–8 (flute solo opening 3 notes) prepared (strict) version.

Video 2 | Bars 7–8 (flute solo opening 3 notes) improvised state of mind version.

Video 3 | Bars 8–12 prepared (strict) version.

Video 4 | Bars 8–12 improvised state of mind.

Video 5 | Bars 165–177 prepared (strict) version.

Video 6 | Bars 165–177 improvised state of mind.

REFERENCES

- Abert, H. (2007). *W. A. Mozart* New Haven, CT: Yale University Press.
- Achermann, P., Rusterholz, T., Dürr, R., König, T., and Tarokh, L. (2016). Global field synchronization reveals rapid eye movement sleep as most synchronized brain state in the human EEG. *R. Soc. Open Sci.* 3:160201. doi: 10.1098/rsos.160201
- Azzara, C. D., and Snell, A. H. (2016). "Assessment of improvisation in music," in *Oxford Handbooks Online* (Oxford: Oxford University Press), 1–24. doi: 10.1093/oxfordhb/9780199935321.013.103
- Berkowitz, A. (2010). *The Improvising Mind. Cognition and Creativity in the Musical Moment*. Oxford: Oxford University Press.
- Boly, M., Sasai, S., Gosseries, O., Oizumi, M., Casali, A., Massimini, M., et al. (2015). Stimulus set meaningfulness and neurophysiological differentiation: a functional magnetic resonance imaging study. *PLoS ONE* 10:e0125337. doi: 10.1371/journal.pone.0125337
- Carhart-Harris, R. (2018). The entropic brain - revisited. *Neuropharmacology*. doi: 10.1016/j.neuropharm.2018.03.010. [Epub ahead of print].
- Carhart-Harris, R. L., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., et al. (2014). The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. *Front. Hum. Neurosci.* 8:20. doi: 10.3389/fnhum.2014.00020
- Carhart-Harris, R. L., and Nutt, D. J. (2010). User perceptions of the benefits and harms of hallucinogenic drug use: a web-based questionnaire study. *J. Subst. Use* 15, 283–300. doi: 10.3109/14659890903271624
- Casali, A., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., and Casali, K. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 5:198ra105. doi: 10.1126/scitranslmed.3006294
- Chirico, A., Serino, S., Cipresso, P., Gaggioli, A., and Riva, G. (2015). When music "flows". state and trait in musical performance, composition and listening: a systematic review. *Front. Psychol.* 6:906. doi: 10.3389/fpsyg.2015.00906
- Cohen, D., and Inbar, E. (2002). "Music imagery as related to schemata of emotional expression in music and on the prosodic level of speech," in *Music Imagery*, eds R. Godoy, and H. Jorgensen (Lisse: Swets and Zeitlinger), 137–159.
- Cone, E. T. (1968). *Musical Form and Musical Performance*. New York, NY: Norton.
- Cover, T., and Thomas, J. (2012). *Elements of Information Theory*. Somerset: Wiley.
- Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. San Francisco, CA: Jossey-Bass.
- Csikszentmihalyi, M. (1996). *Creativity*. New York, NY: Harper Collins.
- Csikszentmihalyi, M. (1997). *Finding Flow: The Psychology of Engagement with Everyday Life*. New York, NY: Basic Books.
- Després, J.-P., Burnard, P., and Stévanec, S. (2017). Expert western classical music improvisers' strategies. *J. Res. Music Educ.* 65, 139–162. doi: 10.1177/0022429417710777
- Dobson, M. (2008). "Exploring classical music concert attendance: the effects of concert venue and familiarity on audience experience," in *Proceedings of the First International Conference of Students of Systematic Musicology*, eds M. M. Marin, M. Knoche, and R. Parncutt (Graz), 14–15. Available online at: <http://www.uni-graz.at/muwi3www/SysMus08/>
- Dolan, D., Sloboda, J. A., Jeldtoft-Jensen, H., Cruts, B., and Feygelson, E. (2013). The improvisatory approach to classical music performance: an empirical investigation into its characteristics and impact. *Music Perform. Res.* 6, 1–38. Available online at: [http://mpronline.net/Issues/Volume%206%20\[2013\]/MPR0073.pdf](http://mpronline.net/Issues/Volume%206%20[2013]/MPR0073.pdf)
- Donnay, G., Rankin, S., Lopez-Gonzalez, M., Jiradejvong, P., and Limb, C. (2014). Neural substrates of interactive musical improvisation: an fmri study of 'trading fours' in jazz. *PLoS ONE* 9:e88665. doi: 10.1371/journal.pone.0088665
- Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., and Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS ONE* 5:e12166. doi: 10.1371/journal.pone.0012166
- Eigeldinger, J. J. (1986). *Chopin: Pianist and Teacher as Seen by his Pupils*. Cambridge: Cambridge University Press.
- Godoy, R. I., and Jorgensen, H. (eds.). (2012). *Musical Imagery*. New York, NY: Routledge.
- Goertzen, V. (1996). By way of introduction: Preluding by 18th_ and 19th-century pianists. *J. Musicol.* 14, 299–337.
- Griffiths, R. R., Richards, W. A., Johnson, M. W., McCann, U. D., and Jesse, R. (2008). Mystical-type experiences occasioned by psilocybin mediate the attribution of personal meaning and spiritual significance 14 months later. *J. Psychopharmacol.* 22, 621–632. doi: 10.1177/0269881108094300
- Grosshauser, T., Tröster, G., Bertsch, M., and Thul, A. (2015). "Sensor and software technologies for lip pressure measurements in trumpet and cornet playing - from lab to classroom," in *Proceedings of the 12th International Conference on Sound and Music Computing* (Maynooth).
- Hamilton, K. (2008). *After the Golden Age: Romantic Pianism and Modern Performance*. Oxford: Oxford University Press
- Hutka, S. A., Bidelman, G. M., and Moreno, S. (2013). Brain signal variability as a window into the bidirectionality between music and language processing: moving from a linear to a nonlinear model. *Front. Psychol.* 4:984. doi: 10.3389/fpsyg.2013.00984
- Klem, G., Lüders, H., Jasper, H., and Elger, C. (1999). "The ten-twenty electrode system of the International Federation," in *Recommendations for the Practice of Clinical Neurophysiology: Guidelines of the International Federation of Clinical Physiology*, eds G. Deuschl and A. Eisen (New York, NY: Elsevier Science), 3–6.
- Leech-Wilkinson, D. (2016). Classical music as enforced utopia. *Arts Humaniti. High. Educ.* 15, 325–336. doi: 10.1177/1474022216647706
- Lopata, J., Nowicki, E., and Joannis, M. (2017). Creativity as a distinct trainable mental state: an EEG study of musical improvisation. *Neuropsychologia* 99, 246–258. doi: 10.1016/j.neuropsychologia.2017.03.020

- Love, S. (2017). An ecological description of jazz improvisation. *Psychomusicology* 27, 31–44. doi: 10.1037/pmu0000173
- Lowis, M. J. (2002). Music as a trigger for peak experiences among a college staff population. *Creativ. Res. J.* 14, 351–359. doi: 10.1207/S15326934CRJ1434_6
- MacLean, K. A., Johnson, M. W., and Griffiths, R. R. (2011). Mystical experiences occasioned by the hallucinogen psilocybin lead to increases in the personality domain of openness. *J. Psychopharmacol.* 25, 1453–1461. doi: 10.1177/0269881111420188
- MacRitchie, J., and McPherson, A. P. (2015). Integrating optical finger motion tracking with surface touch events. *Front. Psychol.* 6:702. doi: 10.3389/fpsyg.2015.00702
- Mézard, M., and Montanari, A. (2009). *Information Physics and Computation*. Oxford: Oxford University Press.
- Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PLoS ONE* 9:e89642. doi: 10.1371/journal.pone.0089642
- Müller, V., and Lindenberger, U. (2011). Cardiac and respiratory patterns synchronize between persons during choir singing. *PLoS ONE* 6:e24893. doi: 10.1371/journal.pone.0024893
- Müller, V., Sängler, J., and Lindenberger, U. (2013). Intra- and inter-brain synchronization during musical improvisation on the guitar. *PLoS ONE* 8:e73852. doi: 10.1371/journal.pone.0073852
- Norgaard, M. (2011). Descriptions of improvisational thinking by artist-level jazz musicians. *J. Res. Music Educ.* 59, 109–127. doi: 10.1177/0022429411405669
- Norgaard, M. (2014). How jazz musicians improvise. *Music Percept.* 31, 271–287. doi: 10.1525/mp.2014.31.3.271
- Noy, L., Levit-Binun, N., and Golland, Y. (2015). Being in the zone: physiological markers of togetherness in joint improvisation. *Front. Hum. Neurosci.* 9:187. doi: 10.3389/fnhum.2015.00187
- Pinho, A., de Manzano, O., Fransson, P., Eriksson, H., and Ullen, F. (2014). Connecting to create: expertise in musical improvisation is associated with increased functional connectivity between premotor and prefrontal areas. *J. Neurosci.* 34, 6156–6163. doi: 10.1523/JNEUROSCI.4769-13.2014
- Pitts, S. E. (2005). What makes an audience? Investigating the roles and experiences of listeners at a chamber music festival. *Music Lett.* 86, 257–269. doi: 10.1093/ml/gci035
- Pitts, S. E., and Spencer, C. P. (2007). Loyalty and longevity in audience listening: Investigating experiences of attendance at a chamber music festival. *Music Lett.* 89, 227–238. doi: 10.1093/ml/gcm084
- Radhakrishnan, N., and Gangadhar, B. (1998). Estimating regularity in epileptic seizure time-series data. *IEEE Eng. Med. Biol. Mag.* 17, 89–94. doi: 10.1109/51.677174
- Sängler, J., Müller, V., and Lindenberger, U. (2012). Intra- and interbrain synchronization and network properties when playing guitar in duets. *Front. Hum. Neurosci.* 6:312. doi: 10.3389/fnhum.2012.00312
- Sängler, J., Müller, V., and Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. *Front. Hum. Neurosci.* 7:234. doi: 10.3389/fnhum.2013.00234
- Schartner, M., Carhart-Harris, R., Barrett, A., Seth, A., and Muthukumaraswamy, S. (2017). Increased spontaneous MEG signal diversity for psychoactive doses of ketamine, LSD and psilocybin. *Sci. Rep.* 7:46421. doi: 10.1038/srep46421
- Schartner, M., Seth, A., Noirhomme, Q., Boly, M., Bruno, M., Laureys, S., et al. (2015). Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. *PLoS ONE* 10:e0133532. doi: 10.1371/journal.pone.0133532
- Shehan Campbell, P., Myers, D., and Sarath, E. (2014). *Transforming Music Study from its Foundations: A Manifesto for Progressive Change in the Undergraduate Preparation of Music Majors*. Report of the Task Force on the Undergraduate Music Major. The College Music Society. Available online at: <http://www.mtosmt.org/issues/mto.16.22.1/manifesto.pdf>
- Steinbeis, N., Koelsch, S., and Sloboda, J. A. (2006). The role of musical structure in emotion: investigating neural, physiological and subjective emotional responses to harmonic expectancy violations. *J. Cogn. Neurosci.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380
- Tervaniemi, M., Janhunen, L., Kruck, S., Putkinen, V., and Huottilainen, M. (2016). Auditory profiles of classical, jazz, and rock musicians: genre-specific sensitivity to musical sound features. *Front. Psychol.* 6:1900. doi: 10.3389/fpsyg.2015.01900
- Thompson, S. (2006). Audience responses to a live orchestral concert. *Musicae Scientiae* 10, 215–244. doi: 10.1177/102986490601000204
- Thompson, S. (2007). Determinants of listeners' enjoyment of a performance. *Psychol. Music* 35, 20–36. doi: 10.1177/0305735607068886
- Wan, X., Cruts, B., and Jensen, H. J. (2014). The causal inference of cortical neural networks during music improvisations. *PLoS ONE* 9:e112776. doi: 10.1371/journal.pone.0112776
- Zhang, X., Roy, R., and Jensen, E. (2001). EEG complexity as a measure of depth of anesthesia for patients. *IEEE Trans. Biomed. Eng.* 48, 1424–1433. doi: 10.1109/10.966601
- Zhang, Z. (2012). Microsoft Kinect Sensor and its effect. *IEEE Multimedia* 19, 4–10. doi: 10.1109/mmul.2012.24
- Ziv, J. (1978). Coding theorems for individual sequences. *IEEE Trans. Inf. Theory* 24, 405–412. doi: 10.1109/TIT.1978.1055911

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 Dolan, Jensen, Mediano, Molina-Solana, Rajpal, Rosas and Sloboda. 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.



Let's Read a Poem! What Type of Poetry Boosts Creativity?

Małgorzata Osowiecka^{1*} and Alina Kolańczyk²

¹ Warsaw Faculty of Psychology, SWPS University of Social Sciences and Humanities, Warsaw, Poland, ² Faculty in Sopot, SWPS University of Social Sciences and Humanities, Sopot, Poland

OPEN ACCESS

Edited by:

Ian Hocking,
Canterbury Christ Church University,
United Kingdom

Reviewed by:

Paul M. Camic,
Canterbury Christ Church University,
United Kingdom
Shelly Marie Kemp,
University of Chester, United Kingdom

*Correspondence:

Małgorzata Osowiecka
mosowiecka@swps.edu.pl;
maggieosa@gmail.com

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 30 April 2018

Accepted: 03 September 2018

Published: 21 September 2018

Citation:

Osowiecka M and Kolańczyk A (2018)
Let's Read a Poem! What Type
of Poetry Boosts Creativity?
Front. Psychol. 9:1781.
doi: 10.3389/fpsyg.2018.01781

Poetry is one of the most creative uses of language. Yet the influence of poetry on creativity has received little attention. The present research aimed to determine how the reception of different types of poetry affect creativity levels. In two experimental studies, participants were assigned to two conditions: poetry reading and non-poetic text reading. Participants read poems (Study 1 = narrative/open metaphors; Study 2 = descriptive/conventional metaphors) or control pieces of non-poetic text. Before and after the reading manipulation, participants were given a test to determine levels of divergent thinking (DT; i.e., fluency, flexibility, and originality). Additionally, in both studies, the impact of frequent contact with poetry was examined. In Study 1 ($N = 107$), participants showed increased fluency and flexibility after reading a narrative poem, while participants who read the non-poetic text showed a decrease in fluency and originality. In Study 2 ($N = 131$) reception of conventional, closed metaphorization significantly lowered fluency and flexibility of thinking (compared to reading non-poetic text). The most critical finding was that poetry exposure could either increase or decrease creativity level depending on the type of poetic metaphors and style of poetic narration. Furthermore, results indicate that long-term exposure to poetry is associated with creativity. This interest in poetry can be explained by an ability to immerse oneself in a poetry content (i.e., a type of empathy) and the need for cognitive stimulation. Thus, this paper contributes a new perspective on exposure to poetry in the context of creativity and discusses possible individual differences that may affect how this type of art is received. However, future research is necessary to examine these associations further.

Keywords: creativity, divergent thinking, metaphor, poetry reception, language

INTRODUCTION

Creativity is often understood in different ways. In an elitist view, creativity means eminent works of art created by great, gifted artists. In contrast, creativity has also been described as a common cognitive process, which can be improved (Finke et al., 1992). This more popular approach has been labeled by Csikszentmihalyi (1996) as "little c Creativity." Previous research (Mednick, 1962) has shown that creative thinking is based on flatter concept hierarchies, enabling remote associations to be more easily made. Csikszentmihalyi states that this kind of creativity is part of everyday human life, and can be observed even in young children. This type of "common" creativity results in more efficient problem solving, better performance on tasks measuring creative potential, and can even bring about the production of outstanding works of art. The current research concentrates on

“little c Creativity,” which can be improved by specific interventions under specific circumstances, and then observed and measured (Guilford, 1950; Finke et al., 1992; Runco, 1999).

In this article, we examined whether the creative potential of a poem can be beneficial for receivers by testing whether one-time reception of poetry can influence the quality of divergent thinking (DT; i.e., multidirectional and/or potentially creative thinking). Additionally, we investigated if this impact depends on the type of poetic metaphors and/or the style of poetic narration.

There are several studies that have examined how humans produce metaphors (Paivio, 1979; Chiappe and Chiappe, 2007; Silvia and Beaty, 2012; Beaty and Silvia, 2013), but little is known about metaphor comprehension, especially within the context of poetry. This research has inspired many books that attempt to teach the skills necessary to generate imaginative and interesting metaphors (e.g., Plotnik, 2007). It may be that the ability to associate remote ideas, facts, and elements of the environment, which is a key factor in metaphor production, may also be a key factor in creativity. Thus, these skills that can be taught to improve metaphorization may also overlap with skills to improve general creative ability.

Most psychological research on poetry has focused on the influence of text structure (i.e., rhythm, rhymes) on emotional reception of poems (e.g., Jakobson, 1960; Turner and Pöppel, 1983; Lerdahl, 2001; Obermeier et al., 2013). Additionally, many studies that have focused on poets' creativity have also collected data revealing links between mental disorders and functioning (e.g., Stirman and Pennebaker, 2001; Djikic et al., 2006). Further, previous research has also examined the relationship between poetic training and creativity (e.g., Baer, 1996; Andonovska-Trajkovska, 2008; Cheng et al., 2010). However, the current manuscript focuses on the influence of poems as creative products that may affect receivers' levels of creative thinking. This influence, however, likely depends on the type of poetry received.

The efficiency of DT is a key measure of idea generation (e.g., Baer, 1996; Runco, 1999; Nęcka, 2012). In contrast to convergent thinking, DT enables problem solving in diverse and potentially valuable ways. It often involves redefining the problem, referring to analogies, redirecting one's thoughts, and breaking barriers in thinking. Previous research has found that spreading activation in the semantic network is indicative of DT (Martindale, 1989; Ashton-James and Chartrand, 2009; Kaufman and Beghetto, 2009). Developing associations between distant ideas is a basic mechanism of creative thinking (Mednick, 1962). For instance, Benedek et al. (2012) provided evidence that the ability to generate remote associations makes creative problem solving easier. Gilhooly et al. (2007) showed that ignoring close associations (but choosing remote ones) and breaking the stiff, typical relationships between ideas plays a crucial role in effective DT. The current studies are based on the hypothesis that the process of DT can be supported by poetry comprehension.

Poetry, which contains remote associations described through metaphors and analogies, combines non-related notions in atypical ways (Lakoff and Johnson, 2003). In general, metaphoric expression often involves mapping between abstract and more concrete concepts (Glucksberg, 2001, 2003); therefore, the comprehension of metaphors requires the activation of a

broader set of semantic associations. This is due to connecting two remote parts of a metaphor (theme and vehicle) into a meaningful expression (Paivio, 1979; Kenett et al., 2018). Poetry reception can involve readiness to notice similarities between remote categories, which can be a crucial ability in generating creative ideas (e.g., Mednick, 1962; Koestler, 1964; Martindale, 1989). Training in metaphorical thinking results in the broadening of categories (Nęcka and Kubiak, 1989), which leads to increased DT (Trzebiński, 1981). Glucksberg et al. (1982) have shown that poetry reading broadens the scope of associations. Metaphor, based on remote associations, provides a new way of understanding reality and human feelings. In addition to fostering multidirectional and creative thinking, metaphor can also help individuals adjust to the surrounding world (Kolańczyk, 1991; Nęcka, 2012). Metaphorization is, structurally, the most essential element of the poetic art (e.g., Lakoff and Johnson, 2003; Kovecses, 2010). Rhythm, syllabification, and word combinations in well-written poetry construct a meaningful whole aside from very remote notions (Csikszentmihalyi, 1996). Thus, poetry comprehension can change readers' DT; however, this impact likely depends on type of poetic metaphors and the narration used by the poet.

Thinking expressed in metaphors always involves the flexible activation and manipulation of acquired knowledge (Benedek et al., 2014); even though metaphors are not always creative, even in poetry. Understanding a conventional metaphor is not intellectually challenging: comprehending such expressions is based on the retrieval of well-known meaning from memory (Kenett et al., 2018). For example, love can be understood metaphorically as a nutrient. The metaphors “starved for affection” and “given strength by love” are not particularly creative, as they are based on a highly conventional metaphor (i.e., love = nutrient). These metaphors are ostensibly viewed as new by receivers of poetry, although they are not flexible or original. Hausman (1989) writes about two specific types of metaphors; one he describes as impoverished, frozen, and closed; the other, he refers to as original, divergent, and open. It seems logical to use terms like closed/convergent and open/divergent when referring to metaphors, which can emphasize a functional dimension of how these types of metaphors are used in poetry and casual language. To the best of our knowledge, however, previous research has never introduced this distinction in terms of differences between metaphors. Instead, Beaty and Silvia (2013) uses the metaphor labels *conventional* (i.e., familiar) and *creative* (i.e., novel).

Until now, no typologies of metaphors have been introduced that highlight differences in how poetry is constructed and how this impacts recipients. It seems that poetry uses at least these two kinds of metaphorization. Both of these can be adaptive for the recipient, because creativity requires both accommodation and assimilation (Ayman-Nolley, 2010). Therefore, recipients' reception of novel and open metaphors could result in more flexible and original thinking, whereas reception of conventional, well known, and closed metaphors could result in less flexible and less creative problem-solving.

In addition to the types of metaphors used, poetry is also characterized by content. One conceptualization of poetry

describes it as a certain type of story, which is a separate and coherent whole, through which people express their thoughts and/or opinions (Heiden, 2014). In this case, the author can bring an abstract idea closer to the reader through narrative imagery. This type of poetry can result in the receiver taking on another's (i.e., the author's) point of view, hence improving creativity. Moreover, this narrative type of poetry is an open task for readers, because understanding is reached based on the receiver's own experience and understanding. The second type, noncreative poetry, is more conservative, and includes variously structured, commonplace (i.e., conventional) metaphors, which are often clichés based on common-sense regularities, and are sometimes the contents of parables or prayers. Metaphors in this type of poetry delineate and conventionalize meaning; they describe the world in ways known to everyone (e.g., Lakoff and Turner, 1989; Gibbs, 1994; Lakoff and Johnson, 2003; Kovecses, 2010).

The general goals of this research were to determine whether the reception of poetry stimulates creative thinking, and whether poetry's impact on creativity varies depending on the type of poetry. Accordingly, we formulated the following research hypotheses:

1. Reception of an unconventional, open metaphor poem will stimulate the generation of creative ideas (i.e., improves DT from baseline).
2. Reception of conventional poetry either will not influence, or will negatively influence the generation of creative ideas (i.e., no increase or decrease in DT from baseline).
3. DT will be increased after the reception of open metaphor poetry, when compared to reading a neutral text.
4. DT will be decreased after reception of conventional poetry, when compared to a neutral text.

In Study 1, participants were exposed to a poem with narrative imagery expressing an author's point of view and utilizing open metaphors. In Study 2, participants were exposed to a conventional poem that employed a biographical approach, comprised of commonplace metaphors and aphorisms.

STUDY 1

Methods

Participants

Participants were recruited from high-school classes. All participants resided in Poland. A total of 107 participants completed the study (M age = 17.46; SD = 1.03; 53 female). Students from the pool were randomly assigned to one of two groups. Upon entering the lab, participants were given a consent form and a brief explanation of the study procedures. The study was conducted in a group setting, with the number of participants ranging from 10 to 15. Participants provided written, informed consent, and were free to withdraw from the research at any time without giving reason or justification for withdrawing. Minors participated in research with written parental consent. Participants received points for behavior as compensation. Their

participation was anonymous. The study was approved by a local ethics committee (clearance number: WKE/S 15/VI/1).

Materials

DT Measurement

To measure DT, participants were administered versions of the Question Generation task (Chybicka, 2001). This task was conducted using a test-retest design (to observe creativity change). Participants listed as many questions as they could regarding an unambiguous picture (baseline image from Chybicka, 2001; post-test, a comparable version from Corbalan and Lopez, 1992). The fluency, flexibility, and the originality of answers were evaluated by three independent judges. Fluency was the total number of meaningful responses given by participant; flexibility (i.e., diversity of categories) was measured as the number of different categories; and originality was calculated as the number of original, novel, and interesting responses.

Poetry—Szymborska's Poem

In Study 1, we chose Szymborska's (2012) poem *Utopia* as an example of narrative, non-rhythmic poetry. In *Utopia*, Szymborska creates a sort of plot or story, which she conveys to the reader in a very metaphorical, condensed form. Szymborska's narration in *Utopia* is characterized by ethical and metaphysical themes (e.g., "As if all you can do here is leave and plunge, never to return, into the depths. Into unfathomable life & The Tree of Understanding, dazzlingly straight and simple, sprouts by the spring called Now I Get It"). Six independent judges, all of which were Polish language teachers, filled in a short scale which contained three questions about affectivity of the chosen poem (e.g., "the poem is neutral"). They confirmed that the poem was emotionally stable, allowing for control over the influence of both rhythm and emotion on participants' creativity.

Control Text

For the control text, we used the description of a cooking device (Speedcook, RPOL, Mielec, Poland). This description approximated the word count of a poem and did not contain any metaphors (e.g., "Our kitchen appliance has a classic, elegant design. This device could replace every cooking appliance, a steam cooking tool, and a juicer"). Device descriptions are often made according to the same pattern and in a comparable way. The description that we used contained close, functional associations between concepts. The text is constructed to provide concrete information to the recipient. The device description was obtained from an Internet website (Wachowicz, 2014).

Contact With Poetry Scale

We developed a scale to measure poetry contact that addressed passion, as well as frequency of reading poetry and taking part in poetic meetings. Agreement/disagreement with statements was assessed. Statements included "I am passionate about poetry," "In my free time, I very often read poems," "I write poems and share my work with others," "I have several favorite poets," "Sometimes, I put down my creative thoughts onto paper," and "I was once an unpublished writer." Participants answered the five items on a 5-point scale from 1 = strongly disagree to 5 = strongly agree.

The reliability of the tool, as measured by internal consistency, was satisfactory (Cronbach's $\alpha = 0.83$).

Procedure

First, participants read introductory information highlighting the importance of their participation in the study and a confidentiality statement (assuring that participants would remain anonymous and encouraging them to answer all questions truthfully). Then, participants received the first version of the Question Generation Task (Chybicka, 2001). Participants wrote questions about a picture printed on a piece of paper for 10 min. Next, participants were randomized into one of two groups: (a) the experimental group, which read the poem; or (b) the control group, which read the cooker description. Participants were instructed to silently read the poem twice, in a calm and attentive manner (Kraxenberger and Menninghaus, 2016). After reading the text, participants answered two questions; one regarding understanding the content ("*I understand the meaning of the text*") and the other an affective estimation of the text ("*In my opinion, the text is pleasant*"). Items were rated on a 6-point scale, with response options ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Then, participants completed a parallel version of the drawing from the Question Generation Task (Corbalan and Lopez, 1992). Finally, participants completed the devised scale concerning contact with poetry. Duration of the entire procedure was approximately 35 min. After completing the scale, participants were debriefed and thanked for their participation. We also collected postal addresses from participants who were interested in the results.

Data were analyzed using SPSS 24 (IBM, Armonk, NY, United States). The data from all participants were included in analyses and a significance level of $p < 0.05$ was adopted for all tests.

Results

All three DT indicators were scored by three independent raters. A Kendall's W of 1.00 was calculated for fluency at both time points; a W of 0.75 and 0.72 for flexibility in the first and the second measurement, respectively; and 0.76 for originality in both measurements (W greater than 0.70 = good concordance). All indicators were analyzed separately via three repeated-measures analyses of variances (ANOVAs) with effect of measurement (first vs. second) as the within-subjects factor and group (poetry vs. description) as the between-subjects factor.

A 2×2 (measurement \times group) repeated measures ANOVA for fluency revealed an interaction [$F(1,105) = 12.12, p < 0.001, \eta^2 = 0.1$], but no main effects. Pairwise comparisons showed a significant improvement in fluency scores on the second measurement compared to the first in the poetry group [$t(56) = 2.57, p = 0.013$; Cohen's $d = 0.35$]. Moreover, the control group differed in fluency across the measurements. Specifically, participants in this group demonstrated significantly lower scores in the second measurement than in the first [$t(52) = 2.44, p = 0.018$; Cohen's $d = 0.35$]. Extended data are shown in **Figure 1**.

A 2×2 (measurement \times group) repeated measures ANOVA for flexibility also revealed an interaction [$F(1,105) = 10.15,$

$p < 0.01, \eta^2 = 0.09$]. Further, a main effect of measurement was observed [$F(1,105) = 17.52, p < 0.001, \eta^2 = 0.14$]. The second picture of the DT task led to more flexible answers ($M = 4.83, SD = 1.63$) than did the first one ($M = 4.25, SD = 1.56$). Two-tailed, paired t -tests for two measurements in the poetry group yielded significant differences [$t(56) = 5.47, p = 0.001$; Cohen's $d = 0.75$]. Extended data are presented in **Figure 2**.

A 2×2 (measurement \times group) repeated measures ANOVA for originality also revealed an interaction [$F(1,105) = 23.03, p = 0.01, \eta^2 = 0.18$]. Additionally, a main effect of measurement was observed [$F(1,105) = 12.12, p < 0.01, \eta^2 = 0.11$]. The first picture in the creativity test triggered more original answers ($M = 2.85, SD = 1.18$) than did the second ($M = 2.34, SD = 1.71$). Two-tailed paired t -tests yielded significant differences between the first and the second measurement only in the description group [$t(50) = 5.09, p < 0.001$; Cohen's $d = 0.75$]. Extended data are shown in **Figure 3**.

To verify how individual differences in poetic interests are connected to DT, we also performed a linear regression analysis predicting DT on the first measurement (before the manipulation). As expected, flexibility was predicted by the level of poetic interests, $F(1,56) = 3.29, p = 0.075, b = 0.24$ (a near-significant trend). However, fluency and originality were not predicted by level of poetic interests. Further, no significant predictions were observed for the second measurement of creativity.

Discussion

Results of the experiment support our hypotheses to a large extent, however, there are some issues that remain to be elucidated. Reading of poetry improved two creativity indicators (fluency and flexibility), while reading of the control (descriptive) text caused a decline in fluency and originality. Although these results are interesting, the question of why reading poetry does not improve originality remains. It is possible that reading this type of poetic narration introduces insufficient changes to the semantic network, so that individuals were unable to improve in the only indicator of product quality (i.e., originality). Additionally, flexibility did not decrease as a result of reading instructions. Likely because the cooker is compared with similar devices, which requires looking at it from different perspectives. Moreover, frequent contact with poetry predicted flexibility. These results suggest that the reception of narrative and open poetry broadens activation of the semantic network and allows for flexible switching between remote categories; however, it is not connected with the creation of very original solutions.

The chosen poem combines both abstract and concrete concepts. The abstract ones (e.g., *obvious, understanding*) are explained in concrete or imaginative terms (e.g., *valley, tree*), which facilitate a distinct view of reality (Kirsch and Guthrie, 1984). Contact with this kind of poetry can diversify experience, which can lead to increased flexibility (Ritter et al., 2012). Hence, poetry reception may result in diverse idea generation. Flexibility is the ability to use various categories beyond the boundaries of their literal meaning. Many researchers agree that reception of poetry inhibits automatic associations, thereby

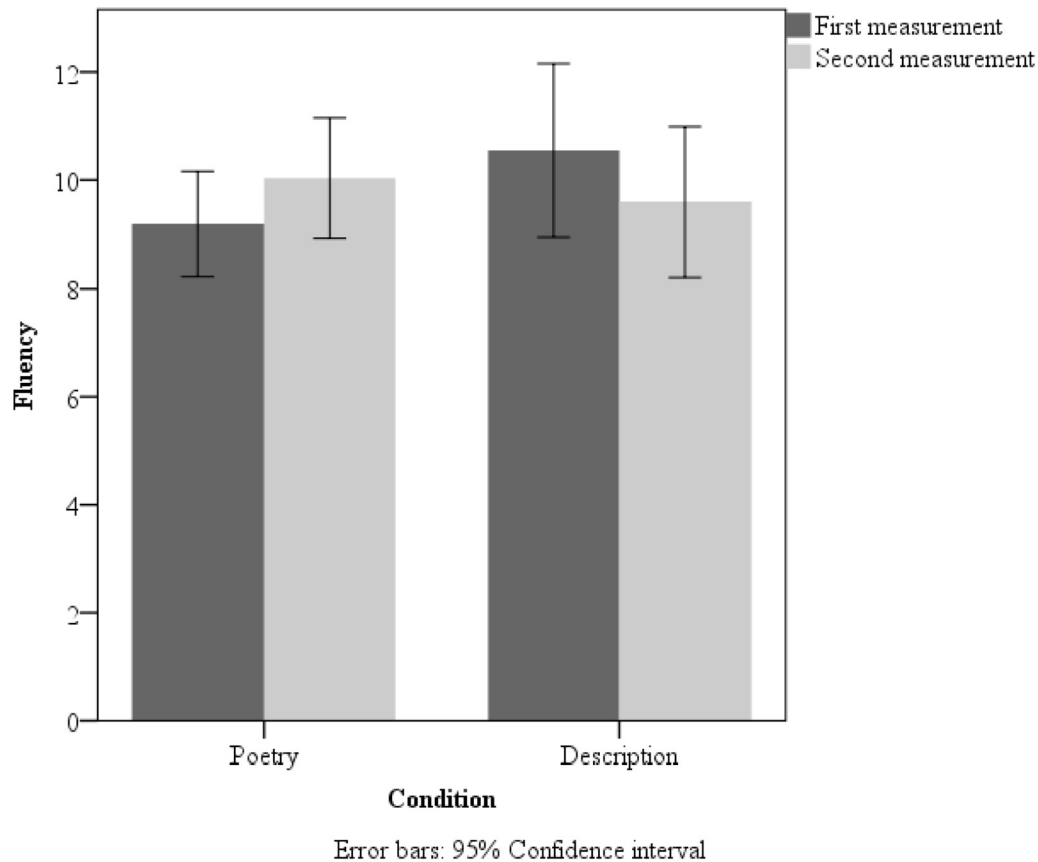


FIGURE 1 | Mean fluency scores in the poetry-reading group and description-reading group in the first and second measurement in Study 1. Error bars: 95% Confidence interval.

producing ideas without value (Kirsch and Guthrie, 1984; Halonen, 1995). Creative thinking is often connected with breaking typical patterns of thinking and seeing the world in another way (Amabile, 1996), which relates to intellectual risk-taking (Nickerson, 1999).

The lack of change in originality scores may be related to the character of the poem. *Utopia* is rather calm, balanced, and narrative. As such, it may be able to weaken resistance to seeing things from another point of view (flexibility). In contrast, reception of such a poem may inhibit original idea production until the whole of the poem is understood. Therefore, the reception of this type of poetry may have a buffering effect on intrinsically motivated original ideas. The purification of the dominant influence of the author's unique perspective is possible in more emotional and cathartic poetry. Thus, increased originality may be more visible after reception of cathartic metaphoric poems, which presents the extraordinary experience of a poet.

Finally, showing that the level of poetic interest predicts flexibility (measured prior to manipulation) is in line with previous research; specifically, that long-term contact with poetry is associated with creative problem solving (McGovern and Hogshead, 1990). As Sternberg and Lubart (1999) claim, people's

interest in poetry can increase creative potential understood as seeing problems in unique ways.

Study 1 showed the positive impact of narrative poetry on DT. Subsequently, Study 2 utilized conventional poetry, with the hypothesis that reception of this type of poetry would not enhance creativity. We wanted also reveal why individuals demonstrate spontaneous contact with poetry, which may be essential for receiving this kind of art, and thus increased performance on tasks requiring DT ability. These elements were empathy (i.e., the tendency to become immersed in the poetry content; Davis, 1983), and need for cognition (NFC; construed as willingness to interact with the cognitively demanding text of a poem; Cacioppo and Petty, 1982). Poems can be challenging cognitive tasks. As such, understanding a poem requires the creation of complex meaning from specific words and exploration of multifaceted ideas (Csikszentmihalyi, 1996).

We predicted that the variables listed above would be crucial for initial DT levels (i.e., baseline, recorded during the first DT test); but that these individual difference effects would disappear after the manipulation. We also predicted that reception of conventional poetry (and the control text) would lead to a poorer performance on the DT task after its reception.

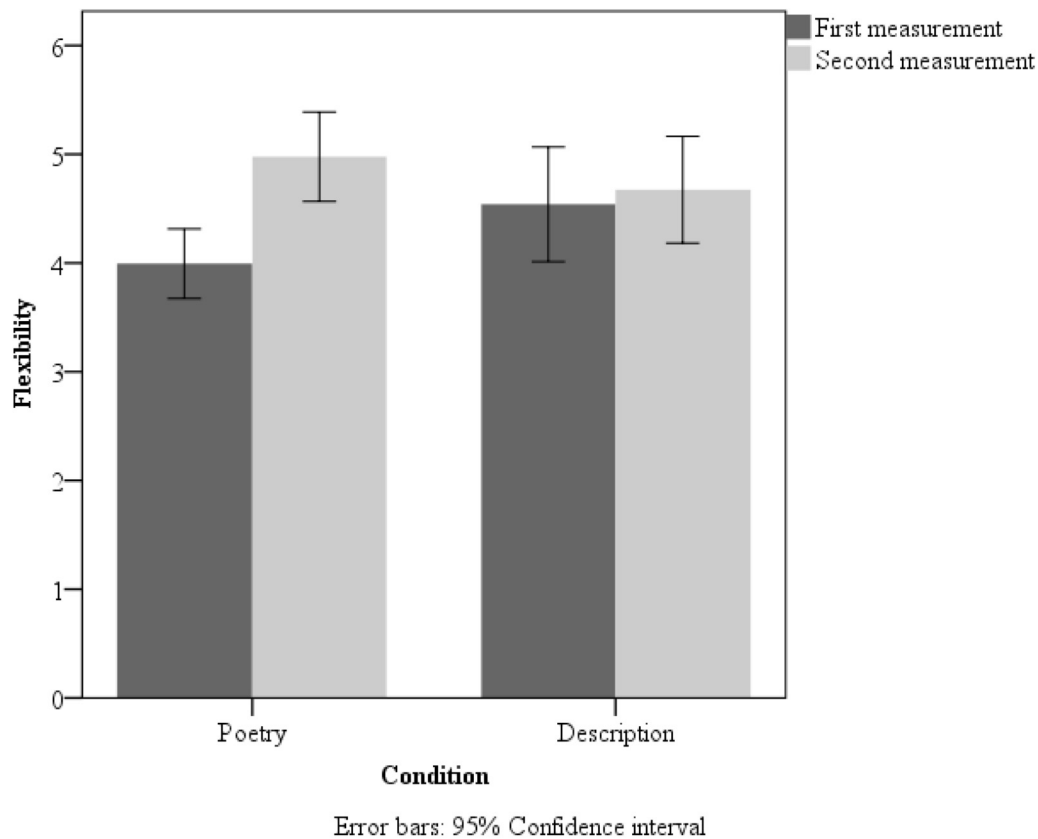


FIGURE 2 | Mean flexibility scores in the poetry-reading group and description-reading group in the first and second measurement in Study 1. Error bars: 95% Confidence interval.

STUDY 2

Methods

Participants

Participants were recruited from high-school classes. All participants resided in Poland. A total of 131 participants completed the study (M age = 16.36; SD = 0.71; 84 female). Students from this pool were randomly assigned to one of two groups. Upon entering the lab, participants were given a consent form and a brief explanation of the study procedures. The study was conducted in a group setting, with the number of participants ranging from 10 to 15. Participants provided written, informed consent, and were free to withdraw from the research at any time without giving reason or justification for withdrawing and received course credit as compensation. Minors participated in research with written parental consent. Participants received course credit for participation, and their participation was anonymous. The study was approved by a local ethics committee (clearance number: WKE/S 15/VI/1).

Materials

DT Measurement

DT measurement protocols for this study were identical to those used in Study 1.

Gustafson's Poem

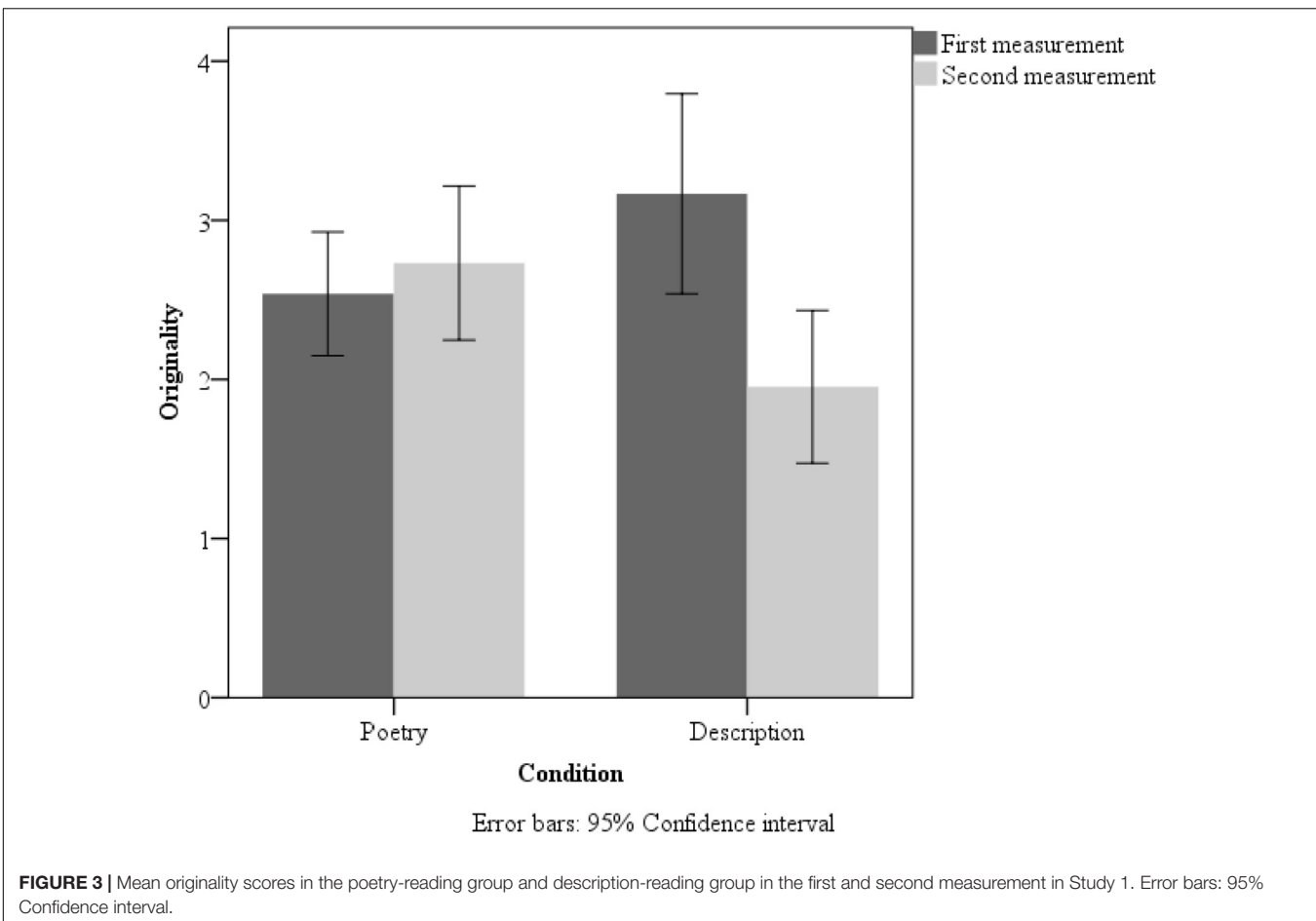
Lars Gustafson's poetry is philosophical; descriptive; and uses well-known metaphors of "life as a machine," which was very popular in the 20th century. We used the Polish version of Gustafsson (2013) poem, *Silence of The World before Bach*, which, in a very descriptive way, presents a biography of Bach and the changes in the world connected with his music/art works. It uses commonplace metaphors, which describe the world in well-known ways (e.g., "*Soprano never in helpless love twined round the gentler movements of the flute*"), making it an excellent example of conventional poetry. The chosen poem does not rhyme and is emotionally stable, which was confirmed by three judges, in a manner similar to Study 1.

Gustafson's Poem Description

For a control text, we created a description of the poem's content. It approximated the word count of the poem and did not contain any metaphors.

Contact With Poetry Scale

This scale was an extended version of the task created for Study 1, which measures passion for poetry, as well as frequency of poetry reading and taking part in poetic meetings (e.g., "*I am passionate about poetry*," "*In my free time I very often read poems*," and "*Poetry is incredibly difficult for me*"). Participants answered



the eight items on a 5-point scale from 1 = strongly disagree to 5 = strongly agree. The reliability of the tool, as measured by internal consistency, was satisfactory (Cronbach's $\alpha = 0.853$).

The Rational Experiential Inventory—NFC (Reflective) Scale

We used the Polish version of the Rational Experiential Inventory (REI; Epstein et al., 1996; Shiloh et al., 2002). This tool consists of two dimensions: an analytical-rational style of thinking and an intuitive-experimental style of thinking. The REI was devised based on the Myers-Briggs Type Indicator (Briggs and Myers, 1976) and the NFC scale (Cacioppo and Petty, 1982), which defines the type of motivation described by the authors as the need for knowledge cognition. The NFC scale was used to build a rational (reflective) REI scale, opposite of the intuition scale. The most important element of this measure for the current study was the NFC scale. The REI is a 40-item Likert scale with response options ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The reliability of this tool, as measured by internal consistency, was satisfactory (Cronbach's α for whole REI = 0.821, α for the NFC scale = 0.743).

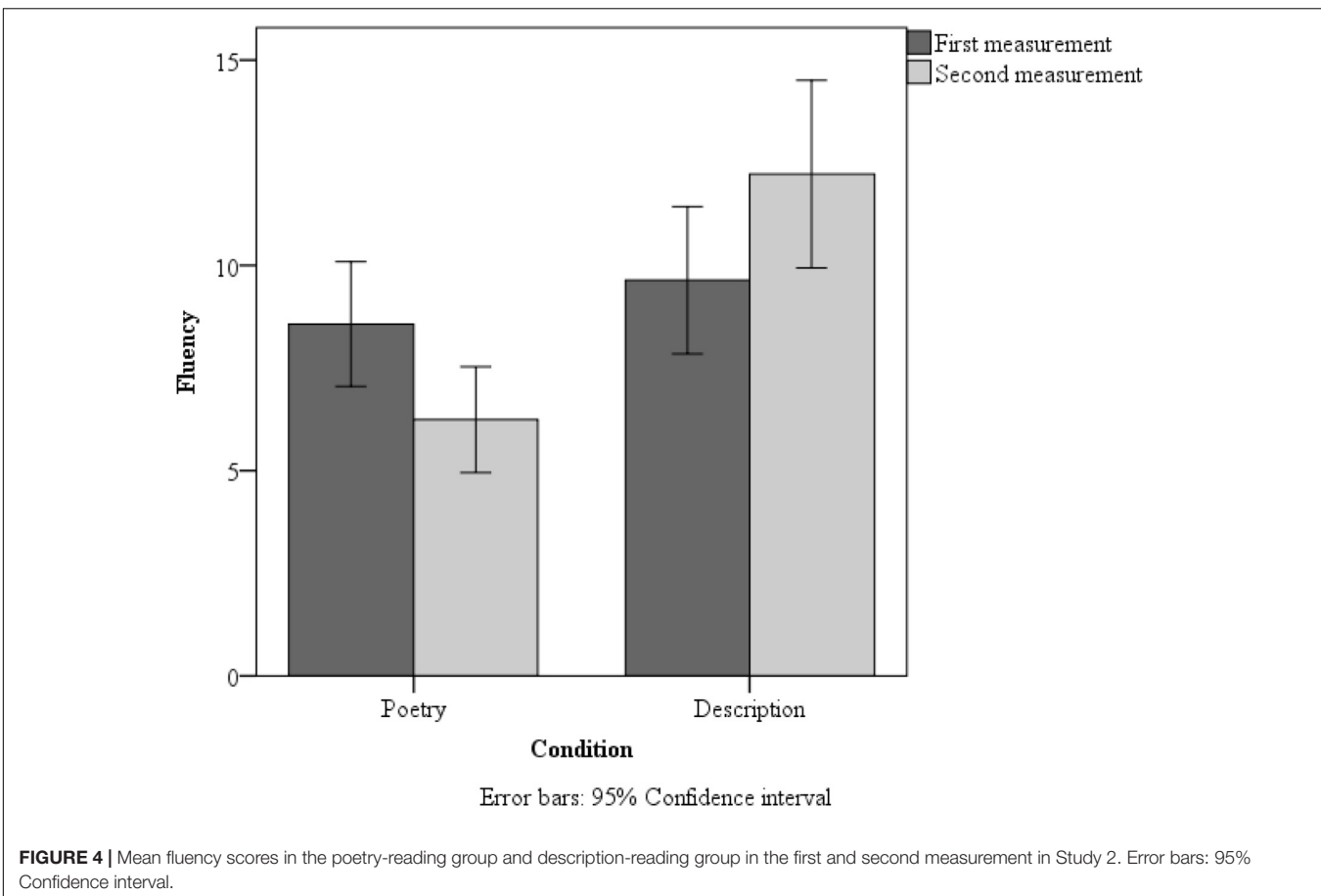
Interpersonal Reactivity Index (IRI)—Fantasy Scale

The IRI is a questionnaire addressing empathy. It consists of four scales: Perspective Taking, Fantasy, Empathic Concern, and

Personal Distress. In the current study, the Fantasy scale was used. This scale measures the tendency to imaginatively transpose oneself into fictional situations, as well as into the feelings and actions of fictitious characters in books, movies, and plays. This scale consists of 7 items (e.g., “*I really get involved with the feelings of the characters in a novel*,” “*I am usually objective when I watch a movie or play, and I do not often get completely caught up in it*”). The IRI involves a 5-point response option scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The reliability of the Fantasy Scale, as expressed by Cronbach's α , was 0.682.

Procedure

Participants first completed the baseline creativity test. Then, participants were randomized into one of two groups; (a) the experimental group that read the poem, and (b) the control group that read the description of its content. Participants read his/her respective documents twice. After the second reading, participants completed the second creativity test and completed the questionnaires listed above, using pen-and-paper procedures. The order of the creativity tests was counterbalanced across participants. After completing the scale, participants were debriefed and thanked for their participation. We also collected postal addresses from participants interested in the results.



Data were analyzed using SPSS 24 (IBM, Armonk, NY, United States). Two participants were excluded from analyses due to lack some data. A significance level of $p < 0.05$ was adopted for all tests.

Results

All three DT indicators were scored by five independent raters. Kendall's $W = 0.9$ for fluency in both measurements; $W = 0.78$ and 0.72 for flexibility in the first and the second measurement, respectively; and $W = 0.7$ for originality in both measurements. All indicators were analyzed separately by means of three repeated-measures ANOVAs with effect of measurement (first vs. second) as the within-subjects factor and group (poetry vs. description) as the between-subjects factor.

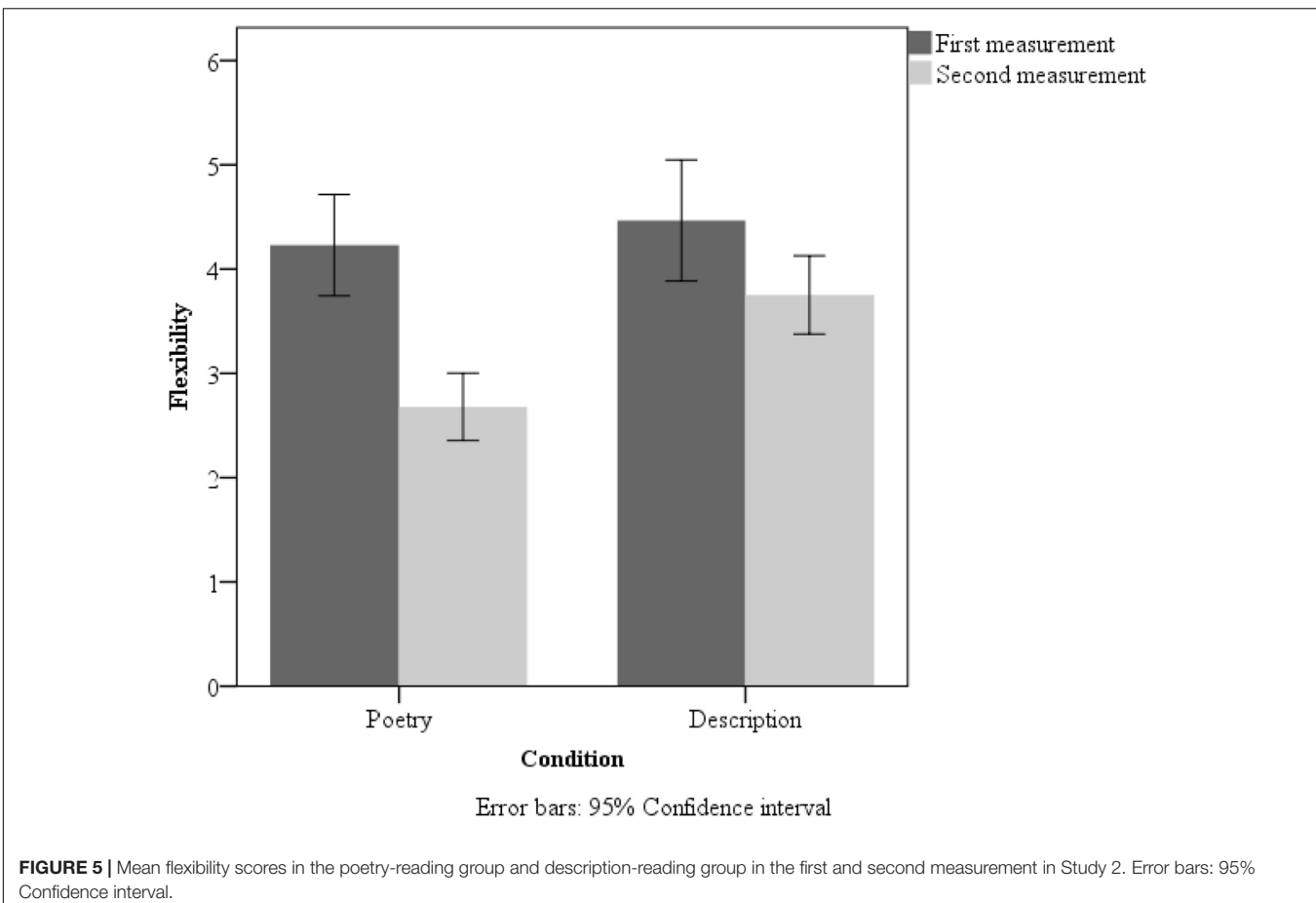
A 2×2 (measurement \times group) repeated measures ANOVA conducted for fluency revealed an interaction [$F(1,127) = 11.56$, $p = 0.01$, $\eta^2 = 0.08$]. Moreover, we found a main effect of Group [$F(1,127) = 12.35$, $p = 0.001$, $\eta^2 = 0.09$]. The poem made people less fluent ($M = 7.41$, $SD = 0.71$) than did the description ($M = 10.93$, $SD = 0.72$). Pairwise comparisons showed that, in the second measurement, the poetry group's fluency was significantly lower than the fluency of the description group [$t(127) = 4.61$, $p = 0.001$; Cohen's $d = 0.84$]. Two-tailed paired t -tests showed that the poetry group demonstrated a significant decrease in scores on the second measurement compared to the first measurement

[$t(65) = 2.52$, $p = 0.014$; Cohen's $d = 0.31$]. Furthermore, the description group demonstrated better scores on the second measurement than on the first [$t(62) = 2.31$, $p = 0.024$; Cohen's $d = 0.29$]. Extended data are shown in **Figure 4**.

A 2×2 (measurement \times group) repeated measures ANOVA for flexibility also revealed an interaction [$F(1,127) = 3.92$, $p = 0.05$, $\eta^2 = 0.03$]. Additionally, we found a main effect of group [$F(1,127) = 28.68$, $p < 0.001$, $\eta^2 = 0.18$]. The description triggered more flexible answers ($M = 4.11$, $SD = 0.17$) than did the poem ($M = 3.45$, $SD = 0.17$). We also found differences between the first and second measurement of flexibility in both the poetry [$t(65) = 5.64$; $p = 0.001$; Cohen's $d = 0.71$] and description groups [$t(62) = 2.21$, $p = 0.031$; Cohen's $d = 0.29$]. Two-tailed paired t -tests showed that flexibility of both groups dropped in the second measurement when we compared its level with the first measurement. Furthermore, we found differences between the poetry and the description groups in the second measurement [$t(127) = 4.34$, $p = 0.001$; Cohen's $d = 0.59$]. Two t -tests showed that poetry reception resulted in lower flexibility scores than description reception in the second measurement. Extended data are presented in **Figure 5**.

A 2×2 (measurement \times group) repeated measures ANOVA for originality yielded not significant interactions or main effects.

Next, we conducted linear regression analyses to determine whether the mean frequency of contact with poetry, fantasy



(empathy factor), and/or NFC predicted DT scores in the baseline measurement. Analyses showed that frequent contact with poetry positively predicted all parameters of DT [fluency, $F(1,127) = 21.49$, $p < 0.001$, $R^2 = 0.15$, $b = 0.38$; flexibility, $F(1,127) = 23.73$, $p < 0.001$, $R^2 = 0.16$, $b = 0.39$; and originality, $F(1,127) = 17.94$, $p < 0.001$, $R^2 = 0.13$, $b = 0.35$]. Further regression analyses yielded no significant associations between DT and fantasy, or DT and NFC.

We tried to explain the observed behavior—contact with poetry—in psychological terms. To elucidate the impacts of personality predictors on contact with poetry, we performed a single multiple regression analysis. The dependent variable was frequency of contact with poetry and the independent variables were fantasy and NFC. Results showed that the two-variables model was significant: $F(2,127) = 10.67$, $p < 0.001$, $R^2 = 0.15$. Fantasy was a slightly stronger predictor of contact with poetry/passion ($b = 0.26$) than was NFC ($b = 0.25$). As predicted, we found no significant effects regarding these variables in the second measurement.

Discussion

We found that contact with conventional, biographical poetry led to decreased indicators of DT. We also observed that people who received this type of poetry demonstrated less fluent and flexible thinking compared with those that read a description

of the same information. These results provide support for our hypothesis that idea generation is less likely after reception of narrative-conventional poetry, and that people are less creative after reading this kind of text, when compared to reading a neutral text.

Kovecses (2010) stated that a large body of poetry is constructed in a very conventional way (i.e., based on conceptual, conventional metaphors that are often used in everyday language). Such conventional metaphors (e.g., life is a journey; death is dark), as a part of our cognitive system, allow us to adapt to reality, but do not necessarily stimulate creativity (Lakoff and Turner, 1989). “The idea that metaphor constrains creativity might seem contrary to the widely held belief the metaphor somehow liberates the mind to engage in divergent thinking” (Gibbs, 1994, p. 7). Poets create novel, non-conventional poems through cognitive transformations: elaboration, extension, questioning, and combining (Lakoff and Turner, 1989). Therefore, it seems that the biographical, closed, and conventional poetry is also insufficient to stimulate creativity.

Our research confirms that contact with poetry, understood as long-term individual interest (not one-time contact), is associated with readers’ creativity. Accordingly, the results showed that frequent contact with poetry could be explained by individual differences, specifically increased ability to become absorbed in

the feelings of characters in a novel, as well as a stronger NFC. We can conclude that the features of the text, as well as the ability to actively perceive the poem, are key factors for appropriate poem reception. Noy and Noy-Sharav (2013) argue that the emotional message of art is always individually perceived. Silvia (2005), who refers to the appraisal theory of aesthetic emotions, claims that the evaluation of art, and not art itself, arouses emotions. Understanding of a poem requires the ability to actively follow and immerse oneself in the poetry content, which is an essential dimension of empathy (Davis, 1983). Experience suggests that absorption and poetry-elicited empathy should impact positively on the aesthetic evaluation of a poem (Garrido and Schubert, 2011; Taruffi and Koelsch, 2014).

Furthermore, curiosity is a key component of emotional motivation (Hoffman, 2006; Silvia, 2005). The recipient should be motivated to comprehend the cognitively demanding content of the poem, which is a determinant of NFC (i.e., an individual's tendency to engage in, and enjoy, effortful cognitive endeavors; Cacioppo and Petty, 1982). In general, we conclude that poetry reception favors pro-creativity states only under certain conditions, and that these conditions should be investigated in future studies.

GENERAL DISCUSSION

Poets describe their emotions and observations, in the form of metaphorical statements, in an effort to better convey their vision of the world to the reader. In two studies, which were conducted using a test/re-test design, we controlled for the impact of two different types of poems, from two renowned artists, to determine what, if any, impact the reception of poetry has on idea generation. Szyborska's narration is intellectually intriguing, with a surprising conclusion. Conversely, Gustafson's narration is a poetic description of the music of a master. The first poet uses open metaphors, while the second conventional ones. We expected, and confirmed to a large extent, that perceiving novel metaphors, based on remote associations (i.e., open metaphors) would result in more creative responses to a problem, whereas reception of well-known metaphors, which reinforce the world view shared by the community (i.e., closed metaphors) would lead to less creative ideas. Even one-time contact with narrative, open poetry improved some aspects of DT. However, we did not observe changes in originality, which is the key indicator of DT efficiency. We attributed this effect to the author's reasoning, aimed at one, surprising punch line.

Despite limitations in the selection of material, we conclude that poetry could be a useful tool for manipulating DT. Specifically, the results of the current studies suggest that poetry improves creativity if it contains open metaphors. However, reading conventional poetry may actually decrease idea generation. It is likely that the selection of poetic and control texts will remain an open problem for future studies on this topic.

We also accounted in these studies for individual differences that are critical for poetry reception. Frequent contact with poetry is associated with a slightly higher level of DT (compared to a lack of involvement in poetry) and could be explained by higher need

for cognition (curiosity) and ability to empathize with poetry content.

Limitations and Future Directions

Although many of our hypotheses about the varied impact of poetry on generating ideas have been confirmed, it became clear that the simple division of metaphors into novel/open and well-known is not enough of a manipulation to affect DT. The narrative structure of the poem introduced limitations to the free and original interpretation of even the most distant, metaphorical associations. Therefore, future studies will seek pro-creative poetry in less structured and more emotional forms of poetic expression, specifically with the development of emotional themes that increase uncertainty and stimulate the reader's imagination (Kozielecki, 2007).

While we showed that the impact of poetry reading on creative thinking depends on the type of poetry, future studies should manipulate the type of poetry utilized in a single study. Specifically, there are more types of poetry (aside from non-conventional and conventional) that could impact the reader in diverse ways that we did not explore. According to Heiden (2014), a fictionalized, narrative text can either address one's understanding of life and a specific challenge found within the individual's personal story (reference to "I"), or be an interpretation of events in the form of a story in general (referenced as "life at large"). Poetry that focuses on feelings, and disregards coherent narration, can be referred to as "cathartic poetry" (omitted in this research). The aim of cathartic poems is not to bring meaning closer, but rather to evoke the reader's emotions. This type of poetry is an open task for readers, because everybody can comprehend it according to his or her own experience and understanding. It can support creativity more than narrative poetry used in the Study 1. Thus, it would be desirable to use narrative, cathartic, and conventional poems in one experimental model.

The current studies showed no increase in originality following poetry exposure. Therefore, it is important to conduct future studies to determine what kind of poetry, as well as what kind of cognitive abilities are necessary to achieve an increase in originality, which is the primary metric in DT.

It is also possible that the effects we observed could be due to the specific poems chosen, rather than the content relating to metaphor styles. This issue can be addressed only by choosing several wide-ranging poems, which differ in terms of both metaphorization style and structure. In addition to the well-structured poetry that we used in the current studies, we will choose poems in future research that are emotional and uncertain.

It is important to note that the control texts used in both of our experiments were not rated by the same judges who rated the poems in terms of affectivity and comprehensibility. Thus, we did not control the same possible factors that were neutralized by selecting and rating poems. Future studies should seek to ensure that all pieces used (both poetry and control) are rated. Additionally, the description of the poem's content that was used as control text in the second study expresses a similar meaning to the poem, but without the use of metaphors. Without

rating the content of both texts (poetry and its description), however, we cannot infer their similarity. To address this, a diverse range of texts included in the final collection should be rated by judges in the same manner as poems, both for affectivity and comprehensibility. In this way, the collection would result in several poems, restricted to the best examples of the three different metaphor styles (i.e., narrative, conventional, and cathartic). Further, the personality determinants of poetry receiving in judges and the receivers should be also be controlled.

In the current studies, creativity was more related to general problem solving than production of creative works (e.g., poetry, fictional stories). In future studies, we intend to check the influence of specific types of poetry reading on creating one's own poems or prose samples. Future research should also explore the underlying mechanism behind how poetry influences creativity. Considering factors like emotions that are a consequence of contact with a poem, as well as individual differences in NFC and empathy, would allow us to construct a model to better describe the impact of poetry on the human mind. Furthermore, we failed to target specific audiences with specific types of poetry, which future studies should attempt. Finally, since the sample comprised high school students it would be difficult to extrapolate the results to a wider population.

DATA AVAILABILITY STATEMENT

Datasets are available upon request. The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

REFERENCES

- Amabile, T. M. (1996). *Creativity in Context*. Boulder, CO: Westview.
- Andonovska-Trajkovska, D. (2008). Creative writing in primary school and creativity as a concept. *Presented at VI Autumn Scientific Conference*, Palo Alto, CA.
- Ashton-James, C. E., and Chartrand, T. L. (2009). Social cues for creativity: the impact of behavioral mimicry on convergent and divergent thinking. *J. Exp. Soc. Psychol.* 45, 1036–1040. doi: 10.1016/j.jesp.2009.04.030
- Ayman-Nolley, S. (2010). A Piagetian perspective on the dialectic process of creativity. *Creat. Res. J.* 12, 267–275. doi: 10.1207/s15326934crj1204-4
- Baer, J. (1996). The effects of task-specific divergent-thinking training. *J. Creat. Behav.* 30, 183–187. doi: 10.1002/j.2162-6057.1996.tb00767.x
- Beaty, R. E., and Silvia, P. J. (2013). Metaphorically speaking: cognitive abilities and the production of figurative speech. *Mem. Cogn.* 41, 255–267. doi: 10.3758/s13421-012-0258-5
- Benedek, M., Beaty, R., Jauk, E., Koschutnig, K., Fink, A., Silvia, P. J., et al. (2014). Creating metaphors: the neural basis of figurative language production. *Neuroimage* 90, 99–106. doi: 10.1016/j.neuroimage.2013.12.046
- Benedek, M., Könen, T., and Neubauer, A. C. (2012). Associative abilities underlying creativity. *Psychol. Aesthet.* 6, 273–281. doi: 10.1037/a0027059
- Briggs, K. C., and Myers, I. B. (1976). *Myers-Briggs Type Indicator: Form F*. Palo Alto, CA: Consulting Psychologists Press.
- Cacioppo, J. T., and Petty, R. E. (1982). The need for cognition. *J. Pers. Soc. Psychol.* 42, 116–131. doi: 10.1037/0022-3514.42.1.116
- Cheng, Y. Y., Wang, W. C. H., Liu, K. S., and Chen, Y. L. (2010). Effects of association instruction on fourth graders' poetry creativity

ETHICS STATEMENT

The study was reviewed and approved by the Ethics Council of SWPS University of Social Sciences and Humanities, Faculty in Sopot, Poland. Written informed consent was obtained from all participants and from the parents of all minors.

AUTHOR CONTRIBUTIONS

MO and AK equally contributed to the study concept and design. Additionally, MO collected the data, developed the line of argumentation, performed the data analyses, and developed a poetry classification. MO and AK approved the final version of the manuscript for submission.

FUNDING

The preparation of this paper was supported by a grant from National Science Centre, Poland No. 2016/21/N/HS6/2868 awarded to MO.

ACKNOWLEDGMENTS

We would like to thank students Katarzyna Rajska, Oskar Wójcik, Katarzyna Gałasińska-Grygorczuk, and Angelika Krause for their help with data collection and creativity rating. We also thank Radosław Sterczynski for his help designing procedures. We would like to thank Editage (www.editage.com) for English language editing.

- in Taiwan. *Creat. Res. J.* 22, 228–235. doi: 10.1080/10400419.2010.481542
- Chiappe, D. L., and Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *J. Mem. Lang.* 56, 172–188. doi: 10.1016/j.jml.2006.11.006
- Chybicka, A. (2001). *Sytuacyjne wyznaczniki twórczości*. Unpublished doctoral dissertation. Gdańsk: Uniwersytet Gdański.
- Corbalan, J., and Lopez, J. J. (1992). Question generation test: a new procedure for measuring creativity. *J. Creat. Behav.* 26, 3–18.
- Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. New York, NY: Harper Collins.
- Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126. doi: 10.1037/0022-3514.44.1.113
- Djikic, M., Oatley, K., and Peterson, J. (2006). The bitter-sweet labor of emoting: linguistic comparison of writers and scientists. *Creat. Res. J.* 18, 195–201. doi: 10.1207/s15326934crj1802-5
- Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *J. Pers. Soc. Psychol.* 71, 390–405. doi: 10.1037/0022-3514.71.2.390
- Finke, R. A., Ward, T. B., and Smith, S. M. (1992). *Creative Cognition: Theory, Research and Applications*. Cambridge, MA: MIT Press.
- Garrido, S., and Schubert, E. (2011). Individual differences in the enjoyment of negative emotion in music: a literature review and experiment. *Music Percept.* 28, 279–296. doi: 10.1525/mp.2011.28.3.279
- Gibbs, R. W. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge: Cambridge University Press.

- Gilhooly, K. J., Fioratou, E., Anthony, S. H., and Wynn, V. (2007). Divergent thinking: strategies and executive involvement in generating novel uses for familiar objects. *Br. J. Psychol.* 98, 611–625.
- Glucksberg, S. (2001). *Understanding Figurative Language: From Metaphors to Idioms*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195111095.001.0001
- Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends Cogn. Sci.* 7, 92–96. doi: 10.1016/S1364-6613(02)00040-2
- Glucksberg, S., Gildea, P., and Bookin, H. B. (1982). On understanding nonliteral speech: can people ignore metaphors? *J. Verbal Learn. Verbal Behav.* 21, 85–98. doi: 10.1016/S0022-5371(82)90467-4
- Guilford, J. P. (1950). Creativity. *Am. Psychol.* 4, 444–454. doi: 10.1037/h0063487
- Gustafsson, L. (2013). *Dziwne Drobne Przedmioty. Wybór Wierszy*. Kraków: Wydawnictwo Znak.
- Halonen, J. S. (1995). Demystifying critical thinking. *Teach. Psychol.* 22, 75–81.
- Hausman, C. (1989). *Metaphor and Art*. New York, NY: Cambridge University Press.
- Heiden, B. (2014). Narrative in poetry: a problem of narrative theory. *Narrative* 22, 269–283. doi: 10.1353/nar.2014.0015
- Hoffman, M. (2006). *Empatia i Rozwój Moralny*. Gdansk: GWP.
- Jakobson, R. (1960). “Linguistik und Poetik in Jakobson,” in *Style in Language*, ed. T. Sebeok (New York, NY: Wiley), 350–377.
- Kaufman, J. C., and Beghetto, R. A. (2009). Beyond big and little: the four C model of creativity. *Rev. Gen. Psychol.* 13, 1–12. doi: 10.1037/a0013688
- Kenett, Y. N., Gold, R., and Faust, M. (2018). Metaphor comprehension in low and high creative individuals. *Front. Psychol.* 9:482. doi: 10.3389/fpsyg.2018.00482
- Kirsch, I. S., and Guthrie, J. T. (1984). Prose comprehension and text search as a function of reading volume. *Read. Res. Q.* 19, 331–342. doi: 10.2307/747824
- Koestler, A. (1964). *The Act of Creation*. New York, NY: Penguin Books.
- Kolańczyk, A. (1991). *Intuicyjność Procesów Przetwarzania Informacji. [Intuitiveness of information processing]*. Gdansk: Wydawnictwo Uniwersytetu Gdańskiego.
- Koveses, Z. (2010). *Metaphor: A Practical Introduction*. Oxford: Oxford University Press.
- Kozielecki, J. (2007). *Psychotransgresjonizm. Nowy Kierunek Psychologii. [Psychotransgressionism. A New Current in Psychological Thought]*. Warszawa: Wydawnictwo Akademickie Żak.
- Kraxenberger, M., and Menninghaus, W. (2016). Affinity for poetry and aesthetic appreciation of joyful and sad poems. *Front. Psychol.* 7:2051. doi: 10.3389/fpsyg.2016.02051
- Lakoff, G., and Johnson, M. (2003). *Metaphors We Live by*. Chicago: University of Chicago Press. doi: 10.7208/chicago/9780226470993.001.0001
- Lakoff, G., and Turner, M. (1989). *More Than Cool Reason: A Field Guide to Poetic Metaphor*. Chicago: University of Chicago Press.
- Lerdahl, F. (2001). *Tonal Pitch Space*. New York, NY: Oxford University Press.
- Martindale, C. (1989). “Personality, situation, and creativity,” in *Handbook of Creativity: Perspectives on Individual Differences*, ed. R. T. Brown (Boston, MA: Springer), 211–232.
- McGovern, T. V., and Hogshead, D. L. (1990). Learning about writing, thinking about teaching. *Teach. Psychol.* 17, 5–10. doi: 10.1207/s15328023top1701_1
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* 69, 220–232. doi: 10.1037/h0048850
- Nęcka, E. (2012). *Psychologia Twórczości. [Psychology of Creativity]*. Gdansk: GWP.
- Nęcka, E., and Kubiak, M. (1989). The influence of training in metaphorical thinking on creativity and level of dogmatism. *Polish Psychol. Bull.* 20, 69–80.
- Nickerson, R. S. (1999). “Enhancing creativity,” in *Handbook of Creativity*, ed. R. J. Sternberg (New York, NY: Cambridge University Press), 392–430.
- Noy, P., and Noy-Sharav, D. (2013). Art and Emotions. *Int. J. Appl. Psychoanal. Stud.* 10, 100–107. doi: 10.1002/aps.1352
- Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., et al. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Front. Psychol.* 4:10. doi: 10.3389/fpsyg.2013.00010
- Paivio, A. (1979). “Psychological processes in the comprehension of metaphor,” in *Metaphor and Thought*, ed. A. Ortony (New York, NY: Cambridge University Press), 150–171.
- Plotnik, A. (2007). *Spunk and Bite: A Writer's Guide to Bold, Contemporary Style*. New York, NY: Random House Reference.
- Ritter, S. M., Damian, R., Simonton, D., van Baaren, R. B., Strick, M., Derks, J., et al. (2012). Diversifying experiences enhance cognitive flexibility. *J. Exp. Soc. Psychol.* 48, 961–964. doi: 10.1016/j.jesp.2012.02.009
- Runco, M. (1999). Metaphors and creative thinking. *Creat. Res. J.* 4, 85–86. doi: 10.1080/10400419109534376
- Shiloh, S., Salto, E., and Sharabi, D. (2002). Individual differences in rational and intuitive thinking styles as predictors of heuristic responses and framing effects. *Pers. Individ. Differ.* 32, 415–429. doi: 10.1016/S0191-8869(01)00034-4
- Silvia, P. J. (2005). Cognitive appraisals and interest in visual art: exploring an appraisal theory of aesthetic emotions. *Empir. Stud. Arts* 23, 119–133. doi: 10.2190/12AV-AH2P-MCEH-289E
- Silvia, P. J., and Beaty, R. E. (2012). Making creative metaphors: the importance of fluid intelligence for creative thought. *Intelligence* 40, 343–351. doi: 10.1016/j.intell.2012.02.005
- Sternberg, R. J., and Lubart, T. I. (1999). “The concept of creativity: prospects and paradigms,” in *Handbook of Creativity*, ed. R. J. Sternberg (New York, NY: Cambridge University Press), 3–15.
- Stirman, S., and Pennebaker, J. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosom. Med.* 63, 517–522. doi: 10.1097/00006842-200107000-00001
- Szymborska, W. (2012). *Wiersze Wybrane*. Krakow: Wydawnictwo.
- Taruffi, L., and Koelsch, S. (2014). The paradox of music-evoked sadness: an online survey. *PLoS One* 9:e110490. doi: 10.1371/journal.pone.0110490
- Trzebiński, J. (1981). *Twórczość a Struktura Pojęć*. Warszawa: PWN.
- Turner, F., and Pöppel, E. (1983). The neural lyre: poetic meter, the brain, and time. *Poetry* 142, 277–307.
- Wachowicz, E. (2014). *Sklep Internetowy*. Available at: <http://speedcook.pl/urzadzenie> [accessed April 20, 2018].

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 PC and handling Editor declared their shared affiliation at time of review.

Copyright © 2018 Osowiecka and Kolańczyk. 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.



What Are the Stages of the Creative Process? What Visual Art Students Are Saying.

Marion Botella*, Franck Zenasni and Todd Lubart

Laboratoire Adaptations Travail-Individu, Université Paris Descartes, Paris, France

OPEN ACCESS

Edited by:

Kathryn Friedlander,
University of Buckingham,
United Kingdom

Reviewed by:

Gareth Dylan Smith,
New York University, United States
Elena Alessandri,
Lucerne University of Applied
Sciences and Arts, Switzerland

*Correspondence:

Marion Botella
marion.botella@parisdescartes.fr

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 25 April 2018

Accepted: 31 October 2018

Published: 21 November 2018

Citation:

Botella M, Zenasni F and Lubart T
(2018) What Are the Stages of the
Creative Process? What Visual Art
Students Are Saying.
Front. Psychol. 9:2266.
doi: 10.3389/fpsyg.2018.02266

Within the literature on creativity in the arts, some authors have focused on the description of the *artistic* process (Patrick, 1937; Getzels and Csikszentmihalyi, 1976; Mace and Ward, 2002; Yokochi and Okada, 2005) whereas others have focused on the *creative* process (Wallas, 1926; Osborn, 1953/1963; Runco and Dow, 1999; Howard et al., 2008). These two types of processes may be, however, somewhat distinct from each other because the creative process is not always dedicated to artistic creation, and productive work in the arts may not always involve creativity, in terms of specifically original thinking. Our goal is to identify the specific nature of the *artistic creative* process, to determine what are the basic stages of this kind of process. This description can then be integrated in a Creative process Report Diary (CRD; Botella et al., 2017) which allows self-observations *in situ* when participants are creating.

Keywords: creative process, stages, visual art students, interviews, Creative process Report Diary

FROM THE EXISTING CREATIVE AND ARTISTIC PROCESSES TO THE ARTISTIC CREATIVE PROCESS

The creative process is defined as a succession of thoughts and actions leading to original and appropriate productions (Lubart, 2001; Lubart et al., 2015). The creative process may be described at two levels: a *macro* level, featuring the stages of the creative process, and a *micro* level, which explains the mechanisms underlying the creative process, e.g., divergent thinking or convergent thinking (Botella et al., 2016). Although the works carried out on micro-processes tend to agree on a set of mechanisms that can be involved in the creative process, work focusing on macroprocesses have not achieved consensus regarding the nature or the number of stages involved in the creative process. **Table 1** shows some of the different models that can be found in the scientific literature, with overlaps or divisions between some stages of the models. In this paper, we treat micro-processes as contents of a more global, macro-level process, which make it possible to describe the construction of a work of art from the beginning (i.e., the wish to create) to the end (exhibiting that work). Moreover, the process can be examined in a psychological and individual or in a socio-cultural perspective (Glăveanu, 2010; Burnard, 2012). In the present study situated in the visual art field, we will consider the artistic creative process as an individual phenomenon.

Art is often considered to be an archetypal domain of creativity research (Schlewitt-Haynes et al., 2002; Stanko-Kaczmarek, 2012), complimented by research on scientific, musical, design-oriented, and literary creativity (Glăveanu et al., 2013). Even if some overlap can be observed between different creative fields, each field has its own specificities (Botella and Lubart, 2015). The purpose of this section is to merge some existing models of the creative process and artistic process to examine what the artistic creative process could be. Obviously, this section cannot be exhaustive but offers a first consideration of the numerous important stages of the artistic creative process.

The process starts by an *orientation*, in which the individual identifies the problem that must be solved (Osborn, 1953/1963), called also a stage of *problem selection* (Busse and Mansfield, 1980) or a *sensitivity to problems* (Guilford, 1956). *Problem definition* involves producing as many questions as possible. For Runco and Dow (1999), *problem-finding* refers to a process of “sensing gaps” (Torrance, 1962)—that is, detecting elements that are lacking. In the same vein, Bruford (2015) proposed a stage of *differentiation* consisting of retaining information that leads to producing something different, involving interpretative and expressive musical differences. Additionally, Mumford et al. (1994) suggested making a distinction between *discovering a problem* (i.e., rejecting problems that are untrue, incorrect, or incomplete; Getzels and Csikszentmihalyi, 1976; Arlin, 1986), *posing the problem* (i.e., finding a correct formulation), and *constructing a problem* (i.e., describing the problem). In the artistic field, Fürst et al. (2012) proposed a model of art production that includes a *goal of creation*.

Then, there is *preparation*, the first stage described in the early macroprocess model by Wallas (1926). Carson (1999) explained that, in this stage, the individual defines the problem (or understands it; Treffinger, 1995) and gathers information in order to solve it. Based on a series of interviews with novelists, Doyle (1998) argued that the creative process begins with an *incident*, when an individual discovers an idea. In the artistic process literature, Mace and Ward (2002) proposed a four-stage model based on interviews with professional artists. For them, the artistic process begins with the *design* of an artistic work. Hence, work is initiated by a more-or-less vague idea or impression. Recently, based also on a series of interviews with professional artists, Botella et al. (2013) identified six stages in the creative process in art, starting by an *idea* or a “*vision*” in which an image, a sight, a sound resonates with the artist.

Before the second main stage described by Wallas (1926), some authors added complementary stages after preparation. Based on a previous review of the literature, Botella et al. (2011) propose a stage of *concentration* (“I am concentrating on the work I have to do”) in which it is possible to focus the creator’s attention on those solutions deemed to be adequate, and to reject the other solutions (Carson, 1999). Osborn (1953/1963) added *analysis*, when the creator takes a step back to identify the relations between ideas and the importance of each idea; and *ideation*, when the individual develops alternative ideas. Busse and Mansfield (1980) indicated also a stage requiring making an *effort* in order to solve the problem.

Then, according to Wallas (1926) and many other authors, *incubation* occurs (Osborn, 1953/1963; Shaw, 1989, 1994; Runco, 1997; Runco and Dow, 1999; Botella et al., 2011). This is a time of solitude and relaxation, where idea associations take place at a subconscious level (Carson, 1999). Recently, Sadler-Smith (2016) reintegrated a fifth stage in the Wallas’ model: *intimation* occurs between incubation and insight. Intimation is described as an “association-train” in a fringe conscious level, between conscious and unconscious levels (p. 346). Cropley and Cropley (2012) revisited as well Wallas’s work and split the stage of incubation into *activation* and *generation*. The process once again becomes conscious in the stage of

ideation, with the generation of further ideas, which are not necessarily judged or assessed. The individual then experiences an *illumination* or *insight* (Eureka!) with the emergence of an idea, an image or a solution (Wallas, 1926; Carson, 1999). Boden (2004) noted that illumination or insight needs previous thought-processes.

Idea generation can take place in various ways according to the different models. Busse and Mansfield (1980) described a stage in which the creator sets the constraints related to the solution of the problem and, then, another stage involving the transformation of these constraints or adaptation of the constraints that are not suitable. For Doyle (1998), there is some form of *navigation* between various knowledge domains, which makes it possible to assess the relevance of this idea. Based on Dewey (1934), Bruford (2015) proposed a *selection* stage in which the creator chooses one option among several, requiring agency and control abilities. In the field of art, Mace and Ward (2002) named this step *idea development* in which the artist structures, completes, and restructures the idea. Botella et al. (2013), through interviews with professional artists identified a stage of *documentation* and *reflection* during which artists gather more information about the materials and technologies required in order to turn their vision into reality. The last stage described by Wallas (1926) is *verification* (Busse and Mansfield, 1980). New ideas are tested and *verified*, leading to the elaboration of a solution and to its production (Carson, 1999). More precisely, Osborn (1953/1963) proposed two distinct phases of *synthesis*, which consists of gathering ideas together and distinguishing relations between them.

Gruber (1989) argued that the four-stage model is incomplete. For Russ (1993), there lacks a stage of application, or deployment of the creative production. Treffinger (1995) added effectively a stage of *idea production*, leading to action by *planning*. This work corresponds to the development and implementation of ideas through a search for solutions (evaluation, selection, and redefinition), and then the acceptance of this solution (promoting an idea, looking for its strengths and drawbacks). This last stage makes it possible to materialize the ideas that have been found and to solve the problem. In this vein, in the field of art, Mace and Ward (2002) described the *realization* of an idea, during which the artist transforms that idea into a physical entity. Botella et al. (2011) also added stages of *planning* (“I am planning my work”), and *production* (“I am producing/composing my ideas”). Results of observations in the art field suggested that the production stage is comprised, in fact, of two stages: a stage that consists of searching for ideas through the creative gesture (sketches, drafts, mock-ups), and then a stage consisting of the realization of an idea that is already constructed (transposing an idea to a concrete medium). The initial stage of “production” describes a similar action, but the underlying cognitive micro-processes are different. In the first case, the goal is to produce in order to formulate an idea whereas in the second case, it is to produce in order to implement an idea that already exists. In a study consisting of interviews of professional artists, Botella et al. (2013) confirmed the stages of first *sketches* to give a material form to the initial project, *testing* the forms and ideas that originated from reflection and preliminary

TABLE 1 | Synthesis of some examples of models of creative process.

Author(s)	Stages					
	Preparation	Incubation	Insight	Incubation	Insight	Insight
Wallas, 1926						
Guilford, 1956	Sensitivity to problems					
Osborn, 1953/1963	Preparation	Incubation	Ideation	Transformation	Synthesis	Evaluation
Busse and Mansfield, 1980	Orientation Selection	Analysis				
Shaw, 1989, 1994	Immersion	Constraints	Efforts	Verification	Creative synthesis	Validation
Mumford et al., 1994	Problem discovery			Incubation	Insight	Explanation
Treffinger, 1995	Problem definition Understanding		Problem construction			
Amabile, 1988, 1996	Preparation	Idea production	Planning			
Runco, 1997	Problem presentation	Answer generation	Answer validation	Outcome		
Doyle, 1998	Incident	Incubation	Insight	Verification	Communication	Validation
Carson, 1999	Information					
Runco and Dow, 1999	Preparation	Incubation	Insight	Verification	Elaboration	Production
Mace and Ward, 2002	Conception	Incubation	Development		Realization	Evaluation
Basadur and Gelade, 2005					Finalization	
Kilgour, 2006	Definition	Combination	Generation	Conceptualization	Optimization	Implementation
Howard et al., 2008	Analysis		Idea generation			
Botella et al., 2011	Preparation	Incubation	Generation	Communication/ Implementation		
Cropley and Cropley, 2012	Preparation	Incubation	Ideation	Verification	Production	Validation
Botella et al., 2013	Idea or "vision"	Activation	Insight	Insight	Verification	Verification
Sadler-Smith, 2016	Documentation and reflection	First sketches	Testing	Provisional objects		Series
	Preparation	Incubation	Intimation	Insight	Verification	

work, and *provisional objects*, “drafts” and almost-finished products. Revisiting Wallas’ model, Cropley and Cropley (2012) mentioned a stage of *communication*, as Bruford (2015) with musicians.

For Osborn (1953/1963), the last stage is *evaluation* (Runco and Dow, 1999; or *assessment* for Bruford, 2015), in which the individual assesses the chosen idea. For Mace and Ward (2002), the final step of the artistic process, called *finalization*, brings the artistic work to conclusion (or *validation* according to Botella et al., 2011; Cropley and Cropley, 2012). The artist reassesses the production and may choose to finish, to elaborate, abandon, delay, store, or destroy it. If the artist believes the mission that was set has been accomplished, the artist may choose to exhibit the production. Recently, professional artists suggested to add one more stage with *series*, transforming a first object to many objects (Botella et al., 2013).

All these models were developed based on rational or empirical approaches. Original works and models from Poincaré and Wallas’ were conceived based, respectively, on their own experience and pragmatic empirical observations. Patrick (1935, 1937) supported Wallas proposal by collecting empirical data in terms of observations and verbal reports of poets and artists who were invited to do a specific creative task. Most of the “stage models” are then based on this kind of rational or empirical analyses, with verbalizations, specifications, and clarifications of the processes by the participants themselves in the majority of cases. Therefore, these models maybe be considered as a specific approach to creativity, distinct from the psychometric, problem finding or cognitive experimental approaches (Kozbelt et al., 2010). Recent studies on the four-stages model of Wallas confirmed again that researchers do not agree on the number of stages: Cropley and Cropley (2012) found seven stages whereas Sadler-Smith (2016) found five stages based on Wallas’ book.

OBJECTIVES

Models of the creative process and of the artistic process do not agree on the nature or on the number of steps involved in a creative artistic process (see Howard et al., 2008). This lack of a consensus could be explained by the fact that (a) the creative process is a complex phenomenon as described by Osborn (1953/1963) who believed that creation is set off by “*stop-and-go*” or “*grab what you can*”-type processes; (b) models of a creative process are constructed based on a specific creative population and a specific creative domain, though these are described as if they were generic and could apply to all domains whether art, science, music, writing, or design. The process is most often described in general terms, as if it should apply to all creative domains, whether it is art, science, music, writing, or design; (c) descriptions of the artistic process do not always take into account the definition of creativity, in particular the contextually rich, situated nature that originality, and appropriateness may have; and (d) the methodologies used were different [be it a review of the literature (Busse and Mansfield, 1980; Botella et al., 2011), a series of interviews with novelists (Doyle, 1998), with

professional artists (Mace and Ward, 2002; Botella et al., 2013), or an applied and consulting-based approach (Carson, 1999)].

The aim of the present study is to question directly some stakeholders of artistic creativity, namely visual art students. However, it is maybe too ambitious to ask them to describe completely their creative process. We suggest that the lack of consensus in the previous studies could be due to the desire to capture all aspects of the creative process in the same study. So, the students interviewed here describe only what constitutes, for them, the stages of their process of artistic creativity. We ask them specifically to list the stages of their process in order to be as exhaustive as possible. This qualitative study makes it possible to identify what stages the students consider relevant in their mental representation of the visual artistic creative process, rather than relying on stages extracted from the scientific literature on creativity. With this study, we will not be able to have a macro vision of the entire artistic creative process but we will construct an inventory of the stages involved to picture this process.

Given the descriptive nature of the present research on the artistic creative process, the findings can be integrated in further work as a part of the Creative process Report Diary (CRD, Botella et al., 2017). The CRD is a useful and relevant analytical tool to assess the creative process in a natural context, when it occurs, allowing ecological validity. It is possible to realize various versions of the CRD depending on the context, the creative field, and any other considerations. The CRD has two parts: a part listing the stages of the creative process (which will be as exhaustive as possible based on the present study) and a part listing factors such as cognitive, conative, emotional, and environmental ones that may come into the creative process (for example, we could assess team work; Peilloux and Botella, 2016). Finally, the CRD allows the creative process to be modeled for individuals *in situ* during all the time needed for their creation. Thus, the purpose of CRD will be to observe the link and the transitions between the stages of the artistic creative process and to examine which factors will be involved at each stage. However, to do that, we need, in the present study, to list as exhaustively as possible all the stages of the visual artistic creative process which will allow a specific CRD to be created to observe the process in further study.

METHODS

Participants

The sample was composed of 28 students in the second year of a visual graphic arts school. Seventeen students were female and 11 were male (*mean age* = 20.9 years old, *sd* = 1.7, *span* = 19–24 years old). The rationale for the choice of this sample was to interview participants with some artistic experience but to avoid a sample habituated to interviews with strongly formatted ideas. In previous research, when we interviewed professional artists (Botella et al., 2013), we noticed some routines in the discourse. Some artists were familiar with interviews and they narrated a story, usually the story of an artwork but sometimes the reports were distanced from their own story and therefore from their own creative process.

Interview Guide

The goal of the study was to construct a list of the stages of the process of visual artistic creativity. Given this, the interview guide was purposely kept short and open, and consisted of only two questions: (1) “how does your creative process generally take place?” and (2) “how would you name the stages that you have just mentioned?”

The interviewer’s follow-up questions allowed the students to describe another stage of their creative process. The main prompts consisted of reformulating the last sentence provided by the participant and asking “When you did [...], what do you do next?” or “Can you describe more precisely what you do when you finish [...]?” It was very important to not induce ideas with our questions so, we just reformulated the words used by the visual art students themselves to help them list the stages of their artistic creative process.

Interviews were semi-structured and lasted 10 min on average. Obviously, the interviews were too short to capture all the complexity of the artistic creative process with its “stop-and-go” or “grab what you can” aspects (Osborn, 1953/1963). However, to make an inventory of the stages it was enough. The added value of this study is to focus the interview on the stages that visual art students themselves considered and how they named them.

Procedure

Ethics approval was not required according to our institution’s guidelines and national regulations. After the participants provided informed consent, the volunteer students were interviewed in their art school, during their course on creativity. This situation made it easier for them to recall the stages of their visual artistic creative process. Participants were led to a separate room to take part in a one-on-one discussion with the interviewer. The interviewer (and then, the analyst) was the first author, with knowledge on the literature about creativity and creative process, who had already realized many interviews mainly with artists (Botella et al., 2013; Glaveanu et al., 2013). The prompts consisted of reformulating what participants said to assure that we did not induce the use of certain terms.

RESULTS

Given our objective was to inventory the stages of the artistic creative process, we analyzed the words employed during the interviews. The terms used by students were grouped in equivalence sets using Tropes software which presents references cited at least three times. The name retained for the category was the most cited term; others citations were used to describe the category. In the first part of the analysis, we focus on the stages of the process of visual artistic creativity that emerged spontaneously from the participants’ discourse. Hence, we will deal with the responses to the first question in the interview guide. In the second part, we will examine the stages named by the students. Finally, we will confront these two analyses, in order to check whether the stages named by the participants do indeed correspond to those referenced in the discourse. It is expected that the names will be very similar for both analyses but this

confrontation serves to cross-check the categorized sets of terms and their labels.

Identifying the Stages of the Process From the Students’ Open Discourse

Based on the students’ responses to the first question in the interview guide, all the terms cited at least three times were listed. It should be noted that the software can already group some terms according to the context: for example, “impossible” and “not possible” are considered as similar. The software can also identify co-occurrences of combined terms, such as “applied art.” Then, terms were grouped by the analyst according to the context in which they appeared (see **Table 2**). The context helped us to identify the terms concerning the creative process. When terms seem to correspond to the same idea, they were grouped together, such as “Sketchpad,” “sketch,” “drawing,” and “writing.” We conducted an ascendant hierarchical classification, grouping two by two the closest words. The number of clusters was not decided in advance and the grouping was stopped when we considered that another aggregation was not relevant. Terms that did not refer to the creative process were not retained (“year,” “art,” “stage,” “have an inclination toward,” “social environment,” etc.).

In **Table 2**, the number of times that a category was cited and how many students referred to this category are indicated because the same student could mention the same category several times. One stage consists of *approaching* the subject matter, taking possession of it, gaining knowledge about the subject-related words used (S14: “So, you go there, you throw yourself”). *Reflection* refers to the students’ efforts for deciphering and understanding the topic. This stage may imply visualized images (S1: “I think, I get things straight for a week”). The stage of *research* involves the student going to the library in order to collect references to artists and to prior work (S4: “I am looking for references to see what has been done. There is a time of documentation”). Then the student constructs a knowledge base of works which have already been produced, before distancing themselves from these works. *Inspiration* is based on one’s impression and experience of a given subject matter (S24: “it’s really how I feel it and I know I’ll be able to continue on it”). Although the term *illumination* was not used, we can note the presence of this stage in students’ reports of “an idea suddenly appearing” or “coming across an idea by accident” (S6: “It’s not totally conscious. It comes like this. Ideas come alone. We feel it. And after that, we try from that to bring this idea in a frame that could be appropriate”). *Trials* correspond to producing notebooks containing sketches. Students record their sketches, and make attempts before they can find an idea (S27: “I try to explore as many things as possible”). *Organization* consists of students ordering, guiding, and organizing their approach by mixing existing ideas and combining them together (S25: “There is an order to be defined”). The student will have to *select* an idea out of all those produced (S25: “I will select what is best”). A work involves inevitably one or more *techniques* (S18: “Whether computer, photoshop or drawing, rush. Really, exploit everything I know as technical before you get to a final thingy”). Depending

TABLE 2 | Categories of references used in the students' discourse.

Category	No. times cited	Percentage of students (%)	Other references used
Approaching the subject matter	245	100	Apprehending, starting up, word, concept, project, topic, text, theme
Research	203	96.43	Artist, basis, library, searching, cinema, knowledge, culture, curiosity documentation, information, internet, book, research, researching, reference, providing information
Trials	198	92.86	Sketchpad, sketch, drawing, writing, trying, mock-up, note, taking notes
Finalization	119	85.17	Stopping, end, final, finalizing, adding, resuming, correcting, reworking, finishing
Specification	80	85.71	Improving, moving forward, changing, continuing, adding depth, detailing, developing, evolving, perfectionism, photoshop, pushing forward, specifying
Realization	98	75	Composition, make concrete, illustration, implementation, practice, production, clean, production, product, volume
Technique	85	71.43	Code, color, materials, matter, means, painting, photography, style, technique, line, typography, use
Inspiration	51	64.29	Impression, inspiration, inspire, feeling
Reflection	46	57.14	Understanding, deciphering, exploiting, reflecting, reflection, vision, visualization
Organization	38	46.43	Training, link, linking, logic, mixing, order, organizing, steering, connection, relationship
Illumination	37	46.43	Coming across an idea
Presentation	22	42.86	Showing, presentation, handing in
Selection	32	39.29	Choosing, choice, selection, selecting
Judgment	19	35.71	Opinion, looking at
Failure	20	28.57	Failure, mistake, throwing away, missing, beginning again, doing over

Categories represent the stage of the visual artistic creative process organized by the percent of citation. Number times cited indicates that a student can refer to a category more than one time. All the others references used to mention a stage are reported in this table.

on individual preferences and on the constraints of the situation, the student will choose to use a particular technique. The product of the creative process is made concrete during the stage of *realization* (S9: “I go directly to the realization with the materials. I take the painting and I do it directly to clean”). The stage of *specification* indicates that the student improves, specifies and adds the finishing touches to the work (S15: “I am improving what I have already drawn. Above all, I simplify. Because I tend to put too much”). *Finalization* refers to the stage in which the work is completed, finished, and voluntarily stopped (S28: “I am very meticulous and I spend a lot of time on the end”). The stage of *judgment* corresponds to assessing the work that has been produced (S27: “Generally, I have to finish in advance so I can look at it for a long time and then see if something is missing or not. Because sometimes, I have the impression that it is not finished at all and, by dint of looking at it, finally I realize that it misses nothing or that it misses things precisely”). The *presentation* is the moment when students present their work to their teachers (S20: “It’s when I show to the teachers”). The stage of *failure* indicates that the student has abandoned something, be it the work or an idea. In the latter case, the student throws away one idea and starts something new, or starts again based on an existing work (S3: “If it’s not good, I do not leave, I start again. It happens to me often when I’m done and it’s ugly, that I know it’s not good, I don’t care, I spend another 8 hours, 10 hours to rework another volume. In general, when I resume it’s still the same theme, but it’s not the same idea”).

Identifying the Stages of the Process Named by Students

This analysis focused on the second question in the interview guide, i.e., how the students named the stages in their visual artistic creative process. Terms were grouped in **Table 3**. From there, we were able to identify 16 stages in the process of visual artistic creativity.

Immersion refers to assimilating the work to be done; it involves listening to the instructions given by the teacher, defining the words in the topic, and entering into the project. *Reflection* relates to a form of brainstorming where the student attempts to understand, to decipher the topic and to reflect upon it. *Research* may focus on artists, documents, books, the Internet, and aims for the students to construct a knowledge base for themselves. *Inspiration* seems to be related to intuition and instinct. *Apparition* refers to ideas being found and appearing of their own accord. *Trials* designate all the try-outs, notes, sketches, notes, and testing made by the students. *Assembly* refers both to attempting a new approach and to the different ideas that emerge from assembling ideas together. The stage of *new ideas* includes different ideas which emerge. The stage of *selection* involves choosing an idea. *Materials* were also mentioned in terms of photography and volume. The stage of *realization* refers to action, composition, concretization, production, and to the transfer of an idea to a medium. The stage of *specification* can be viewed as increasing the depth of analysis, developing the work, and correcting it. *Finalization* is the completion of the work. The

TABLE 3 | Categories used in naming the stages of the creative process.

Category	No. times cited	Percentage of students (%)	Other references used
Research	58	78.57	Artist, searching, knowledge, document, information, internet, book, research, researching, reference
Trials	36	64.29	Try-outs, drafts, notebook, croquis, sketch, trial, trying, notes, testing
Realization	27	64.29	Act, action, application, composition, concretization, illustration, implementation, cleaning up, production, transcription, transposition
Selection	12	42.86	Choice, selection
Finalization	12	35.71	Ending, stopping, final, finalization, finalizing, finishing, finishing touches, corrections
Reflection	14	32.14	Brainstorming, understanding, deciphering, reflection, reflecting
Immersion	10	32.14	Assimilation, definition, defining, listening, immersion, impregnation, reception, entering
Specification	10	28.57	Detailing, detail, developing, development, specification
Assembly	9	28.57	Assembly, assembling, setting up ideas
Presentation	7	21.43	Explaining, justifying, speaking, presentation, hand-in
Materials	7	17.86	Photo, volume, materials
Examination	5	17.86	Taking a step back, examining, questioning
Settling	6	14.29	Decanting, digesting, letting time pass, taking a break
Teacher	6	14.29	Teacher
Inspiration	4	14.29	Inspiration, instinct, intuition
Apparition	6	10.71	Apparition, finding, coming
New idea	3	10.71	New, different ideas, differentiation

Categories represent the stage of the visual artistic creative process organized by the percent of citation. Number times cited indicates that a student can refer to a category more than one time. All the others references used to mention a stage are reported in this table.

stage of *examination* indicates taking a step back from the work, formulating an analysis of the work, and questioning one's own work. *Presentation* refers to the fact that students must justify, explain, and present their work. The fact that students let the work *settle*, digest and breathe may refer to the concepts of breaks and incubation. Finally, the *teacher* was also cited as a part of the stages of the process of artistic creativity when students ask for help because they are stuck or when they need reference.

Confronting the two Analyses and Identifying the Stages in the Process of Visual Artistic Creativity

This confrontation allowed us to verify that the students had indeed described all the stages in their creative process, thus validating the number and nature of steps involved in the process to integrate these in the CRD (see **Table 4**). Fourteen stages appear both in the free discourse and the stages named by the students, one stage was mentioned only in the discourse, and two stages were mentioned when naming the stages of the process. In the end, 17 different stages were retained. Only the stage referring to teacher was not retained because the teacher corresponds more to a social support than a stage of the process. Additionally, the teacher can be partially included in the stage of research as a source of knowledge.

In the stage of *immersion*, the goal is to apprehend the topic at hand and to listen to the instructions given by the teacher. Some students may sometimes feel the need to define the words and concepts present in the topic (S1: "What I do personally, I take the words and I take a few days or even a week depending on the time of the project to get things straight, think about it

because sometimes there are topics that are very vague like that and we understand not at all. And then it gets more and more precise."). Such an approach allows them to "soak up" the topic and jump into the fray and start themselves off (S18: "The thing is, I often tend to get into an idea. When you give me a subject or what. I guess right now the thing and what I could do with it."). *Reflection* makes it possible to understand what should be done, and to decipher the teacher's requirements. Mental work may sometimes begin with visualizing an image. This image may guide the student throughout the process (S20: "Me, I cannot start looking for a word if I do not visualize the final "what." Even if I will redo after..."). During the stage of *research* the students learn to search for artists, references, documents, and work already produced about the topic that they are apprehending. A solid knowledge base and a culture regarding prior work might help create new and original ideas (S15: "The teachers give us research. Because when we come here, we do not necessarily have a culture in terms of graphics, anyway. They give us references to go see. This is because, often, it is sometimes references of choreographers and it goes a little beyond the field of visual arts and graphics. And suddenly, it allows to compare universes. And then we improve what we do."). *Inspiration* occurs when an idea emerges slowly and gradually. According to the students, it is based on instinct, impressions, and feelings (S14: "Sometimes you feel that you have a lot of data and from that, you can start to grab something"). Although the word *illumination* was never mentioned, the literature places a strong emphasis on this stage. It is translated in the interviews as "apparition," "coming across an idea," and "hey, there's an idea!" where the idea sometimes comes from an unknown place (S5: "Sometimes it comes alone.");

TABLE 4 | Confrontation between the two analyses.

Analysis of the discourse	Analysis of the stages named	Stages retained
Apprehending the subject matter: Apprehending, starting up, word, concept, project, topic, text, theme Reflection: Understanding, deciphering, exploiting, reflecting, reflection, vision, visualization Research: Artist, basis, library, searching, cinema, knowledge, culture, curiosity documentation, information, internet, book, research, researching, reference, providing information Inspiration: Impression, inspiration, inspire, feeling Illumination: Coming across an idea, coming Trials: Sketchpad, sketch, drawing, writing, trying, mock-up, grading, grade Organization: Training, link, linking, logic, mixing, order, organizing, steering, connection, relationship Selection: Choosing, choice, selection, selecting Technique: Code, color, materials, matter, means, painting, photography, style, technique, line, typography, use Realization: Composition, make concrete, illustration, implementation, practice, production, clean, product, volume Specification: Improving, moving forward, changing, continuing, adding depth, detailing, developing, evolving, perfectionism, photoshop, pushing forward, specifying Finalization: Stopping, end, final, finalizing, adding, resuming, correcting, reworking, finishing Judgment: Opinion, looking at Presentation: Showing, presentation, handing in Failure: Failure, mistake, throwing away, missing, beginning again, doing over	Immersion: Assimilation, definition, defining, listening, immersion, impregnation, reception, entering Reflection: Brainstorming, understanding, deciphering, reflection, reflecting Research: Artist, searching, knowledge, document, information, internet, book, research, researching, reference Inspiration: Inspiration, instinct, intuition Apparition: Apparition, finding, coming Trials: Try-outs, drafts, notebook, croquis, sketch, trial, trying, notes, testing Assembly: Assembly, assembling, setting up ideas New idea: New, different ideas, differentiation Selection: Choice, selection Materials: Photography, volume, materials Realization: Act, action, application, composition, concretization, illustration, implementation, cleaning up, production, transcription, transposition Specification: Detailing, detail, developing, development, specification Finalization: Ending, stopping, final, finalization, finalizing, finishing, finishing touches, corrections Examination: Taking a step back, examining, questioning Presentation: Explaining, justifying, speaking, presentation, hand-in Settling: Decanting, digesting, letting time pass, taking a break	Immersion: Impregnating oneself with the topic, defining the topic, listening to instructions Reflection: Brainstorming, understanding, visualizing Research: Searching for references, documents, artists, information Inspiration: An idea gradually formed, instinct Illumination: “Aha, there’s an idea!,” emergence of an idea Trials: Sketches, note-taking, testing, trying out Assembly: Mixing or assembling ideas together Ideation: Finding and exploring new ideas Selection: Choosing ideas Technique: Choosing a technique, a style, a typography Realization: Production, composition, concretization, cleaning up Specification: Refining and detailing Finalization: Adding finishing touches, correcting, ending Examination: Taking a step back and examining one’s production Presentation: Explaining and justifying one’s production Break: Letting everything rest, digesting, decanting Abandoning: Putting aside or abandoning an idea

S21: “I did not look. It fell on me in fact. And so after, you have to bounce back.”). The use of notebooks gather the students’ *trials*, their sketches and their notes. They allow the students to try out and test an image. More importantly, the teachers examine the notebooks to follow the evolution of the students’ work. Notebooks show students’ train of thought, how they achieved a particular work (S2: “These ideas, I always put them in my notebook to show them to the teacher.”). *Assemblies* of ideas are the result of logical connections that the student establishes between several existing ideas. Thus, it corresponds to the direction which the student wishes to give to the production and future work (S3: “I try to mix everything together”). The stage of *ideation* was not mentioned in the discourse. It was only mentioned when students were naming the stages. *Selection* refers to classifying and sorting ideas. The goal here is to choose which ideas can be exploited, and which, on the other hand, should be set aside (S24: “It’s hard to choose, on which track to

go”). *Technique* is a very important aspect for aspiring artists. They must comply with codes, rules, find a typography, a style of their own. Although this stage was rarely named as such by the students, it is very present in their discourse (S27: “I put in some technique. For example, I had been taught a little about the technique of collage, I had exploited this thing after because I liked it. I tried to distort it from school in my own way.”). *Realization* refers to translating an idea into an image. It is at this point that the composition and production of a material work take shape (S18: “I try to realize it at best”). The stage of *specification* reveals the improvements, the added details, the changes, and corrections made to the work underway. At this point, students add details that they had not necessarily planned initially (S23: “When I have something that I like, I dig it even more to see if I can exploit it”). *Finalization* refers to the point at which the student decides that the work is done. The work is complete, or almost at the point of completion (S17: “It’s never

finished. For renderings, there is a fixed date and there it is finished. But just for a grade. But in general, we always have stuff to add, photos to resume, stuff to put back. Generally, we do it if we have a jury at the end of the year. And here, we try to finalize the project of the beginning of the year.”). The term *judgment* was not explicitly mentioned either. However, it can be found in the terms of taking a step back, questioning one’s work, observing it with great attention, and thus assessing it (S3: “I look at [my work]. I think instead of teachers. If I was a teacher, if I look at, if there is something wrong, if there is a stain, if I see that there is something wrong, if it is not good, well cut, I’ll start all over again.”). Although this stage was not directly mentioned in the students’ discourse, the stage of the *break* also seems to exist. Its goal is to let the ideas rest, digest, settle and “breathe.” The discourse suggests also the presence of trial and error. Because the word “failure” seems a little strong, we retain the term of “*abandoning*,” whose connotations are less negative (S3: “Sometimes I change my idea and sometimes, when I work, it’s not possible like that”).

DISCUSSION

The goals of this study were to determine the nature and number of stages present in the creative visual artistic process in order to build a specific CRD. Twenty-eight art students were asked to describe their process of visual artistic creativity and to name its stages. By comparing the discourse of these art students and the names they gave to the various stages of their work, we identified 17 stages.

Immersion is present in several existing models. It corresponds to preparation in Wallas’ (1926) model (see **Table 5** for a synthesis). Wallas views preparation as a preliminary analysis which makes it possible to define and set the problem. The same idea is present in Carson’s (1999) consulting-centric model and in the work on the creative process of actors (Blunt, 1966; Nemiro, 1997, 1999). Osborn (1953/1963) speaks instead of orientation, in which the individual identifies the problem that is to be solved. Shaw (1989, 1994) proposes also the term “immersion.” *Reflection* is typically included in preparation. Osborn proposes a stage when the individual takes a step back to examine the connections that exist between different ideas. More recently, this stage of reflection was identified in interviews with professional artists (Botella et al., 2013). The stage of *research* is required by the school of art (S8: “We have a lot of instructions from the teachers who help us. We must go through research.”). Research is also generally included in preparation. It should be noted that in Treffinger’s model (Treffinger, 1995), preparation is called understanding. The goal here is for the individual to search for information regarding the problem at hand. Also, Runco (1997) mentions a stage of information. Here, the research stage could help visual art students to differentiate their own work from previous ones (Bruford, 2015). In the interviews with professional artists (Botella et al., 2013), this search stage was coupled with reflection, as a search for means (i.e., material or technological) to transform the initial idea into a real production.

Inspiration corresponds to intuition and metacognition (Cropley, 1999). Amongst other things, it allows us to identify which approach will be more efficient than another. Policastro

TABLE 5 | Correspondence between the stages retained in the present study and the existing stages in research field.

Stages retained in the present study	Correspondence with existing stages
Immersion	<i>Preparation</i> (Wallas, 1926; Osborn, 1953/1963; Blunt, 1966; Amabile, 1988; Nemiro, 1997, 1999; Carson, 1999; Botella et al., 2011) <i>Sensitivity to problems</i> (Guilford, 1956) <i>Orientation</i> (Osborn, 1953/1963) <i>Immersion</i> (Shaw, 1989, 1994) <i>Problem discovery</i> (Mumford et al., 1994) <i>Problem presentation</i> (Amabile, 1988) <i>Problem finding</i> (Runco and Dow, 1999) <i>Conception</i> (Mace and Ward, 2002) <i>Definition</i> (Kilgour, 2006)
Reflection	<i>Preparation</i> (Wallas, 1926; Osborn, 1953/1963; Blunt, 1966; Amabile, 1988; Nemiro, 1997, 1999; Carson, 1999; Botella et al., 2011) <i>Analysis</i> (Osborn, 1953/1963; Howard et al., 2008) <i>Efforts</i> (Busse and Mansfield, 1980) <i>Problem definition</i> (Mumford et al., 1994) <i>Conception</i> (Mace and Ward, 2002) <i>Generation</i> (Basadur and Gelade, 2005; Howard et al., 2008) <i>Goal of creation</i> (Fürst et al., 2012) <i>Documentation and reflection</i> (Botella et al., 2013)
Research	<i>Preparation</i> (Wallas, 1926; Osborn, 1953/1963; Blunt, 1966; Amabile, 1988; Nemiro, 1997, 1999; Carson, 1999; Botella et al., 2011) <i>Efforts</i> (Busse and Mansfield, 1980) <i>Problem definition</i> (Mumford et al., 1994) <i>Understanding</i> (Treffinger, 1995) <i>Information</i> (Runco, 1997) <i>Conception</i> (Mace and Ward, 2002) <i>Documentation and reflection</i> (Botella et al., 2013) <i>Differentiation</i> (Bruford, 2015)
Inspiration	<i>Intuition and metacognition</i> (Cropley, 1999) <i>Idea or vision</i> (Botella et al., 2013) <i>Intimation</i> (Sadler-Smith, 2016)
Illumination	<i>Insight</i> (Wallas, 1926; Ghiselin, 1952; Gruber and Davis, 1988; Weisberg, 1988; Shaw, 1989, 1994; Runco, 1997; Carson, 1999; Botella et al., 2011) <i>Idea production</i> (Treffinger, 1995) <i>Answer generation</i> (Amabile, 1988) <i>Incident</i> (Doyle, 1998) <i>Conceptualization</i> (Basadur and Gelade, 2005)
Trials	<i>Transformation</i> (Busse and Mansfield, 1980) <i>Navigation</i> (Doyle, 1998) <i>Idea development</i> (Mace and Ward, 2002) <i>First sketches and testing</i> (Botella et al., 2013)
Assembly	<i>Divergent thinking</i> (Guilford, 1950) <i>Combination</i> (Kilgour, 2006)
Ideation	<i>Ideation</i> (Osborn, 1953/1963; Carson, 1999; Botella et al., 2011) <i>Idea generation</i> (Kilgour, 2006) <i>Generation</i> (Cropley and Cropley, 2012)
Selection	<i>Selection</i> (Busse and Mansfield, 1980; Bruford, 2015) <i>Concentration</i> (Carson, 1999; Botella et al., 2011)
Technique	<i>Constraints</i> (Busse and Mansfield, 1980)
Specification	<i>Elaboration</i> (Berger et al., 1957; Carson, 1999) <i>Explanation</i> (Shaw, 1989, 1994) <i>Planning</i> (Treffinger, 1995; Botella et al., 2011) <i>Optimization</i> (Basadur and Gelade, 2005) <i>Documentation</i> (Botella et al., 2013)

(Continued)

TABLE 5 | Continued

Stages retained in the present study	Correspondence with existing stages
Realization	<i>Synthesis</i> (Osborn, 1953/1963; Shaw, 1989, 1994) <i>Problem construction</i> (Mumford et al., 1994) <i>Production</i> (Treffinger, 1995; Carson, 1999; Botella et al., 2011) <i>Realization</i> (Mace and Ward, 2002) <i>Implementation</i> (Basadur and Gelade, 2005; Howard et al., 2008) <i>Provisional objects</i> (Botella et al., 2013)
Finalization	<i>Finition</i> (Mace and Ward, 2002)
Judgement	<i>Verification</i> (Wallas, 1926; Busse and Mansfield, 1980; Armbruster, 1989; Runco, 1997; Carson, 1999; Botella et al., 2011) <i>Evaluation</i> (Osborn, 1953/1963; Runco and Dow, 1999; Howard et al., 2008) <i>Validation</i> (Amabile, 1988; Shaw, 1989, 1994; Runco, 1997; Botella et al., 2011; Cropley and Cropley, 2012) <i>Assessment</i> (Bruford, 2015)
Presentation	<i>Outcome</i> (Amabile, 1988) <i>Communication</i> (Runco, 1997; Howard et al., 2008; Cropley and Cropley, 2012)
Break	<i>Incubation</i> (Wallas, 1926; Patrick, 1937; Osborn, 1953/1963; Dreistadt, 1969; Shaw, 1989, 1994; Smith and Blankenship, 1989, 1991; Smith and Vela, 1991; Russ, 1993; Runco, 1997; Carson, 1999; Runco and Dow, 1999; Botella et al., 2011)
Withdrawal	<i>Abandoning</i> (Mace and Ward, 2002)

(1995) defines intuition as an implicit form of information processing, which is intended to anticipate and guide creative research. According to her, intuition may allow an unconscious shift from incubation to illumination. However, intuition was never considered a stage in the creative process or in the artistic process. Therefore, it is a stage that is specific to the current study. As described by the students, the inspiration stage is close to the stage on intimation added between incubation and insight (Sadler-Smith, 2016). It is surprising and interesting that visual art students consider inspiration as a stage of their creative process. So, a replication of this study will be necessary to confirm if it is really a stage or if it is a factor involved in the creative process. The word “illumination” was not mentioned by the students as such. Numerous authors have previously shown that the illumination stage was seldom mentioned by students in art. Doyle (1998) has described illumination as an accident, where the solution emerges in a sudden and unexpected way (Wallas, 1926). Hence, the description that the students made of this stage might be termed illumination: the idea comes or appears in an unexpected manner. Other authors believe that this experience of illumination would, in most cases, be more gradual than sudden (Ghiselin, 1952; Gruber and Davis, 1988; Weisberg, 1988). Although it is possible that illumination is not a part of all creative processes, or that the creators might not always be conscious of it, the stage of illumination remains a key stage in the creative process, because it is at this stage that the idea takes shape.

The *trials*, tests, and fiddling made by students may correspond to the stage of idea development in Mace and Ward's model (Mace and Ward, 2002). In their description of the artistic process, Mace and Ward argue that, during the development of an idea, the artist will structure, complete, and restructure the idea. Authors indicate that this trial stage will allow artists to form a more precise idea of the initial project for themselves. This stage is worked in Art school with sketchpads.

Assembly corresponds to the microprocess of divergent thinking, in which ideas are assembled and mixed together. In contrast, convergent thinking makes it possible to focus on a single idea (Guilford, 1950). This mode of thinking allows individuals to find the one and only solution to a problem. The generation of ideas that have not yet been checked and assessed corresponds to *ideation* (Carson, 1999). Osborn (1953/1963) mentions a stage of synthesis, which consists of putting ideas together and distinguishing relations between them.

Selection refers to concentration (Carson, 1999). Concentration makes it possible to focus the attention of the individual on those solutions deemed to be adequate, and to reject other solutions. No model emphasizes the stage of choosing a *technique*. Yet, the artist must identify the technique that will allow them to make the idea materialize in the best possible way. During the interviews with professional artists, technical issues were included in the stage of documentation (Botella et al., 2013). However, in the present study, because 71.43% of the students mentioned this stage in their discourse and 17.86% named it directly, we decided to consider “technique” as a specific stage of the visual artistic creative process. In further studies, it will be interesting to explore if this stage is specific to visual arts or if it is a more common stage concerning other creative domains.

Specification might correspond to elaboration. Berger et al. (1957) defined elaboration as the individual's ability to provide detail to the ideas produced. This stage may also tie in with creative explanation, whose goal is for the artist to explain the ideas (Shaw, 1989, 1994).

Realization refers to the creative production (Treffinger, 1995) or to creative synthesis (Shaw, 1989, 1994). The goal here is to make the idea concrete. “Technique” is generally included in this stage. However, it seems that production points to the act of creating and to the gestures involved rather than to the cognitive or emotional choice of a technique. Mace and Ward (2002) speak also of realization, i.e., the transformation of an idea into a “physical entity.” They note that for some physical arts and for a wide variety of artistic media it is necessary to have a detailed idea of what the artist is going to do. Hence, some decisions—such as, for example, those related to the choice of a technique—should be anticipated.

Finalization corresponds, at least in part, to the finition phase in Mace and Ward (2002). The authors argue that finalization implies that the individual has decided that his/her work is finished. If the artist considers the work to be successful and satisfactory and they may choose to exhibit it. In that case, the stage of finalization also includes hanging up or exhibiting the work.

The stage of *judgement* of the creative production is very often named in models of the creative process. In particular, Wallas (1926) writes about verification, where the individual assesses the idea that has emerged. At this stage, one must take a step back from one's work and assess it. Verification may be of two kinds: "internal" verification, i.e., a comparison between the idea that has been produced and the idea formed during illumination or "internal" verification, which consists of anticipating the reactions of the audience (Armbruster, 1989). According to Busse and Mansfield (1980), verification may take place earlier during the process, as the individual first verifies the ideas and then elaborates a solution. Other authors have argued that judgment occurs at a later stage. For example, Osborn (1953/1963) considers that evaluation is the moment when the individual evaluates the chosen idea. When describing the creative process, Osborn (1953/1963) mentions the stage of analysis, in which the individual takes a step back to examine the connections that form between ideas and their importance. In contrast, Shaw (1989, 1994) addresses the concept of validation, thus emphasizing the importance of this stage. According to him, personal validation consists of appreciating one's own work and in using the experience acquired over the course of this process to generate a new creative process. In addition to personal validation, there exists a collective level of validation. The latter deals with the evaluation of a creative production by peers, by an audience or by a critic. Collective validation can only lead to a new process if there is acceptance of the evaluation that has been formulated. If the production is validated, it can then be followed by a series in which the idea is extended to several works (Botella et al., 2013).

The stage of *presentation* is not typically described as such in models of the creative process or of the artistic process; its goal is to present the work to teachers. In the case of professional artists, this would refer more to the sale of a work. However, recent models included a communication stage (Runco, 1997; Howard et al., 2008; Cropley and Cropley, 2012).

The term "*break*" which has emerged in the stages named by students might correspond to incubation. As we have seen, this stage is very difficult to assess and to take into account (Botella et al., 2011), even though it is essential (Patrick, 1937; Dreistadt, 1969; Smith and Blankenship, 1989, 1991; Smith and Vela, 1991), especially to the expression of artistic creativity (Russ, 1993). The words used by the students highlight some unconscious associations. Indeed, they talk about letting their ideas rest, letting them digest and decant. Incubation is always difficult to evaluate, because it relies in most cases on unconscious work. Finally, although the stage of *withdrawal* is a subject of research, it is not included in most models of the creative process. Only Mace and Ward (2002) take into account a clear possibility of abandoning the process at any time. Even if the process is brutally interrupted, the artist develops continuously new knowledge. This knowledge is the result of a perpetual, dynamic interaction with artistic practice. Artists extend and refine their repertoire of skills, techniques, and knowledge. Also they sharpen their artistic interests and personality. New ideas can emerge in this work, to be reused later.

CONCLUSION

Although this study was limited by the interview method—and thus focused on students' implicit theories of their own creative process—it allowed us to identify multiple stages in the process of visual artistic creativity. Because of the implicit theories and the number of models suggesting a linear sequence of stages, sometimes with some loops or cycles possible, it seems too ambitious to understand the sequence of the stages from interviews. The present study invites us to rethink what composes an artistic creative process. Even if we already have a long list of models, none is complete and satisfactory. It is possible that we may need to construct and maintain a list of all the stages of the creative process which can then be adapted to each domain, given that the creative process may vary depending upon the area in question (Baer, 1998, 2010; Botella and Lubart, 2015). Given this uncertainty, continued research into the creative process is indicated. For now, the present list of stages of the visual artistic creative process could help teachers in their coursework. During the interviews, students indicated that the stages of research and the use of the diary notebook were required by their art school. This appears as a limitation of the present study. We are not sure if art students described the prescriptive stages in their Art school or their real stages of creation. The question was oriented how their creative process generally takes place but because they are art students and they were interviewed in their art school, some prescriptive stages appears in their discourse. However, during the interviews, some students had specified if the stage is prescriptive and we indicated this point throughout this paper. With the updated list, teachers could propose other exercises to guide art students for all the stages. Moreover, outside an educational context, the demand for consultancy to stimulate business creativity is increasing (see Berman and Korsten, 2010), and the current research may also provide a helpful template for the effective management of creative processes in this area of industrial innovation. However, we have to be careful about the use of such a list. By conceptualizing the creative process, are we actually at risk of creating a "uniform" prescriptive model of how to be creative? We can hypothesize that some creative process are more adapted to some creative individuals but it would be counterproductive to try to force all individuals to engage in the same process. The creative process varies across fields (Botella and Lubart, 2015) and probably also across culture, creators' personalities, and tasks.

These stages and more precisely their sequence should be validated in the field, by observing students as they carry out artistic work—notably to determine the exact succession of the stages—using a tool like the CRD. Moreover, it will be interesting to observe the collaborative creative process as well as to situate the process in a more global socio-cultural approach. As we saw in the introduction, the creative process can be described using micro-level or macro-level approaches and more globally takes place in a particular socio-cultural context. These approaches could be used directly during observations of the creative process and associated with cognitive, conative, emotional, and environmental factors involved in the process.

ETHICS STATEMENT

All subjects gave written informed consent in accordance with the Declaration of Helsinki.

REFERENCES

- Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Res. Organ. Behav.* 10, 123–167.
- Amabile, T. M. (1996). *Creativity in Context*. Boulder, CO: Westview Press.
- Arlin, P. K. (1986). “Problem finding and young adult cognition,” in *Adult Cognitive Development: Methods and Models*, eds R. A. Mines and K. S. Kitchener (New York, NY: Praeger), 22–32.
- Armbruster, B. B. (1989). “Metacognition in creativity,” in *Handbook of Creativity*, eds E. P. Torrance, J. A. Glover, R. R. Ronning and C. R. Reynolds (New York, NY: Plenum), 177–182. doi: 10.1007/978-1-4757-5356-1_10
- Baer, J. (1998). The case for domain specificity of creativity. *Creat. Res. J.* 11, 173–177. doi: 10.1207/s15326934crj1102_7
- Baer, J. (2010). “Is creativity domain specific?,” in *The Cambridge Handbook of Creativity*, eds J. C. Kaufman and R. J. Sternberg (Cambridge: Cambridge University Press), 321–341. doi: 10.1017/CBO9780511763205.021
- Basadur, M., and Gelade, G. A. (2005). Modelling applied creativity as a cognitive process: theoretical foundations. *Kor. J. Think. Prob. Solving* 15, 13–41.
- Berger, R. M., Guilford, J. P., and Christensen, P. R. (1957). A factor-analytic study of planning abilities. *Psychol. Monogr.* 71, 1–29. doi: 10.1037/h0093704
- Berman, S., and Korsten, P. (2010). *Capitalizing on Complexity: Insights from the Global Chief Executive Officer Study*. Somers, NY: IBM.
- Blunt, J. (1966). *The Composite Art of Acting*. New York, NY: Macmillan.
- Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms*. Abingdon: Routledge. doi: 10.4324/9780203508527
- Botella, M., Glaveanu, V. P., Zenasni, F., Storme, M., Myszkowski, N., Wolff, M., et al. (2013). How artists create: creative process and multivariate factors. *Learn. Individ. Diff.* 26, 161–170. doi: 10.1016/j.lindif.2013.02.008
- Botella, M., and Lubart, T. (2015). “Creative processes: art, design and science,” in *Multidisciplinary Contributions to the Science of Creative Thinking*, eds G. E. Corazza and S. Agnoli (Singapore: Springer), 53–65.
- Botella, M., Nelson, J., and Zenasni, F. (2016). “Les macro et micro processus créatifs [The macro and micro creative processes],” in *Créativité et apprentissage [Creativity and learning]*, ed I. Capron-Puozzo (Louvain-la-Neuve: De Boeck), 33–46.
- Botella, M., Nelson, J., and Zenasni, F. (2017). It is time to observe the creative process: how to use a creative process Report Diary (CRD). *J. Creat. Behav.* doi: 10.1002/jocb.172
- Botella, M., Zenasni, F., and Lubart, T. I. (2011). A dynamic and ecological approach to the artistic creative process in arts students: an empirical contribution. *Empir. Stud. Arts* 29, 17–38. doi: 10.2190/EM.29.1.b
- Bruford, W. (2015). Making it Work: Creative Music Performance and the Western Kit Drummer. Unpublished dissertation, University of Surrey.
- Burnard, P. (2012). *Musical Creativities in Practice*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199583942.001.0001
- Busse, T. V., and Mansfield, R. S. (1980). Theories of the creative process: a review and a perspective. *J. Creat. Behav.* 14, 91–132. doi: 10.1002/j.2162-6057.1980.tb00232.x
- Carson, D. K. (1999). “Counseling,” in *Encyclopedia of Creativity*, Vol. 1, eds M. A. Runco and S. R. Pritzker (New York, NY: Academic Press), 395–402.
- Cropley, A. J. (1999). “Definitions of creativity,” in *Encyclopedia of Creativity*, Vol. 1, eds M. A. Runco and S. R. Pritzker (New York, NY: Academic Press), 511–524.
- Cropley, D. H., and Cropley, A. J. (2012). A psychological taxonomy of organizational innovation: resolving the paradoxes. *Creat. Res. J.* 24, 29–40. doi: 10.1080/10400419.2012.649234
- Dewey, J. (1934). *Art as Experience*. New York, NY: Penguin Books.
- Doyle, C. L. (1998). The writer tells: the creative process in the writing of literacy fiction. *Creat. Res. J.* 11, 29–37. doi: 10.1207/s15326934crj1101_4
- Dreistadt, R. (1969). The use of analogies and incubation in obtaining insights in creative problem solving. *J. Psychol.* 71, 159–175. doi: 10.1080/00223980.1969.10543082
- Fürst, G., Ghisletta, P., and Lubart, T. (2012). The creative process in visual art: a longitudinal multivariate study. *Creat. Res. J.* 24, 283–295. doi: 10.1080/10400419.2012.729999
- Getzels, J. W., and Csikszentmihalyi, M. (1976). “Concern for discovery in the creative process,” in *The Creativity Question*, eds A. Rothenberg and C. R. Hausman (Durham, NC: Duke University Press), 161–165.
- Ghiselin, B. (1952). *The Creative Process*. Berkeley, CA: University of California Press.
- Glăveanu, V. P. (2010). Paradigms in the study of creativity: introducing the perspective of cultural psychology. *New Ideas Psychol.* 28, 79–93. doi: 10.1016/j.newideapsych.2009.07.007
- Glaveanu, V. P., Lubart, T., Bonnardel, N., Botella, M., De Biasi, P.-M., De Sainte Catherine, M., et al. (2013). Creativity as action: findings from five creative domains. *Front. Educ. Psychol.* 4:176. doi: 10.3389/fpsyg.2013.00176
- Gruber, H. E. (1989). “The evolving systems approach to creative work,” in *Creative People at Work: Twelve Cognitive Case Studies*, eds D. B. Wallace and H. E. Gruber (New York, NY: Oxford University Press), 3–24.
- Gruber, H. E., and Davis, S. N. (1988). “Inching our way up Mount Olympus: the evolving systems approach to creative thinking,” in *The Nature of Creativity*, ed R. J. Sternberg (New York, NY: Cambridge University Press), 243–270.
- Guilford, J. P. (1950). Creativity. *Am. Psychol.* 5, 444–454. doi: 10.1037/h0063487
- Guilford, J. P. (1956). Structure of intellect. *Psychol. Bull.* 53, 267–293. doi: 10.1037/h0040755
- Howard, T. J., Culley, S. J., and Dekoninck, E. (2008). Describing the creative design process by the integration of engineering design and cognitive psychology literature. *Des. Stud.* 29, 160–180. doi: 10.1016/j.destud.2008.01.001
- Kilgour, M. (2006). Improving the creative process: analysis of the effects of divergent thinking techniques and domain specific knowledge on creativity. *Int. J. Business Soc.* 7, 79–107.
- Kozbelt, A., Beghetto, R. A., and Runco, M. A. (2010). “Theories of creativity,” in *The Cambridge Handbook of Creativity*, Vol. 2, eds J. C. Kaufman and R. J. Sternberg (New York, NY: Cambridge University Press), 20–47.
- Lubart, T. I. (2001). Models of the creative process: past, present and future. *Creat. Res. J.* 13, 295–308.
- Lubart, T. I., Mouchiroud, C., Tordjman, S., and Zenasni, F. (2015). *Psychologie de la Créativité (Deuxième Edition) [Psychology of Creativity]*. Paris: Armand Colin.
- Mace, M.-A., and Ward, T. (2002). Modeling the creative process: a grounded theory analysis of creativity in the domain of art making. *Creat. Res. J.* 14, 179–192. doi: 10.1207/S15326934CRJ1402_5
- Mumford, M. D., Reiter-Palmon, R., and Redmond, M. R. (1994). “Problem construction and cognition: applying problem representations in ill-defined domains,” in *Problem Finding, Problem Solving, and Creativity*, ed M. A. Runco (Norwood, NJ: Ablex), 3–39.
- Nemiro, J. (1997). Interpretative artists: a qualitative exploration of creative process of actors. *Creat. Res. J.* 10, 229–239.
- Nemiro, J. (1999). “Acting,” in *Encyclopaedia of Creativity*, Vol. 1, eds M. A. Runco and S. R. Pritzker (New York, NY: Academic Press), 1–8.
- Osborn, A. F. (1953/1963). *Applied Imagination, 3rd Edn.* New York, NY: Scribners.
- Patrick, C. (1935). Creative thought in poets. *Arch. Psychol.* 178, 1–74.
- Patrick, C. (1937). Creative thought in artists. *J. Psychol.* 4, 35–73. doi: 10.1080/00223980.1937.9917525
- Peilloux, A., and Botella, M. (2016). Ecological and dynamical study of the creative process and affects of scientific students working in groups. *Creat. Res. J.* 28, 165–170. doi: 10.1080/10400419.2016.1162549

AUTHOR CONTRIBUTIONS

MB methodology, interviews, analyses, and writing; FZ methodology and writing; and TL methodology and writing.

- Policastro, E. (1995). Creative intuition: an integrative review. *Creat. Res. J.* 8, 99–113. doi: 10.1207/s15326934crj0802_1
- Runco, M. A. (1997). *The Creativity Research Handbook*. Cresskill, NJ: Hampton Press.
- Runco, M. A., and Dow, G. (1999). "Problem finding," in *Encyclopedia of Creativity*, Vol. 2, eds M. A. Runco and S. R. Pritzker (New York, NY: Academic Press), 433–435.
- Russ, S. W. (1993). *Affect and Creativity: The Role of Affect and Play in the Creative Process*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Sadler-Smith, E. (2016). Wallas' four-stage model of the creative process : more than meets the eye ? *Creat. Res. J.* 27, 342–352. doi: 10.1080/10400419.2015.1087277
- Schlewitt-Haynes, L. D., Earthman, M. S., and Burns, B. (2002). Seeing the world differently: an analysis of descriptions of visual experiences provided by visual artists and nonartists. *Creat. Res. J.* 14, 361–372. doi: 10.1207/S15326934CRJ1434_7
- Shaw, M. P. (1989). The eureka process: a structure for the creative experience in science and engineering. *Creat. Res. J.* 2, 286–298. doi: 10.1080/10400418909534325
- Shaw, M. P. (1994). "Affective components of scientific creativity," in *Creativity and Affect*, eds M. P. Shaw and M. A. Runco (Westport: Ablex Publishing), 3–43.
- Smith, S. M., and Blankenship, S. E. (1989). Incubation effects. *Bull. Psychon. Soc.* 27, 311–314. doi: 10.3758/BF03334612
- Smith, S. M., and Blankenship, S. E. (1991). Incubation and the persistence of fixation in problem solving. *Am. J. Psychol.* 104, 61–87. doi: 10.2307/1422851
- Smith, S. M., and Vela, E. (1991). Incubated reminiscence effects. *Mem. Cogn.* 19, 168–176. doi: 10.3758/BF03197114
- Stanko-Kaczmarek, M. (2012). The effect of intrinsic motivation on the affect and evaluation of the creative process among fine arts students. *Creat. Res. J.* 24, 304–310. doi: 10.1080/10400419.2012.730003
- Torrance, E. P. (1962). *Guiding Creative Talent*. Englewood Cliffs, NJ: Prentice-Hall. doi: 10.1037/13134-000
- Treffinger, D. J. (1995). Creative problem solving: overview and educational implications. *Educ. Psychol. Rev.* 7, 301–312. doi: 10.1007/BF02213375
- Wallas, G. (1926). *The Art of Thought*. New York, NY: Harcourt, Brace and Company.
- Weisberg, R. W. (1988). "Problem solving and creativity," in *The Nature of Creativity: Contemporary Psychological Perspectives*, ed R. J. Sternberg (Cambridge: Cambridge University Press), 148–176.
- Yokochi, S., and Okada, T. (2005). Creative cognitive process of art making: a field study of a traditional Chinese ink painter. *Creat. Res. J.* 17, 241–255.

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 Botella, Zenasni and Lubart. 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.



Conceptualising and Understanding Artistic Creativity in the Dementias: Interdisciplinary Approaches to Research and Practise

Paul M. Camic^{1,2*}, Sebastian J. Crutch^{1,3}, Charlie Murphy¹, Nicholas C. Firth^{1,4}, Emma Harding^{1,3}, Charles R. Harrison¹, Susannah Howard^{1,5}, Sarah Strohmaier^{1,2}, Janneke Van Leewen^{1,3}, Julian West^{1,6}, Gill Windle^{1,7}, Selina Wray^{1,8} and Hannah Zeilig^{1,9} on behalf of the Created Out of Mind Team†

¹ Created Out of Mind, Wellcome Collection, London, United Kingdom, ² Salomons Centre for Applied Psychology, Canterbury Christ Church University, Canterbury, United Kingdom, ³ Dementia Research Centre, UCL Institute of Neurology, University College London, London, United Kingdom, ⁴ Centre for Medical Image Computing, Department of Computer Science, University College London, London, United Kingdom, ⁵ Living Words, Folkestone, United Kingdom, ⁶ Royal Academy of Music, London, United Kingdom, ⁷ Dementia Services Development Centre, Bangor University, Bangor, United Kingdom, ⁸ UCL Institute of Neurology, University College London, London, United Kingdom, ⁹ College of Fashion, University of the Arts London, London, United Kingdom

OPEN ACCESS

Edited by:

Philip A. Fine,
University of Buckingham,
United Kingdom

Reviewed by:

Boris Forthmann,
Universität Münster, Germany
Genevieve Cseh,
Buckinghamshire New University,
United Kingdom

*Correspondence:

Paul M. Camic
paul.camic@canterbury.ac.uk

†<http://www.createdoutofmind.org>

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 26 March 2018

Accepted: 10 September 2018

Published: 03 October 2018

Citation:

Camic PM, Crutch SJ, Murphy C, Firth NC, Harding E, Harrison CR, Howard S, Strohmaier S, Van Leewen J, West J, Windle G, Wray S and Zeilig H (2018) Conceptualising and Understanding Artistic Creativity in the Dementias: Interdisciplinary Approaches to Research and Practise. *Front. Psychol.* 9:1842. doi: 10.3389/fpsyg.2018.01842

Creativity research has a substantial history in psychology and related disciplines; one component of this research tradition has specifically examined artistic creativity. Creativity theories have tended to concentrate, however, on creativity as an individual phenomenon that results in a novel production, and on cognitive aspects of creativity, often limiting its applicability to people with cognitive impairments, including those with a dementia. Despite growing indications that creativity is important for the wellbeing of people living with dementias, it is less well understood how creativity might be conceptualised, measured and recognised in this population, and how this understanding could influence research and practise. This paper begins by exploring prevailing concepts of creativity and assesses their relevance to dementia, followed by a critique of creativity and dementia research related to the arts. Perspectives from researchers, artists, formal and informal caregivers and those with a dementia are addressed. We then introduce several novel psychological and physiological approaches to better understand artistic-related creativity in this population and conclude with a conceptualisation of artistic creativity in the dementias to help guide future research and practise.

Keywords: dementia, creativity, dance, visual art, music, poetry, psychophysiology

THE DEMENTIAS AND CREATIVITY

The terms “creativity” and “dementias” are not two words that often find themselves linked. When asked about what word comes to mind when thinking about the dementias it is rare if not unheard of for creativity to be identified (Brotherhood et al., 2017; van Leeuwen et al., 2017a). Part of this disconnect is the result of years of creativity research that has focused on eminent creators in science and industry, university undergraduate psychology students as part of a course requirement, artists of various sorts, and gifted “geniuses” with very little research exploring

creativity and people with mental or physical health problems, the exception being the apparently “mad” artist (e.g., Csikszentmihalyi, 1997a; Chad et al., 2007; King Humphry, 2010; Bellas et al., 2018). The development of the “mini and little c” creativity models (Kaufman and Beghetto, 2009), among other recent advances which we will address allows for a more extensive exploration of creativity across different physical and mental health conditions. This paper examines the concept of artistic creativity and the dementias with an aim to encourage researchers, practitioners and policy makers to generate more research, enact arts and health policies and develop arts and dementia care programmes to help shape dementia care internationally.

Brief Overview of the Dementias

Recognition of the dementias (*pl.*) and their earliest impacts has been slowed by traditional definitions of dementia which emphasise impairment of memory and criteria which require cognitive impairment sufficient to compromise social and occupational functioning (American Psychiatric Association, 2000). Many diseases can result in a progressive dementia syndrome. The most common causes both in the elderly and in younger people are Alzheimer’s disease (AD), vascular disease, frontotemporal lobe degenerations (FTLD), and dementia with Lewy bodies (DLB). A number of dementias are associated with particular symptom profiles (e.g., DLB: hallucinations, cognitive fluctuations and Parkinsonian gait; semantic dementia: impaired language comprehension and semantic memory). However, heterogeneity in the dementias is increasingly acknowledged, with contemporary Alzheimer’s disease criteria describing not only the classical amnesic presentation, but also atypical presentations affecting visual perception, language or behaviour/executive functions (McKhann et al., 2011; Dubois et al., 2014). Atypical presentations and rarer dementias highlight the range of cognitive skills which may become vulnerable in anyone with a dementia as the condition progresses. Equally this heterogeneity serves to underline the relative preservation of certain skills and abilities well into a disease course when other aptitudes may be perceived to be profoundly compromised. It is against this complex, evolving cognitive background that different forms of individual and collective creativity in people with dementia must be considered.

Prevailing Concepts of Creativity and the Dementias

The idea of creativity is surprisingly recent. As Pope (2005) argues in his historical and critical guide to the concept the first recorded usage of creativity in English occurs only in 1875. Thus, the emergence of the concept coincided with the late Romantic period and was closely associated with the arts (Williams, 1988) and with the notion of the “artistic genius.” Even recent conceptualisations from both psychological and neurological perspectives tend to link creative processes to specific, original and tangible acts of production that are associated with individual motivations (e.g., Csikszentmihalyi, 1997a; Palmiero et al., 2012).

These are of relevance in that the myth of the “creative individual”, the “genius,” is a powerful motif shaping social understandings of creative activities (Runco, 1987). This hegemonic narrative not only informs shared ideas about age and creativity (McMullan and Smiles, 2016) but of central relevance for our discussion here, also influences the ways in which notions of creativity relate (or more pertinently do not relate) to people living with a dementia. Focusing on the characteristics and capacities of an individual defined as particularly creative, the narrative understands creativity as something psychologically inherent to a creative individual (Osborne, 2003). Recognising creativity and the production of creative acts as collective as well as individual (Becker, 2004) and also associated as much with process as product (Plucker and Beghetto, 2004), we explore the opportunities and constraints that are experienced by people living with a dementia in a variety of contexts and the ways in which these may extend our understandings of artistic creativity. The ways in which social practise (i.e., how individual and contexts codetermine each other) are situated or how central cognition seems to be in our understanding of creativity, are not fixed (Barb and Plucker, 2002, p. 169) but part of an ongoing debate about how to define creativity. Locating creativity primarily as a cognitive domain limits, however, the applicability of creativity as a construct in dementia research and care. As cognitive capacities decline and become less and less accessible it is important that researchers and clinicians do not assume that the potential for creative activity is eliminated.

The absence of a precise definition of a concept such as creativity can be problematic for research but arguably, it may also be that a universal definition of creativity and specifically, creativity and the arts, limits its applicability across people and environments and a more situated perspective is necessary (Clarke et al., 2018). For example, there are aspects of the definition offered by Plucker et al. (2004) that fit well across dementias (process, environment, and social context) but one aspect, aptitude, does not; the latter not necessarily being salient to everyday artistic creativity for this population. Whilst an in-depth review of the multiple prevailing definitions of creativity is beyond the scope of this article, four appear highly relevant to conceptualising the arts, creativity and dementias.

In an attempt to incorporate cross-cultural variations in Western and Eastern perspectives the four-criterion construct of creativity (Kharkhurin’s, 2014) uses the attributes of novelty, utility, aesthetics and authenticity to develop a matrix to compare creative products from “different areas of human endeavour across the arts, sciences and business” (p. 349). Two of the components resonate well with dementias. Utility refers to creative work perceived as such by the producer of the work and the recipient, producing a landmark in social or cultural environment and addressing moral issues. Secondly, authenticity, taken from Confucian aesthetics, is particularly noteworthy and reflects a process of bringing new responses into existing ideas to reflect an individual’s own essence at a moment in time (Tu, 1985). These components expand the concept of creativity to include the role of a socio-cultural context, individual perceptions

and responses from others that build on a more inclusive concept of creativity beyond cognitive factors.

Drawing on Rowlands (2010) ideas of an embodied, embedded and extended mind, Glăveanu (2013) sought to situate and contextualise creativity and developed the five A's framework which, he argues, represents "a fundamental change of epistemological position. In light of sociocultural sources, the actor (creator) exists only in relation to an audience, action cannot take place outside of interactions with a social and material world (affordances), and artefacts embody the cultural traditions of different communities" (p. 71). This framework is relevant to our discussion in that it outlines the inherently interrelated nature of the various aspects of creative endeavour. Above all, his framework places the creator (in our work the person with a dementia) in a broad context of material, social and cultural phenomena and relations. Glăveanu's framework represents a more fully systemic and situated theoretical model for understanding contextually how and when artistic creativity might take place across the spectrum of the dementias. For example, a person putting several words together poetically in an advanced stage of Alzheimer's disease as an expressive response to listening to music in the context of receiving residential care, could easily be minimised as a chance event. Yet, given that this person may not have spoken for months her poem might provide insights into her experience of living with dementia and can be understood as a creative response *at this point in her life*.

Glăveanu (2013) and Kharkhurin's (2014) contributions also blend well with the concepts of little and mini-c creativity, introduced by Beghetto and Kaufman (2007) and Richards (2007), respectively, which is the third perspective we draw upon. Little-c creativity, also referred to as "everyday creativity," results in creating something new that has originality and meaningfulness (Richards, 2007) and mini-c creativity is "the novel and personally meaningful interpretation of experiences, actions, and events" (Beghetto and Kaufman, 2007, p. 73). Although appearing quite similar, mini-c creativity is an internal process that consists of ideas and connexions that may not always be visible to anyone except the creator and can be challenging to measure, understand and value in the dementias.

A fourth perspective that contributes to our understanding of creativity and dementias is the heuristic approach proposed by Batey's (2012), which is oriented toward developing a framework for measuring creativity across three axes: the level to be measured (e.g., individual, group, community), the facet of creativity to be assessed (process, press, product, and trait) and the measurement approach (e.g., objective, self-rating). The inherent flexibility of this framework offers the possibility of developing longitudinal research; it fits well across different types of dementia, addresses challenges in measuring and understanding creativity as impairment increases over time and takes into consideration changes in the home, community, hospital and residential environments, (e.g., settings). Batey's approach also provides a useful measurement strategy that can be used across the three frameworks cited above (Beghetto and Kaufman, 2007; Richards, 2007; Glăveanu, 2013; Kharkhurin's, 2014).

A Snapshot of Dementia and Creativity Research

Over the past 10 years, there has been an increasing interest in research on dementia, the arts and creativity across different disciplines (Palmiero et al., 2012). Creative expression in artistic activities such as painting or making music, for example, has been found to be an important way for people with a dementia to express and access emotions even when cognitive abilities are diminishing (McLean, 2011; Zeilig et al., 2014). Rather than as a form of treatment for cognitive decline, creative activities involving the arts are often used in the context of therapy as part of the treatment of behavioural and emotional problems in dementias (Cowl and Gaugler, 2014). Previous research argued that art therapy was a potentially beneficial non-pharmacological intervention for dementia to improve quality of life (Mimica and Kalini, 2011). However, optimal conditions in the design of art interventions for the dementias to foster creativity need to be identified (Chancellor et al., 2014). This was reflected in a recent review of studies on art therapies and dementia revealing incoherent methodologies and tools used to assess creativity where a majority of studies focused on and judged the final product (e.g., Joy and Furman, 2014), for instance a completed picture or other artwork, rather than the process of engaging with the creation of art (Crutch et al., 2001; Beard, 2011).

Different forms of arts-based creative expression have been adopted for dementia populations (e.g., visual art making, playing music and singing, storey-telling, poetry). Ullán et al. (2012), examining an art making educational programme for people with mild to moderate dementia, discovered that participants showed surprised satisfaction at being capable of making art and of having created something with their hands, which appeared to reinforce a more positive self-image. Additionally, in a blocked randomisation design with individuals with moderate to severe dementia engaging in singing, listening to, and creating music, a reduction of agitated behaviours was observed during the intervention as well as at 1-month follow-up (Lin et al., 2011). Finally, (Fritsch et al., 2009) through randomised matched pairs incorporated storytelling as a creative intervention with nursing home residents with dementia and their carers, and discovered, compared to the control group, those using the creative intervention showed significant increases in pleasure, engagement and alertness, interacted more with nursing home staff, and socialised more. In a follow-up study with the same intervention, significantly improved communication skills both with carers and peers were also observed in people with a dementia who had participated in a creative expression intervention through storytelling (Phillips et al., 2010).

In a review of studies and case reports on creativity in dementia, Palmiero et al. (2012) discovered that although people with dementia were generally found to be able to express artistic creativity, divergent thinking was considered to be affected in both artistic and non-artistic people with a dementia in the sense that those with a dementia were found to be less inventive in creating novel art products. For instance, previous research

observed alterations in visual art productions in individuals with different forms of dementia and although drawings by individuals with Alzheimer's disease were closest to drawings of healthy controls, individuals were found to use more muted colours and included fewer details (Rankin et al., 2007). However, Ullán et al. (2012) argue that more simplistic forms of artistic expression do not necessarily mean less creativity.

Furthermore, creativity and creative expression have been found to look different depending on the type of dementia and its corresponding area of the brain as well as the context of the creative activity. Based on a review by Grettton and ffytche (2014), it appears possible that a unique artistic signature exists for each type of dementia diagnosis with different expressions of creativity in visual art depending on the area of damage in the brain. Research looking at creativity and dementia with Lewy bodies (DLB), which examined drawings of a visual artist before and after the onset of the dementia, discovered a gradual decline in all artistic qualities except for novelty as the disease progressed (Drago et al., 2006). Art produced by individuals with semantic dementia has previously been described as being “bizarre” and “distorted” and failing tests of divergent thinking (Rankin et al., 2007). Lower ability levels for creative expression have also been identified in individuals with a diagnosis of frontotemporal dementia (Joy and Furman, 2014) due to degeneration of the frontal and temporal regions of the brain. According to de Souza et al. (2010; p. 3733) any form of artistic expression is thought to be due to “involuntary behaviours” rather than as an expression of purposeful creativity. However, the question arises, even though artistic expression changes after onset of dementia, does this imply a reduction of creativity or a different form of creativity? Likewise, what type of creativity is being considered? For the purposes of this paper, we are interested in understanding everyday artistic creativity (Richards, 2007), most decidedly being of the little-c or mini-c variety where the focus is on the non-expert (Kaufman and Beghetto, 2009).

Co-creativity – Mapping the Concept

Like artistic creativity in people with dementias “co-creativity” is a nascent concept that has yet to be fully theorised. Nonetheless the term is steadily gaining in popularity, indeed the closely allied phrase “co-creation” can be found in various contemporary media (Zeilig et al., 2018). However, there is currently no agreed definition of co-creativity and therefore the concept itself remains somewhat indistinct. The emphasis in business and design contexts is upon the transfer of value from an end (or predefined) product to a shared process in which all those involved play an integral role in bringing something that is mutually valued into existence (Branco et al., 2017).

Artistic co-creativity as theorised and practised with people with a dementia shares some similarity to the understandings offered by design and business, in particular the possibility that distinctions can be erased between the artist-producer and participant-artist (Zeilig et al., 2018). Equally, the emphasis on the equal contribution of all involved is pertinent. However, it fundamentally differs conceptually in that the objective is not to co-design a product or work toward a single composition or performance. The work of Matarasso (2017) has been

informative here. He similarly discusses co-creation in the context of arts-based projects and how artists do not instruct but rather disperse the authority associated with their skills, thereby privileging the creative process over an end product. However, this is not to imply that lone creativity does not also involve intense and embodied engagement with the processes of creating. As cogently outlined by Banfield and Burgess (2013) in their reconceptualisation of Csikszentmihalyi's (1997b) “flow” experience within artistic practise, process is key for individual artists too. These authors suggest that flow, an integral part of the creative process, is particularly important for visual artists who work in two dimensions (Banfield and Burgess, 2013, p.74). The distinction in terms of co-creativity is that creative process and allied experiences of flow are more likely to be shared between two people or by multiple people at group events.

Thus, although there is not currently a single agreed definition for co-creativity it is characterised by a number of key features including centrally, a focus on shared process, the absence of a single author (hence unity and shared ownership), inclusivity, reciprocity and relationality. Co-creativity relies on dialogic and empathic approaches (Sennett, 2012) where through the process of exchange, understandings are expanded, although not necessarily resolved. This is in contrast to dialectic encounters which tend to lead to closure (Sennett, 2012, p. 24). Above all, it contrasts with notions of the lone creative genius that have tended to dominate views of creativity.

The role and value of the creative arts for people living with a dementia has been widely appreciated (Young et al., 2016; Camici et al., 2017; Windle et al., 2017), yet it has not explicitly focused on the ability of people with a dementia to interact and engage as co-creators. This may also reflect different disciplinary aims and theoretical perspectives, and the location of the majority of theories of creativity within a cognitive framework (Plucker and Beghetto, 2004) but may also be linked with dominant perceptions that people with a dementia are less capable of creative interactions (Basting and Killick, 2003; Ullán et al., 2012). There is thus a nascent but steadily growing recognition that people living with a dementia may be able to engage co-creatively with the arts (Kontos et al., 2017, p. 188).

“...individuals with dementia can make recognisably creative contributions despite the absence of sensical language.”

Co-creativity using the arts extends an invitation to participate in an aesthetic process and allows unique opportunities for communication and expression. The possibility that co-creativity can challenge the dominant biomedical perspective that associates the dementias with irretrievable loss and decline by creating opportunities for creative agency is a foundational premise of the projects presented below. As a process and as a tool or strategy for self-actualisation, in which micro-acts of artistic creativity gain significant importance within a group setting, co-creative activity may therefore be positively associated with the maintenance and promotion of various aspects of health and wellbeing (Price and Tinker, 2014) as well as providing important opportunities for playfulness and fun.

HOW DO PEOPLE WITH A DEMENTIA PERCEIVE CREATIVITY?

A search of the literature revealed no studies that examined how people with a dementia perceive and appreciate *their own* artistic creativity. We have found this omission to be problematic in that creativity has become defined by others (e.g., researchers, clinicians, the general public) without taking into consideration the perspectives and experiences of those living with a dementia. One recent systematic review (Nyman and Szymczynska, 2016, p.104) identified the pursuit of new leisure activities (including the arts) as a way for people with dementia “to avoid becoming stagnant . . . and to create a new path . . . (whilst) leaving a legacy for younger generations,” yet absent was how those with a dementia value or understand their own creativity. Although changing, the perspectives of people with a dementia have historically not been taken into consideration when planning services or undertaking research (Wilkinson, 2001). Any conceptualisation of creativity and dementia, we argue, needs to take into consideration the perspectives of those with a dementia along with caregivers, both formal and informal. As part of the development of our understanding of creativity and the dementias we felt it essential to seek the perspectives of people with a dementia and caregivers about this topic. In preparation for this article the authors sought to broaden their understanding of artistic creativity and the dementias beyond the research literature by having a series of conversations with people with a dementia and caregivers. Not designed as a research project that sought to generate new data, the following questions helped to form our conversation: What does creativity mean to you in your day to day life? How do you personally understand artistic creativity? How does creativity impact dementia and how does dementia impact creativity? Is creativity always something positive, and if not, when is it not positive? **Supplementary Table S1** provides a sample of responses, which along with previous and ongoing research, have contributed to our conceptualisation of how artistic creativity is experienced by those with a dementia and caregivers.

CREATIVITY IN CONTEXT

Over a 2-year period (2016–2018) the authors, an interdisciplinary group of researchers, artists and media professionals, have been involved in a series of art experiments at Created Out of Mind¹ a Wellcome Trust funded project examining the potential of different art forms and cultural activities to help better understand the experience of the dementias and likewise, to appreciate how the dementias might influence our understanding of artistic creativity. This section reports on several of those ongoing and novel initiatives and presents new methodologies that have not yet been used in creativity and dementia research. These diverse projects occurred across different dementias and levels of impairment in community and residential care settings as well as in more

traditional laboratory environments and in public forums. All projects have been ethically reviewed and approved by faculty ethics panels at either University College London or Canterbury Christ Church University. Some of these projects have been presented at conferences, others will be written up for journal articles whilst others are early days research that will be further developed.

Creative Opportunities in Dementia Care Environments

About one-third of people with dementia live in residential care and approximately two thirds of people who live in care homes are thought to have dementia (Department of Health, 2013). Care homes face many conflicting pressures involved in delivering day-to-day care, often described as task focussed, and despite best intentions, there is often limited scope for staff and residents to engage in meaningful activities together. Although problems in measuring creativity in this environment are pronounced, nevertheless, there is a growing recognition of the capacity of care homes for establishing artistic/creative residency programmes. In many instances this is motivated by a wish to improve the quality of life of those living with dementia (e.g., Cutler et al., 2011) and there is increasing evidence supporting the role of the arts across a range of positive outcomes (e.g., Windle et al., 2016).

Co-creativity and Advanced Dementia

Helping to provide a stimulating and creative caring residential environment for those with advanced dementias has often been overlooked or simply not considered as part of national dementia care policies (All Party Parliamentary Group on Arts, Health and Wellbeing [APPG], 2017). The practise of *Music for Life* founded in 1993 by Linda Rose has, however, placed a particular emphasis on working with people with advanced dementias. The intention to create community and shared experience through the use of musical improvisation has many parallels with a co-creative approach and is framed by both mini- and little- c creativity (Kaufman and Beghetto, 2009). By improvising pieces of music together (the genesis of creative expression as described through mini-c creativity) professional musicians, people with advanced dementias and professional care staff are engaged in musically responding to each other through what we have labelled as taking creative risks (e.g., picking up an instrument and playing for the first time; conducting the group for a brief period of time; responding musically to a musician or other group member). As dementias progress, many but not all (e.g., those with frontotemporal lobe type dementia) may lose confidence, interest and optimism in their abilities. Attending an arts group where everyday creativity (little-c creativity) and interaction with other members and facilitators is encouraged, may need to be gradually introduced in order to reduce anxiety and encourage participation and creative risk taking. In doing so members have the opportunity to relate to one another in ways that they might not do so usually and beyond the usual restrictions of their perceived roles. By shifting the emphasis onto relationship and communication processes rather than achieving a specified outcome, an ability

¹<http://www.createdoutofmind.org/>

and desire to engage in mutual exchange is revealed. In the project, *Music for Life 360*, several novel technologies were used to capture psychophysiological information, through wearable data collection devices (see section “Psychophysiological Responses to Creativity for People Living with Dementia”), and group interaction processes recorded through 360-degree video cameras (360fly, Canonsburg, PA, United States). The use of a 360-degree camera allows simultaneous interactions to be captured and later more fully understood through slowed-down (0.25 s per frame) video analysis using a software programme. This has enabled greater clarity in ascertaining the extent to which people living with advanced dementias are responding to co-creative interactions, whereas observational methods are more influenced by vocal and motor responses and possible biases of observers (Zeilig and West, 2017). The question of whether or how moments of shared creative experience affect us, regardless of our stage of life and cognitive ability, is addressed. Indeed, the idea that highly trained professional musicians might be stimulated and influenced by their creative interactions with people with advanced dementias could be a meaningful illustration of the concepts of creative and relational agency where the creators are interdependently engaged with a social and material world within a cultural context of artefact production (Glăveanu, 2013). The artefact production in this context (singing) is *both* process and product.

Residential Caregiver Involvement in Creative Activity

Equally important, professional caregivers’ experiences of creativity in practise is a powerful tool toward enhancing care quality. These can enhance client-carer interactions, validating the personhood in residents with dementia (Broome et al., 2017). For example, Basting et al. (2016) describe how they enacted a depiction of *The Odyssey* in the day-to-day running of care facility. This engaged residents, staff and family members in a uniquely creative way to improve quality of life and showed how the arts can transform environments.

Working to reach socially isolated residents within the care environment (e.g., bed bound, those displaying distressing behaviours), one such programme, *Living Words*, developed a 7-stage residency process. Residencies to date have taken place in 24 residential settings, with 820 participants and include using the “listen out loud” method (Gardner, 1983) to co-create an individual book of poetry with each participant focusing on their emotional experiences rather than cognitive abilities, which may vary greatly across participants.

Influenced by Kaufman and Beghetto (2009) mini-c (“genesis of creative expression,” p. 2), Richards (2007) everyday creativity (little-c creativity) model and Batey’s (2012) heuristic framework, creativity was explored through relationship building and the process of constructing poems together. As an example, Sherman was known to shout and interrupt people, banging his fist on a table. Artists were told that he was “incoherent” and had “challenging behaviours”. Through working with a *Living Words* (2014) artist who wrote down and then read his words to him, he began to express his feelings: “I am scared. . . I don’t know where

I am.” The validation of his emotions, words and even the fist banging led to him verbalise more (mini-c creativity), while his banging and shouting lessened. This creative relationship enabled staff to better understand Sherman the person, rather than just his dementia. This supports previous findings that through creativity in dementia, “feelings of peace may be generated” Zeilig et al. (2014, p. 26).

Another resident, Sally, spoke very quietly and in metaphor. This made it hard for staff to hear and understand her. On seeing the *Living Words* book she co-created and hearing her words read to her, staff reported being able “to see” the meaning in her words. For example, staff realised when Sally spoke of machines she was talking about brains; when she informed them that “the world is talking” she was referring to the care home. Sally’s voice became louder and she expressed joy in sharing her book, “One becomes a little more alive . . . Not just hanging there.”

Profiles in Paint and Single Yellow Lines

Taking a brush to canvas is an artistic activity available to most people with a dementia, regardless of previous visual arts experience. Guided by Glăveanu’s 5-A’s framework, one method we have begun using to capture artistic creativity through painting, in the context of different dementias, has been to invite people with an interest in art-making to arrange a group of 12 objects and independently produce a still life painting of their arrangement. The first exploratory study, *Profiles in Paint*, involved four people with different diagnoses of dementia, behavioural variant frontotemporal dementia (bvFTD), primary progressive aphasia (PPA), posterior cortical atrophy (PCA), typical Alzheimer’s disease (tAD) and a control group of four people without a diagnosis (Harrison et al., 2017). All artists received the same materials and instructions and the procedural framework allowed comparisons to be made between the works. For example, the artist with bvFTD approached the exercise in a way that accentuated their individual artistic interests whilst the artist with PPA created a structure to communicate relationships between the objects. The artist with PCA and the artist with tAD both found some of the objects perceptually challenging but this also allowed for a greater focus on the sensual qualities of the medium. Giving people with a dementia a choice over object arrangement also allowed a cooperative interaction to occur with the researcher that facilitated further understanding of perceptual, emotional and motivational aspects of creativity.

Since 2016, the *Single Yellow Lines* project has been examining the creative potential of painting a line. Initially 55 people who attended Rare Dementia Support Groups (PCA/PPA/FTD) were invited to paint a straight line on one canvas and a line of their choice on a second canvas. A further 99 people without a dementia at public events have painted their own straight and expressive lines. The straight lines are initially being examined in laboratory and cultural venue environments as a potential measure for the spatial disruptions people with PCA experience. However, it is interesting that due to the decentralisation of perceptual experience associated with PCA, the expressive lines made by people with this diagnosis have also appeared the most expressive to many observers (e.g., neuropsychologists, artists, general public). For people whose verbal language skills are

compromised the expressive line may also offer opportunities to communicate in another form, using images, words or metaphors.

We are continuing to investigate if the paintings made in these projects may be indicative of common symptomatic features of different dementias. Through public engagement events we have also observed how paintings have been powerful tools for communicating different experiences of the dementias to diverse audiences, ranging from neuroscientists to the general public. The projects aim to broaden the debate on the concept and manifestation of creativity in the dementias and seek to challenge the assertion that definitive interpretations about artistic creativity can be made in relation to diagnostic criteria. As with some definitions of creativity discussed earlier, it is perhaps in the *process* of creating that is felt most intensely (mini-c creativity) and because of this, the pleasures that are manifest in painting are not necessarily compromised in the context of a dementia.

The Neuronal Disco: Dancing Connexions Between Art and Science

Creativity research also has a role to play in conveying scientific complexities in dementia research to a wider audience outside of academia. One area of this research looks at how artistic responses to various aspects of brain abnormalities can offer audiences new insights into the mechanisms supporting the growth and degeneration of brain cells. For example, in order to investigate why abnormalities in the protein tau can lead to neuronal death in familiar Alzheimer's disease (fAD) and frontotemporal dementia (FTD), fibroblasts (skin cells) generated from participants carrying genetic mutations linked to disease are reprogrammed into induced pluripotent stem cells (iPSC). These iPSC can subsequently be differentiated into any cell type of interest, including neurons, which can be grown in both 2D and 3D culture formats (Arber et al., 2017). Comparisons between the neurons grown from participants with and without dementia can then be used to understand the earliest changes in disease cultures.

Grounded in Kaufman and Beghetto (2009) "Four C Model," a new component of this research also investigates how researchers and artists might effectively convey scientific information (Big C-creativity) through creative activities with people living with FTD and Familial Alzheimer's (fAD) as well as reflecting on the profound personal, ethical and metaphysical implications that these technologies present. As part of an initial pilot study a visual and performance artist began to consider how she could represent and embody (Pro-C creativity) what was growing in the laboratory in a form which would dynamically convey the earliest stages of cellular change and encourage public dialogue and discussion about the dementias. Researching ways to animate each change and structure of the cell development through choreographed formations of growth and degeneration, identified music, movements and groupings which could express different morphologies of dementias through a kind of cellular hybrid of country and disco dancing. The resulting *Neuronal Disco* (little-c), was subsequently trialled as a form of public engagement dance initiative to encourage

people of different ages to discuss dementia (Murphy and Wray, 2016).

Devised initially as a creative exercise to better understand these cellular processes, the *Neuronal Disco* evolved into a playful participatory event intended to engage public audiences in the science and aims of this research. Artist and scientist team leaders guide participants through each stage of the research in a series of choreographed groupings which mirror cellular mechanisms and transformations at different scales, performing axonal transport using illuminated balloons as vesicles and coloured streamers to create neuronal networks and tangles (mini-c) (Murphy and Wray, 2016). Appropriating rituals and accessories from rave and party contexts, participants were invited to wear small lights placed on all five fingers in colours matching the stains used to identify particular proteins, while their sound and light bracelets lit up in response to themed music (Wray and Murphy, 2017).

The *Neuronal Disco* invites a broad audience to consider the impacts of dementia on a molecular level through playful physical enactments. Abstract laboratory-based processes (mostly off limits to the public) are transformed into accessible group interactions which are informed by the laboratory team's perspectives (who perform this work on a day to day basis) and the artist's perspective (who has observed her own cells being transformed).

Performing each stage of the research together as a group helps us to creatively interpret and conceptualise the molecular dimensions of dementia research, offers insights into the science behind this research and opens up a new perspective on how we think about and visualise life altering diagnoses. Through the use of public engagement in dancing (mini-c creativity) where no previous dance experience is expected, the general public participates in an enjoyable creative activity as they learn about some of the laboratory science in dementia research. These types of activities also have the potential of shaping public attitudes toward the dementias, lessening stigma and supporting dementia friendly communities (All Party Parliamentary Group on Arts, Health and Wellbeing [APPG], 2017).

PSYCHOPHYSIOLOGICAL RESPONSES TO CREATIVITY FOR PEOPLE LIVING WITH DEMENTIA

Understanding creative experiences through psychophysiological measures has the potential to allow researchers to more fully comprehend physiological responses across periods of time, different dementia diagnoses and impairment severity. These measures are not dependent on cognitive ability and can be used longitudinally across the progression of dementia to assess reactions and responses to different art forms (e.g., playing music, poetry, singing, and painting) (Harding et al., 2017) during mini and little-c creative activities in individual and group settings. Psychophysiological measures have been shown to correlate with involvement during creative practise in a wide range of arts activities (e.g., De Manzano et al., 2010; Tschacher et al., 2012; Tröndle et al., 2014). Such measures offer an objective measure of participants' involvement or engagement in creative

practises complementary to more subjective self-report measures such as visual rating scales and interviews, during earlier and middle stages of dementia, and with video recording and other observational tools during later stages when impairment is severe. Recent advances in wearable technology have decreased costs and increased accuracy of unobtrusive devices so that they are now similarly accurate in emotion recognition tasks (Ragot et al., 2017). To better understand psychological and physiological responses to creative arts activities by those with a dementia, wearable technology has been used to continuously measure psychophysiological changes *during and across activities* (Bourne et al., 2017). Empatica E4 wristbands (Empatica, Cambridge, MA, United States), watch-sized devices, were employed to measure the following (Brotherhood et al., 2017):

- *3-axis accelerometer*: Provides information about levels of physical activity.
- *Electrodermal activity (EDA, an indication of arousal)*: A measure of emotional and sympathetic response useful for detection of levels of emotional and physiological arousal.
- *Blood volume pulse (BVP, used to derive heart rate)*: Used as a measure of heart rate, which may indicate excitement, stress and/or increased physical activity.
- *Peripheral skin temperature (an indication of stress)*: Similarly, to arousal, a measure of stress for determining level of stress in a wide range of activities.

Due to their high sampling rate wristbands such as the E4 collect vast amounts of continuous data capturing psychophysiological responses during creative activities, which can be collected unobtrusively across community and residential care settings. Because participants appear not to be aware they are wearing the devices this potentially makes these measurements more representative of a creative experience than an experimental condition. The unobtrusive nature of the devices also permits the collection of meaningful levels of baseline data which aid interpretation and analysis.

The interpretation of physiological data is not straightforward. For example, as well as participation, increased activity levels could signal agitation (e.g., fidgeting, attempts to leave the room), and emotional arousal could be positive or negative, and even when negative, this could be an engaged and meaningful response to a challenging artwork, and possibly indicate an embodied form of “flow state” (wide Banfield and Burgess, 2013) or a feeling of disgust accompanied with a desire to withdraw from the activity at hand. Difficulties with interpretation arguably make isolated use of such measures problematic (Thomas et al., 2018). Furthermore, there is far less experimental control and far greater complexity in creative arts activities than in carefully controlled psychology experiments. Ideally such data should be interpreted alongside supplementary observational field notes or video data to re-contextualise moments of physiological activation. The issues of interpretation also raise important questions about hypothesis development of creative involvement and whether such activities are studied and measured with the intention of improving wellbeing, quality of life, levels of emotional engagement or communication between a person with dementia

and their family member. Batey’s (2012) creativity framework is useful here to help situate the level, facet and measurement approach. As an objective measure continuous physiological measurement lends itself to examining process over a specified time period in individuals, dyads and groups. It also can be combined with other objective measures and subjective ratings to produce a more comprehensive assessment of creativity.

Issues surrounding interpretation also have a bearing on the analytic approach taken with such data. In the early literature on EDA (previously termed galvanic skin response), heart rate and other measures such as electromyography, the prevailing approach was to hypothesise response increases as markers of anxiety, stress, threat-detection and other tension, (e.g., Darrow, 1936; Dittes, 1957; Fowles, 1980). This has contributed to implicit assumptions that higher psychophysiological markers equate with someone being more stressed, anxious or uncomfortable. Secondly, engagement with the arts or other creative processes is much less clearly delineated as being wholly negative or positive, stressful or pleasant, and it seems that the level and quality of engagement itself would be the most appropriate proxy for any measurement of the quality of the experience; whether that be feelings of great tension while grappling with a new medium or composition choices, increased heart rate when joining an improvised dance or playing a piece to the point of crescendo.

In agreement with the majority of the literature in this field (Thomas et al., 2018), we have found that psychophysiological measures are useful in the context of understanding process responses whilst participating in creative activities (Bourne et al., 2017). In particular, using wearable devices to complement mixed-methods approaches to creative involvement and activity we are able to provide quantitative data to test various hypotheses, some across discrete moments in time.

VISUAL THINKING STRATEGIES

Perhaps one of the most valuable aspects of art, in any form, is that it creates an ambiguous space of being able to create in which there are no right or wrong answers. Yet the feeling of getting it wrong is unfortunately an experience many people living with a dementia can often relate to (Batsch and Mittelman, 2012). There is therefore a need for clinical assessments of dementia that minimise creating a sense of failure, taking into account a person’s rich life experiences and looking at their current difficulties as well as their functional capabilities. One way to approach this problem is to investigate the potential of the arts-based facilitated learning method Visual Thinking Strategies (VTS) to help people living with a dementia create meaning through viewing visual art, whilst also promoting social wellbeing and potentially serving as a valuable diagnostic tool for clinicians (van Leeuwen et al., 2017b). VTS lends itself to Kharkhurin’s (2014) four criterion construct of creativity involving attributes of novelty, utility, aesthetics, and authenticity (meaning). It also draws on Batey’s (2012) heuristic framework that provides flexibility in designing research with different measurement approaches, studying individuals within group settings (level), while focusing on the facets of process, trait or press.

Visual Thinking Strategies is constructed as a moderated group discussion which allows people to create meaning based on their personal observations of visual art. The moderator uses clearly described techniques to carefully structure the discussion: (1) asking participants to identify visible references for their thoughts and pointing these out, (2) neutrally paraphrasing each comment, and (3) connecting the comment to the ongoing discussion. In education, neuro-rehabilitation and museum settings, VTS has been shown to improve written and spoken language skills as well as social, observation and critical reflection skills (Housen, 2002; Naghshineh et al., 2008; Miller et al., 2013; Hailey et al., 2015).

We are exploring if VTS can enable people living with a dementia to express their personal experiences and feel socially connected without relying on memory or previous knowledge. The ideal context for a VTS conversation is a small group setting with the art object present in its original form and viewed under optimal lighting and spatial conditions. However, in order to operationalise the complex interaction between social context, visual thought processes and moderating techniques at play in VTS a computer-based eye tracking paradigm (Isaacowitz et al., 2006) has been designed to monitor these interactions. People are shown visual artworks and complex images on a computer monitor and the eye-tracker records what their eyes are looking at and in which order eye movements occur. In separate experiments people are shown each artwork for various amounts of time. In one experiment they are being played audio recordings of other people reflecting on the artworks while they are looking, in another they are being asked to personally reflect on the artworks with the 3 VTS questions. The focus of this novel method is on *how* people create personal meaning in relation to what they see, hear and communicate. This approach allows people to express themselves freely, lessening the concern they are getting it wrong, often a commonly voiced concern of people with a dementia.

The ultimate aim of this methodology is to harness its findings into guidelines for cultural VTS programmes tailored to people living with a dementia as well as developing a validated diagnostic assessment tool for clinicians, which lessens the distress and discomfort often experienced in current neuropsychological assessment.

CONCEPTUALISING CREATIVITY IN THE DEMENTIAS

Creativity research in psychology has a long history of being constructed through a cognitive lens; we argue that this is problematic for those with a dementia and others with neurocognitive disorders because it potentially devalues their capacity to be creative. As cognitive capabilities decrease it is essential to examine situational, social and environmental components—in addition to *or instead of* cognitive components—to better understand the value of mini and little c models of creativity (Kaufman and Beghetto, 2009) and they might pertain to people with a dementia (Plucker and Beghetto, 2004; Palmiero et al., 2012; Young et al., 2016).

Even as artistic expression may change over the course of the dementias (Crutch et al., 2001), and as cognitive abilities decline, there remain possibilities for artistic creativity to develop. Moreover, as Ullán et al. (2012) noted, simpler forms of artistic expression should not be equated with a lower level of creativity. Whilst there may or may not be reductions in creative activity in a specific art form (e.g., oil painting, glass blowing, ballroom dancing) during any phase of the dementias, this does not imply that alternative forms of creative activity cannot be developed.

The cognitive dominance in creativity research has been reinforced by theoretical assumptions that are not always applicable to this population. Quantitative approaches to creativity often involve measuring levels of memory, motivation, perception and behaviour that vary tremendously across the types of dementia and corresponding levels of impairment, making the use of questionnaires and scales as data gathering tools unreliable or invalid. Qualitative research has mostly relied on structured interviews and observations, with inherent assumptions about a person's capacity to verbally respond to questions and reflect on recent activities, both of which greatly diverge across the dementias. Underpinning this is the often-unspoken assumption by some researchers and clinicians that people with a dementia are not creative, nor can they continue to learn or participate meaningfully in new activities (Bellás et al., 2018).

More recently arts-related programmes in dementia care have been recommended for health and social care, charities and local communities to implement (e.g., All Party Parliamentary Group on Arts, Health and Wellbeing [APPG], 2017). Research relating to artistic creativity in the dementias has tended to focus on understanding the participatory aspects of specific art activities (e.g., Zeilig et al., 2014; Camic et al., 2016; Unadkat et al., 2017; Windle et al., 2017) within the context of healthcare or public health outcomes. However, in order to fully appreciate the complexity and potential of artistic creativity across different art forms, types of dementia, contexts most suitable to enhance and stimulate creativity, as well as approaches to measurement, we believe it is essential to conceptualise creativity in the dementias as a process that is not solely dependent on cognitive aptitude or skills, and to free it from the domain-general vs. domain-specific dichotomy that is “one of the most enduring controversies” in creativity research (Plucker and Beghetto, 2004, p. 153).

Going beyond this debate, considerable evidence from non-dementia research supports the idea that creativity has both specific *and* general components, yet a third component, the social environment of the individual (Amabile, 2013) is also fundamental to a conceptualisation of creativity in dementia. For people with a dementia, the social environment can help foster creativity. In particular, co-creativity is characterised by social interaction between two or more people in a supportive environment (including: home, public space, community centre, residential care, palliative care).

Rather than seeing creativity as necessitating an end product, creativity in the dementias emphasises process and experience (Killick and Craig, 2012), whereas co-creativity adds components such as mutual endeavour, relational interactions and notions of shared creativity. The emphasis on artistic creative *process* rather than on creative *outcomes*, is a necessary shift away from

pre – post measurement of specific variables at given points in time. This shift allows new forms of measurement to be considered, such as obtaining continuous psychophysiological measures of specific moments in time; undertaking longitudinal ethnographic research looking at both the development of and changes in creativity; using eye tracking devices to better understand what is being seen in the moment; investigating the relationship between creative activity and wellbeing (e.g., Strohmaier and Camic, 2017). Emphasis on process over outcomes we argue, is also a more ethical way to research artistic creativity in individuals with a dementia because it places less emphasis and demand on production and end point measurement, whilst giving more attention to encouraging enjoyment, collaboration, exploratory trial and error and discovering what is possible, rather than establishing what is not.

ETHICS STATEMENT

This projects reported in this article were carried out in accordance with the recommendations of the British Psychological Society research guidelines and approved by faculty ethics committees at University College London and Canterbury Christ Church University. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

PC developed the idea for the article, which was further discussed and refined by SC, CM, EH, CH, SS, JW, GW, and HZ. All authors

contributed sections including revisions. HZ, GW, SC, and PC critically revised further draughts. All authors read and approved the final manuscript.

FUNDING

This study was supported by Wellcome Trust grant # 200783/Z/16/Z to SC (P-I).

ACKNOWLEDGMENTS

We would like to thank the many people with a dementia, their family members and care staff for their participation in one or more of the projects described in this article, and for helping to inform how we as researchers and artists are coming to understand creativity in the dementias. Special thanks to Tracy Shorthouse and Kerrie Marshall for answers to our many questions. We would also like to acknowledge the support of: AgeUK Camden, Alzheimer's Society, BBC, Jewish Care, Rare Dementia Support Groups, Resonate Arts, Royal Society for Public Health, Wallace Collection, Wellcome Collection, and Wigmore Hall.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01842/full#supplementary-material>

REFERENCES

- All Party Parliamentary Group on Arts, Health, and Wellbeing [APPG]. (2017). *Creative Health: The Arts for Health and Wellbeing*. London: APPG.
- Amabile, T. M. (2013). "Componential theory of creativity," in *Encyclopedia of Management Theory*, ed. E. H. Kessler (London: Sage).
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edn. Washington, DC: American Psychiatric Association.
- Arber, C., Lovejoy, C., and Wray, S. (2017). Stem cell models of Alzheimer's disease: progress and challenges. *Alzheimer's Res. Ther.* 13, 1–17. doi: 10.1186/s13195-017-0268-4
- Banfield, J., and Burgess, M. (2013). A phenomenology of artistic doing: flow as embodied knowing in 2D and 3D professional artists. *J. Phenomenol. Psychol.* 44, 60–91. doi: 10.1163/15691624-12341245
- Barb, S. A., and Plucker, J. (2002). Smart people or smart contexts? Talent development in an age of situated approaches to learning and thinking. *Educ. Psychol.* 37, 165–182. doi: 10.1207/S15326985EP3703_3
- Basting, A., and Killick, J. (2003). *The Arts and Dementia Care: A Resource Guide*. New York, NY: The National Centre for Creative Aging.
- Basting, A., Towey, M., and Rose, E. (2016). *The Penelope Project. An Arts-Based Odyssey to Change Elder Care*. Iowa City, IA: University of Iowa Press.
- Batey, M. (2012). The measurement of creativity: from definitional consensus to the introduction of a new heuristic framework. *Creat. Res. J.* 24, 55–65. doi: 10.1080/10400419.2012.649181
- Batsch, N. L., and Mittelman, M. S. (2012). *World Alzheimer's Report 2012: Overcoming the Stigma of Dementia*. London: Alzheimer's Disease International.
- Beard, R. L. (2011). Art therapies and dementia care: a systematic review. *Dementia* 11, 633–656. doi: 10.1177/1471301211421090
- Becker, H. (2004). *Art Worlds*. Berkeley: University of California Press.
- Beghetto, R. A., and Kaufman, J. C. (2007). Toward a broader conception of creativity: a case for mini-c creativity. *Psychol. Aesthet. Creat. Arts*, 1, 73–79. doi: 10.1037/1931-3896.1.2.73
- Bellas, S., Balmer, A., May V., Keady, J., Buse, C., Capstick, A., et al. (2018). Broadening the debate on creativity and dementia: a critical approach. *Dementia* doi: 10.1177/2F1471301218760906 [Epub ahead of print].
- Bourne, P., Crutch, S. J., Fancourt, D., and Camic, P. M. (2017). "Choral Singing and dementia: using physiological and psychological measures of stress and wellbeing," in *Proceedings of the Powerful Partners: Advancing Dementia Care through the Arts and Sciences Conference* (London: Royal Society for Public Health).
- Branco, R. M., Quental, J., and Ribeiro, O. (2017). Personalised Participation: an approach to involve people with dementia and their families in a participatory design project. *Int. J. Co-Creat. Des. Arts* 13, 127–143. doi: 10.1080/15710882.2017.1310903
- Broome, E., Denning, T., Schneider, J., and Brooker, D. (2017). Care staff and the creative arts: exploring the context of involving care personnel in arts interventions. *Int. Psychogeriatr.* 29, 1979–1991. doi: 10.1017/S1041610217001478
- Brotherhood, E., Ball, P., Camic, P. M., Evans, C., Fox, N., Murphy, C., et al. (2017). Preparatory planning framework for created out of mind: shaping perceptions of dementia through art and science. *Wellcome Open* 2:108. doi: 10.12688/wellcomeopenres.12773.1
- Camic, P. M., Baker, E., and Tischler, V. (2016). Theorising how art gallery interventions impact people with dementia and their caregivers. *Gerontologist* 56, 1033–1041. doi: 10.1093/geront/gn v063

- Camici, P. M., Hulbert, S., and Kimmel, J. (2017). Museum object handling: a health-promoting community-based activity for dementia care. *J. Health Psychol.* doi: 10.1177/1359105316685899 [Epub ahead of print].
- Chad J. K., Lea A. L., and Brown, C. (2007). Creative styles and gender roles in undergraduates students. *Creat. Res. J.* 19, 273–280. doi: 10.1080/10400410701397396
- Chancellor, B., Duncan, A., and Chatterjee, A. (2014). Art therapy for Alzheimer's disease and other dementias. *J. Alzheimer's Dis.* 39, 1–11. doi: 10.3233/JAD-131295
- Clarke, A. E., Friese, C., and Washburn, R. S. (2018). *Situational Analysis*. London: Sage.
- Cowl, A. L., and Gaugler, J. E. (2014). Efficacy of creative arts therapy in treatment of Alzheimer's disease and dementia: a systematic literature review. *Activ. Adaptat. Aging* 38, 281–330. doi: 10.1080/01924788.2014.966547
- Crutch, S. J., Issacs, R., and Rosser, M. N. (2001). Some workmen blame their tools: artistic change in individual with Alzheimer's disease. *Lancet* 357, 2129–2133. doi: 10.1016/S0140-6736(00)05187-4
- Csikszentmihalyi, M. (1997a). *Creativity: The Psychology of Discovery and Invention*. New York, NY: HarperCollins.
- Csikszentmihalyi, M. (1997b). *Finding Flow*. New York, NY: Basic Books.
- Cutler, D., Kelly, D., and Silver, S. (2011). *Creative Homes: How the Arts Can Contribute to the Quality of Life in Residential Care*. London: Baring Foundation.
- Darrow, C. W. (1936). The galvanic skin reflex (sweating) and blood-pressure as preparatory and facilitative functions. *Psychol. Bull.* 33, 73–94. doi: 10.1037/h0051940
- De Manzano, Ö., Theorell, T., Harmat, L., and Ullén, F. (2010). The psychophysiology of flow during piano playing. *Emotion* 10, 301–311. doi: 10.1037/a0018432
- de Souza, L. C., Volle, E., Bertoux, M., Czernecki, V., Funkiewiez, A., Allali, G., et al. (2010). Poor creativity in frontotemporal dementia: a window into the neural bases of the creative mind. *Neuropsychologia* 48, 3733–3742. doi: 10.1016/j.neuropsychologia.2010.09.010
- Department of Health (2013). *A State of the Nation Report on Dementia Care and Support in England*. London: UK Department of Health. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/262139/Dementia.pdf
- Dittes, J. E. (1957). Galvanic skin response as a measure of patient's reaction to therapist's permissiveness. *J. Abnorm. Soc. Psychol.* 55, 295–303. doi: 10.1037/h0048306
- Drago, V., Crucian, G. P., Foster, P. S., Cheong, J., Finney, G. R., Pisani, F., et al. (2006). Lewy body dementia and creativity: case report. *Neuropsychologia* 44, 3011–3015. doi: 10.1016/j.neuropsychologia.2006.05.030
- Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., et al. (2014). Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol.* 13, 614–629. doi: 10.1016/S1474-4422(14)70090-0
- Fowles, D. C. (1980). The three arousal model: implications of Gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. *Psychophysiology* 17, 87–104. doi: 10.1111/j.1469-8986.1980.tb00117.x
- Fritsch, T., Kwak, J., Grant, S., Lang, J., Montgomery, R. R., and Basting, A. D. (2009). Impact of timeslips, a creative expression intervention program, on nursing home residents with dementia and their caregivers. *Gerontologist* 49, 117–127. doi: 10.1093/geront/gnp008
- Gardner, H. (1983). *Frames of Mind: The Theories of Multiple Intelligences*. New York, NY: Basic Books.
- Glăveanu, V. P. (2013). Rewriting the language of creativity: the Five A's framework. *Rev. Gen. Psychol.* 17, 69–81. doi: 10.1037/a0029528
- Gretton, C., and flytche, D. H. (2014). Art and the brain: a view from dementia. *Int. J. Geriatr. Psychiatry* 29, 111–126. doi: 10.1002/gps.3975
- Hailey, D., Miller, A., and Yenawine, P. (2015). "Understanding visual literacy: the visual thinking strategies approach," in *Essentials of Teaching and Integrating Visual and Media Literacy: Visualizing Learning*, eds D. M. Baylen and A. D'Alba (Berlin: Springer International Publishing), 49–73. doi: 10.1007/978-3-319-05837-5_3
- Harding, E., Sullivan, M. P., and Crutch, S. J. (2017). "Engaged or exasperated? Interpreting physiological data in dementia research," in *Proceedings of the Powerful Partners: Advancing Dementia Care through the Arts and Sciences Conference* (London: Royal Society for Public Health).
- Harrison, C. R., Carton, A. M., Brotherhood, E. V., Hardy, C. J. D., Cohen, M. H., Warren, J. D., and Crutch, S. J. (2017). Profiles in paint: contrasting responses to a common artistic exercise by people with different dementias. *Arts Health* [Epub ahead of print]. doi: 10.1080/17533015.2017.1402795
- Housen, A. (2002). Aesthetic thought, critical thinking and transfer. *Arts Learn. J.* 18, 99–132.
- Isaacowitz, D. M., Wadlinger, H. A., Goren, D., & Wilson, H. R. (2006). Selective preference in visual fixation away from negative images in old age? An eye-tracking study. *Psychol. Aging* 21, 40–48. doi: 10.1037/0882-7974.21.1.40
- Joy, F. P., and Furman, L. (2014). "Progressive change in formal qualities of art produced over the course of frontotemporal dementia," in *Proceedings of the 122nd Annual Convention 7–10 August* (Washington, DC: American Psychological Association).
- Kaufman, J. C., and Beghetto, R. A. (2009). Beyond big and little: the four C model of creativity. *Rev. Gen. Psychol.* 13, 1–12. doi: 10.1037/a0013688
- Kharkhurin's, A. V. (2014). Creativity. 4in1: four-criterion construct of creativity. *Creat. Res. J.* 26, 338–352. doi: 10.1080/10400419.2014.929424
- Killick, J., and Craig, C. (2012). *Creativity and Communication in Persons with Dementia*. London: Jessica Kingsley Publishers.
- King Humphry, E. (2010). *The Mad Artist's Brain: The Connection between Creativity and Mental Illness: Scientific American*. New York, NY: Springer Nature. Available at: <https://www.scientificamerican.com/article/the-mad-artists-brain/?print=true>
- Kontos, P., Miller, K., and Kontos, A. (2017). Relational citizenship: supporting embodied selfhood and relationality in dementia care. *Sociol. Health Illness* 39, 182–198. doi: 10.1111/1467-9566.12453
- Lin, Y., Chu, H., Yang, C. H., Chen, C. H., Chen, S. G., Chang, H. J., et al. (2011). Effectiveness of group music intervention against agitated behavior in elderly persons with dementia. *Int. J. Geriatr. Psychiatry* 26, 670–678. doi: 10.1002/gps.2580
- Living Words (2014). *The Things Between Us: Words and Poems of People Experiencing Dementia*. Kent: Shoving Shepard Books.
- Matarasso, F. (2017). *Co-Creation, from "A Restless Art"*. Available at: <https://arestlessart.com/2017/03/09/co-creation/> [accessed March 9 2017].
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R. Jr., Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* 7, 263–269. doi: 10.1016/j.jalz.2011.03.005
- McLean, J. (2011). *An Evidence Review of the Impact of Participatory Arts on Older People*. London: Mental Health Foundation.
- McMullan, G., and Smiles, S. (2016). *Late Style and its Discontents: Essays in Art, Literature and Music*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198704621.001.0001
- Miller, A., Grohe, M., Khoshbin, S., and Katz, J. T. (2013). From the galleries to the clinic: applying art museum lessons to patient care. *J. Med. Human.* 34, 433–438. doi: 10.1007/s10912-013-9250-8
- Mimica, N., and Kalini, D. (2011). Art therapy may be beneficial for reducing stress-related behaviours in people with dementia—Case report. *Psychiatr. Danubina* 23, 125–128.
- Murphy, C., and Wray, S. (2016). *Introducing the Neuronal Disco and Brains in a Dish*. London: Created Out of Mind, Wellcome Collection. Available at: www.createdoutofmind.org
- Naghshtineh, S., Hafler, J. P., Miller, A. R., Blanco, M. A., Lipsitz, S. R., Dubroff, R. P., et al. (2008). Formal art observation training improves medical students' visual diagnostic skills. *J. Gen. Intern. Med.* 23, 991–997. doi: 10.1007/s11606-008-0667-0
- Nyman, S. R., and Szymczyńska, P. (2016). Meaningful activities for improving the wellbeing of people with dementia: beyond pleasure and meeting fundamental psychological needs. *Perspect. Public Health* 136, 99–107. doi: 10.1177/1757913915626193
- Osborne, T. (2003). Against "creativity": a philistine rant. *Econ. Soc.* 32, 507–525. doi: 10.1080/0308514032000141684
- Palmiero, M., Di Giacomo, D., and Passafiume, D. (2012). Creativity and dementia: a review. *Cogn. Process.* 13, 193–209. doi: 10.1007/s10339-012-0439-y

- Phillips, L. J., Reid-Arndt, S. A., and Pak, Y. P. (2010). Effects of a creative expression intervention on emotions, communication, and quality of life in persons with dementia. *Nurs. Res.* 59, 417–425. doi: 10.1097/NNR.0b013e3181faff52
- Plucker, J., Beghetto, R. A., and Dow, G. (2004). Why isn't creativity more important to educational psychologists? Potential, pitfalls, and future directions in creativity research. *Educ. Psychol.* 39, 83–96. doi: 10.1207/s15326985ep3902_1
- Plucker, J. A. and Beghetto, R. A. (2004). "Why creativity is domain general, why it looks domain specific, and why the distinction does not matter" in *Creativity: From Potential to Realization*, eds R. J. Sternberg, E. L. Grigorenko, and J. L. Singer (Washington, DC: American Psychological Association), 153–168
- Pope, R. (2005). *Creativity: Theory, History, Practice*. Oxford: Routledge.
- Price, K. A., and Tinker, A. M. (2014). Creativity in later life. *Maturitas* 78, 281–286. doi: 10.1016/j.maturitas.2014.05.025
- Ragot, M., Martin, N., Em, S., Pallamin, N., and Diverrez, J. M. (2017). Emotion recognition using physiological signals: laboratory vs. wearable sensors. *Adv. Intell. Syst. Comput.* 608, 15–22. doi: 10.1007/978-3-319-60639-2_2
- Rankin, K. P., Liu, A. A., Howard, S., Slama, H., Hou, C. E., Shuster, K., et al. (2007). A case-controlled study of altered visual art production in Alzheimer's and FTLD. *Cogn. Behav. Neurol.* 20, 48–61. doi: 10.1097/WNN.0b013e31803141dd
- Richards, R. (2007). "Everyday creativity: our hidden potential," in *Everyday Creativity and New Views of Human Nature*, ed. R. Richards (Washington, DC: American Psychological Association), 25–54.
- Rowlands, M. (2010). *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. Cambridge, MA: MIT Press.
- Runco, M. A. (1987). The generality of creative performance in gifted and nongifted children. *Gifted Child Quart.* 31, 121–125. doi: 10.1177/001698628703100306
- Sennett, R. (2012). *Together: The Rituals, Pleasures & Politics of Cooperation*. London: Penguin.
- Strohmaier, S., and Camic, P. M. (2017). "Conceptualising what we mean by wellbeing in the dementias," in *Proceedings of the Powerful Partners: Advancing Dementia Care through the Arts and Sciences Conference* (London: Royal Society for Public Health).
- Thomas, G. E., Crutch, S., and Camic, P. M. (2018). Measuring physiological responses to the arts in people with a dementia. *Int. J. Psychophysiol.* 123, 64–73. doi: 10.1016/j.ijpsycho.2017.11.008
- Tröndle, M., Greenwood, S., Kirchberg, V., and Tschacher, W. (2014). An integrative and comprehensive methodology for studying aesthetic experience in the field: merging movement tracking, physiology, and psychological data. *Environ. Behav.* 46, 102–135. doi: 10.1177/0013916512453839
- Tschacher, W., Greenwood, S., Kirchberg, V., Wintzerith, S., van den Berg, K., and Tröndle, M. (2012). Physiological correlates of aesthetic perception of artworks in a museum. *Psychol. Aesthet. Creat. Arts* 6, 96–103. doi: 10.1037/a0023845
- Tu, W. M. (1985). *Confucian Thought: Selfhood as Creative Transformation*. Albany, NY: State University of New York Press.
- Ullán, A. M., Belver, M. H., Badia, M., Moreno, C., Garrido, E., Gomez-Isla, J., et al. (2012). Contributions of an artistic educational program for older people with early dementia: an exploratory qualitative study. *Dementia* 12, 425–446. doi: 10.1177/1471301211430650
- Unadkat, S., Camic, P. M., and Vella-Burrows, T. (2017). Understanding the experience of group singing for couples where one partner has a diagnosis of dementia. *Gerontologist* 47, 468–478. doi: 10.1093/geront/gnv698
- van Leeuwen, J., Crutch, S. J., Suarez Gonzalez, A., Boomgaard, J., and Warren, J. (2017a). "Looking at the bigger picture: visual art and the social brain," in *Proceedings of the Powerful Partners: Advancing Dementia Care through the Arts and Sciences Conference* (London: Royal Society for Public Health).
- van Leeuwen, J., Murphy, S. C., Harrison, C., and Camic, P. M. (2017b). *Shaping Public Perception of Dementia Through the Arts and Sciences*. London: Wellcome Collection Reading Room.
- Wilkinson, H. (ed.). (2001). *The Perspectives of People with Dementia*. London: Jessica Kingsley.
- Williams, R. (1988). *Keywords: A Vocabulary of Culture and Society*. London: Harper Collins.
- Windle, G., Gregory, S., Howson-Griffiths, T., Newman, A., O'Brian, D., and Goulding, A. (2017). Exploring the theoretical foundations of visual art programmes for people living with dementia. *Dementia* doi: 10.1177/1471301217726613 [Epub ahead of print].
- Windle, G., Newman, A., Burnholt, V., Woods, B., O'Brian, D., Baber, M., et al. (2016). Dementia and imagination: a mixed-methods protocol for arts and science research. *BMJ Open* 6, 1–12. doi: 10.1136/bmjopen-2016-011634
- Wray, S., and Murphy, C. (2017). "Disco morphologies: dancing connections between art and science," in *Proceedings of the Powerful Partners: Advancing Dementia Care Through the Arts Conference* (London: Royal Society for Public Health).
- Young, R., Camic, P. M., and Tischler, V. (2016). The impact of community-based arts and health interventions on cognition in people with dementia: a systematic literature review. *Aging Ment. Health* 20, 337–351. doi: 10.1080/13607863.2015.101108
- Zeilig, H., Killick, J., and Fox, C. (2014). The participative arts for people living with a dementia: a critical review. *Int. J. Ageing Later Life* 9, 7–34. doi: 10.3384/ijal.1652-8670.14238
- Zeilig, H., West, J., van der Byl, and Williams, M. (2018). Co-creativity: possibilities for using the arts with people with a dementia". *Qual. Ageing Older Adults* 19, 35–145. doi: 10.1108/QAOA-02-2018-0008
- Zeilig, H., and West, J. (2017). "Pleasures, perils and possibilities: an exploration of co-creativity," in *Powerful Partners: Advancing Dementia Care through the Arts and Sciences Conference* (London: Royal Society for Public Health).

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 Camic, Crutch, Murphy, Firth, Harding, Harrison, Howard, Strohmaier, Van Leewen, West, Windle, Wray and Zeilig. 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.



Portrait of an Artist as Collaborator: An Interpretative Phenomenological Analysis of an Artist

Ian Hocking*

School of Psychology, Politics and Sociology, Canterbury Christ Church University, Canterbury, United Kingdom

The subjective experience of being an artist was examined using interpretative phenomenological analysis (IPA), focusing on the perspective of the artist but interpreted by me, a psychologist, from my perspective as an artistic collaborator. Building upon a literature that has hitherto focused on clinical, elderly, or vulnerable participants, I interpreted superordinate themes of *Process* (Constraint, Playfulness, Movement) and *Identity* (The Ill-Defined Artist, Becoming, Mixing Identities, Choosing an Identity, Calling, Collaboration, and Outsider). These themes are broadly similar to the existing literature, but emphasise identity while de-emphasising self reflection and the need to become an “insider.”

OPEN ACCESS

Edited by:

Timothy L. Hubbard,
Arizona State University, United States

Reviewed by:

Andrew Patrick Allen,
Maynooth University, Ireland
Casimiro Cabrera Abreu,
Queen's University, Canada
Angelos Mouzakitis,
University of Crete, Greece

*Correspondence:

Ian Hocking
ian.hocking@canterbury.ac.uk

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 08 March 2018

Accepted: 25 January 2019

Published: 12 February 2019

Citation:

Hocking I (2019) Portrait of an
Artist as Collaborator: An
Interpretative Phenomenological
Analysis of an Artist.
Front. Psychol. 10:251.
doi: 10.3389/fpsyg.2019.00251

Keywords: case study, creativity, collaboration, artist, interpretative phenomenological analysis

PORTRAIT OF AN ARTIST AS COLLABORATOR: AN INTERPRETATIVE PHENOMENOLOGICAL ANALYSIS OF AN ARTIST

Modern psychology has had a long association with artistic works, examining the psychological characteristics of, for example, architecture (Woelfflin, 1886, as cited by Jarzombek, 2000) and expressionism (Worringer, 1911). With a movement toward the Gestaltist approach (Perls et al., 1951), the field emphasised internal representation, as well as therapy. Interest waned in the 1970s amid criticisms that art itself is too subjective an experience to render using the ostensibly objective framework of psychological theory, and with individual reactions to art being too variable.

The positivist approach, which characterises much of contemporary psychology, argues that observation and experiment are the only sources of substantive knowledge (Colman, 2015). Under this auspice psychology has explored, for instance, aesthetic preference and appreciation (e.g., symmetry and compositional balance; see Lindell and Mueller, 2011). Meanwhile, wider creativity research has explored personality-based, cognitive, contextual, psychometric, psychoanalytic, and pragmatic approaches (Mayer, 1999), but commentators advocate that more dialogue between these areas is needed (Nelson and Rawlings, 2007).

One reason for the separation of quantitative and qualitative streams is the tractability of creative phenomena—broadly defined—at different levels. We can see this separation most clearly in memory research. Our understanding of low-level aspects of memory is well advanced (Baddeley, 2012), but higher-level, and potentially more meaningful research into, say, how memories inform our identity is less coherent, partly hampered by the nature of the phenomenon: it is less suitable to a quantitative, cumulative discipline. In the case of creativity, if the parameters of a creative task are set by experimenters—and thus the motivational and emotional aspects creativity are rendered

more artificial—the creative process will be undermined, or at least changed substantially from the process as it manifests in real life. We know, for instance, that individuals have been shown to perform better in problem-construction activities that correspond to their own values and interests (Mace and Ward, 2002). This is an issue, then, of the applicability of much general, quantitative creativity work to those creative individual involved in a particular field of expertise. In recent years, qualitative approaches have grown in popularity (see Smith and Osborn, 2015). These emphasise a deeper, more meaningful analysis of phenomena, and commonly feature a thorough treatment of verbal texts (i.e., any object that can be read).

When we look at research on artists being artists, qualitative approaches predominate, and, among various themes, *identity* in its broadest sense is particularly important. Johnson and Wilson (2005), for instance, studied women who were following a multi-generational discipline of textile handcraft. The study combined questionnaires, historical research and participant observation. Across several meaning-construction themes such as production, and the use of what was produced, identity was overarching; the role of producing the textiles gave them, in some sense, their “place in the world” (p. 118). A similar underlying principle was discovered in a narrative inquiry investigating how unpaid arts and crafts contribute to retired people’s sense of occupational identity (Howie et al., 2004). In this study, where six creative industry participants looked back on their lives, the maintenance of their creative identities was founded upon the social embeddedness of practise, an awareness of themselves and their skills changing over the lifetime, a complementary awareness of certain qualities in themselves being stable, and the opportunity to remain reflective on how their creative products gave them a sense of self, and of their life’s journey. Process—i.e., the sense of identity as a changing, responsive quality—is emphasised in an ethnographic study by feminist author Clark/Keefe (2014). An artist herself, identity for Clark/Keefe is what one *becomes*. Spence and Gwinner (2014) provide a similar narrative on the relationship between art, identity, and mental health by an artist living with mental illness, written in conjunction with an Artist in Residence. One of the important things to come out of this research was the notion of an individual maintaining a duality between their artistic identity and their identity as a person with a mental illness; they are not, therefore distinct, though might be presented as such to the outside world. The notion of being an “outsider” is also important, but in the context of attempting to become an “insider.” This is further emphasised by Perruzza and Kinsella (2010), who reviewed the literature on the usefulness of creative arts occupations for therapeutic practise; they identified several important factors, including collaboration, efficacy, and benefits for individual identity—all implicated, to greater and lesser degrees, in participants’ “sense of self” (p. 265)—as well as their social identity. The importance, again, of identity was reiterated in Reynolds et al. (2011), who studied twelve older female visual artists living with arthritis, finding evidence that their artistic activities helped maintain a positive outlook. Finally, Reynolds and Vivat (2010) examined another sample of older women living with chronic fatigue syndrome (also known as ME); a thematic

analysis suggested that the women fell into two groups. For some, their creative works enabled them to recover some of the previous identity that their illness had diminished; for others, their art provided them with a more positive identity, and this group felt that they had become artists. Thus not only is identity central to creative individuals, but *becoming*, or making the transition to artist, can be important too.

Elsewhere, Mace (1997) used a Grounded Theory approach to explore professional artists in New Zealand. They found that *movement*, as a metaphor, was important because each artwork develops over time. The process of artistic creation is viewed as a continuous cycle of problem-finding and problem-solving; communication between these two elements is crucial. For artists, the exploratory stages are sometimes the most engaging, where natural playfulness and freedom add to the enjoyment. Mace and Ward (2002) extended these findings with another Grounded Theory analysis of artists. As before, the emphasis was on a model of artists’ creative process during a time when, importantly, they were producing their own artwork rather than anything specified by researchers. They identified four stages of development: conception, idea development, making the artwork, and finishing. Again they emphasise *movement*, not necessarily linear, sometimes cycling from broad conception to finished artwork. Similarly, physical *constraints*, helpful or not, are critical in the production of art and often shape the nature of the final piece. Another concept suggested is exploration or *playfulness*: being motivated by enjoyment and keeping options open. This stage-like conception of the creative process has antecedents including Wallas (1926), who proposed preparation, incubation, illumination and verification. Preparation involves breaking down the problem and identifying which skills and knowledge will be required to progress with it. Incubation requires setting aside the problem. Illumination is characterised as a sudden insight into the solution, which is then tested during the verification stage. Some consider problem finding to be “pre”-stage (e.g., Amabile, 1996). Others focus on the distinction between implicit and explicit process, which are consistent with two processes: a fast, automatic mechanism and a slow, deliberate one (Allen and Thomas, 2011).

The Current Study

The above studies are drawn from groups of individuals where responses are pooled by researchers uninvolved in the artistic process itself. Multiple participants can be useful in making conclusions more generalisable, for instance in research on visual artists, for which we know a great deal about the relationship between creativity and perceptual abilities, drawing skills, autobiography and personality (Locher, 2010). This is less informative within a qualitative context; here, we are just as interested in how a given reality is constructed. Furthermore, much of this literature, because it has sought adults spending much of their time in purely artistic endeavours (i.e., in the production of artefacts that are typically novel and valuable), has necessarily tended toward groups of retirees, or those recovering from illness. This is not the approach of the present study. In June of 2015, I was contacted by the organisers of a contemporary art festival who wished to embed an artist within the environment of

a researcher examining creativity from a quantitative perspective. This allowed me to study a young professional creative individual over a 12-month period, communicating by way of regular face-to-face meetings, telephone conversations, a shared blog, as well as email; finally, I conducted three interviews investigating themes based on the psychological literature and concepts that appeared to be important from our communications. Crucially, this artistic process was collaborative, allowing me to go beyond the typical “outsider” perspective to an ethnographic or participant observation approach, and addressing the call of Freeman (2014) that artists and psychologists should collaborate in their pursuit of understanding creativity. Locher (2010) observes that our knowledge of the artistic process primarily comes from archival case studies and real-life case studies. The former typically involves the examination of working draughts, such as the those for Picasso’s *Guernica* (see Weisberg, 2004). Clearly, a limitation of this approach is that the work is not captured *in vivo*. Real-life case studies avoid this limitation by analysing the artwork from beginning to end (e.g., Miall and Tchalenko, 2001). Locher goes on to observe that factors related to autobiography, motivation, culture and history will contribute to a final artwork, and the complex interplay between them may be less suited to an experimental approach. Gruber’s evolving systems approach takes a similar stance, where the construction of meaning is emphasised, along with close study of the creator, and consideration of the sociohistorical milieu (Gruber, 1980).

The current study takes the approach of Interpretative Phenomenological Analysis (IPA), a qualitative technique that helps us understand how participants make sense of their personal and social world (Smith and Osborn, 2015). The term “phenomenology” is used in the broad sense of being concerned with subjectivity, rather than the narrower sense in which it is used in phenomenological psychology, which is the application of Continental philosophical phenomenology to academic psychology (Valle et al., 1989). The focus of IPA is a nexus of specific experiences, events and states. It draws heavily upon phenomenology, the philosophical study of consciousness, experience, and the structures that support them. Although phenomenologists do not agree amongst themselves on a formal definition of the term, the current paper takes the approach that phenomenological investigation should be systematic, involve reflection and study, and that the phenomena concerned should be those arising from acts of consciousness. This follows from the work of Husserl (1931) and later thinkers such as Ricoeur (1990) who underscored the complex relationship between meaning, narrative and forms of identity. Nelson and Rawlings (2007) characterise this approach as being more about the “whatness” than the “whyness.” One issue with IPA, which we should bear in mind, is that the descriptions of a person’s internal states and behavioural processes are necessarily limited by their ability to accurately introspect on the processes that generate them (Perkins, 1981, as cited by Mace and Ward, 2002). We should also bear in mind that much artistic work might be intuitive and thus implicit; indeed, for some researchers, this is a hallmark of creativity (Nelson and Rawlings, 2007; cf. Allen and Thomas, 2011).

The construction of meaning is a complex process, and no less so in the context of IPA. We can consider two sources of meaning-making—the individual(s) under study and the person conducting the analysis—but these must be set against the wider complexities of meaning-making in an extra-individual world (Berger and Luckmann, 1991). IPA can be conducted on groups of individuals or a single individual. For instance, in the current study, I, Ian, as a psychologist, will attempt to explore meaning making with the artist, whom I will call Jane; the scope of her experience will be her life in contemporary art, as well as our artistic collaboration. The key aspects of IPA are: (i) an inductive approach, where hypotheses and prior assumptions are avoided; (ii) participants tend to be experts in the area of interest and have the ability to describe their thoughts, commitments and feelings; (iii) researchers reduce experiential data complexity through rigorous and systematic analysis; and (iv) analyses include both an individual, idiographic perspective as well as a sense of commonality with other data (Reid et al., 2005). A successful and valid analysis is interpretative (subjective, with no attempt to be “factual”), transparent (where the journey from data to interpretation is clear) and plausible (to the participant, to the researcher, and to general readers). Throughout this process, I bore in mind Mace and Ward (2002) observation: “. . .the genesis of artwork arises from a complex context of art making, thinking, and ongoing experience” (p. 182).

As there is no prescribed approach for phenomenological methods, it has been argued that it should adapt to the unique qualities of the phenomenon under study (Wertz, 1983). The late stage interviews should be seen in the context of a long term collaboration with the artist; as such, it represents the “tip of the iceberg.” As Tzanidaki and Reynolds (2011) have argued, sample size has traditionally not been seen as an indicator of quality in the qualitative approach because rich data and nuanced analysis often trump quantity. Reid et al. (2005) also warn against assuming a linear relationship between number of participants and the value of research. Further, in the present paper, it is not the intention to present data that has been sampled from a notional population of artists; this is not a strength of the phenomenological approach, even within larger samples, and it is difficult to imagine what the larger population of artists actually would be, given the particularly individual ways in which artists go about their work (and is arguably a central issue for studies where artists of from differing disciplines are mixed, e.g., Nelson and Rawlings, 2007). For this reason, and others, Smith (2004) has argued that there can be advantages for smaller sample sizes and case studies, such as the multiple case studies examining the role of art-making in identity maintenance for those living with cancer (Reynolds and Prior, 2006).

Some personal identifying information has been changed.

METHODS

Design

This study uses IPA to analyse transcribed interviews conducted with an artist in the 11 month of a 12-month artistic collaboration. As well as these transcripts, analysis was informed

by a shared blog, emails, telephone conversations and face-to-face meetings. However, excerpts from the interviews alone are presented here; it was agreed early in the process that making our general communications subject to study would have introduced a harmful self-consciousness to the project.

Participant and Procedure

The case study involves one artist, pseudonymously called Jane, with whom I worked on a contemporary art installation. The installation involved listening to Jane's recreations of telephone calls to psychics, with the psychics trying to predict the nature of the installation. The installation took place in a blacked-out hut.

We spoke three times over 2 weeks for a total interview time of 3 h 30 min. A small amount of the transcription was done by a student intern and myself, but the majority by a graduate student. The final text base was 28,000 words.

Given the importance of "bracketing" presuppositions in IPA, the author underwent an initial self-reflective process that focused on the artistic collaboration from their perspective, their own artistic endeavours (in this case, novel writing) and the creativity literature (cf. a similar approach taken by Nelson and Rawlings, 2007). Additionally, throughout the artistic collaboration and during data collection, a reflective diary was used to assist in the process of reflection on the interviewer's thoughts and feelings (cf. Savin-Baden and Fisher, 2002); this was not to eliminate bias, which is inevitable, but acknowledge the presence of the researcher in the research process, helping to identify themes, and helping to enhance the research process (Finlay and Gough, 2008). This reflection brought home several points, which are personal to me and, whether or not they are factually correct, describe my views: I feel that art and artists are crucial to a functioning society, given the human need for expression and the value placed on the products of these expressions; my quantitative approach is perceived by most artists as reductionist; as a published novelist, I have some common ground with artists; openness on my part was crucial in the collaboration; the typical psychologist-participant power dynamic needed to be minimised but acknowledged.

To guide the interviews, I used a semi-structured format based on core, open categories: history, which focused on personal biography; views, which focused on what art, creativity and practise meant to the artist; and collaboration, which addressed previous collaborations as well as the present one. I made sure to touch upon the following concepts: work/life balance, identity, nature of creativity, collaboration, quantification, documentation, narrative, privacy, power/authority, prediction, and flow. I would introduce these by saying, for instance, "Now, I want to talk about identity. How would you describe your identity?" Or I might say, "Tell me about the role that work/life balance plays in your art." As we spoke, I made notes to record my thoughts, help think of further questions, and to guide my later interpretation of the interview transcript. The transcript was then read carefully and annotated with notes on particular meanings, which were then collected into the themes below.

While this paper examines individual components of experience, it does not present experience as a separate entity alongside other concepts; I see it as a higher level construct

that draws upon all concepts. Together, these comprise my interpretation of Jane's experience.

The study received ethical clearance from the Research Governance Committee of Canterbury Christ Church University (Ref: 16/SAS/277C). The case study was conducted with the full informed consent of the artist. This was signed prior to the interviews. She has also viewed and approved this final version of the article.

RESULTS AND DISCUSSION

Jane has read this manuscript. I have maintained the broad direction of my interpretation, but some of her comments are included as footnotes.

Jane is a professional artist in her mid-thirties. She started out as a painter but soon became interested in more contemporary forms of expression. For her, a key transition point was saving enough money to attend a "studio" programme abroad, after which she worked on what she now considers to be an artistic performance of the type she now pursues.

That's the first [artistic] work. That marks the point where I thought "I've found something interesting that isn't just drawing or dealing with something in a slightly. . ." Looking back, that feels like the work that marked the start of being an artist. (396)

SUPERORDINATE CATEGORY: PROCESS

Subordinate Category: Constraint

Jane appears to view orthodoxy as something that can be pushed against, tested, or broken. She sees orthodoxy as arbitrary and sometimes limiting. Challenging orthodoxy can be seen in some of her works, such as an installation that involved her wearing all her clothes at the same time. On the face of it, this is absurd, but can make the audience wonder why a particular way of dressing should be absurd, and what this might say about consumerism.

I suppose there's that, sometimes I want to respond to things, just like the idea of going crazy or doing something stupid, kind of breaching those "norms" which comes back to that normative ways of doing things. (882)

This corresponds with the artist as an explorer who isn't necessarily concerned, from the outset, where they might end up, which Nelson and Rawlings (2007) characterise as an attitude of risk-taking, of "engagement in a process of exploration without knowing exactly what is being looked for" (p. 222).

When she worked for her previous employer, Jane didn't like the constraints imposed by the system surrounding the job, particularly having to move in a direction that wasn't entirely consistent with her political position.

I found it very constraining and now I really enjoy what I do, and I don't quite know what that says about me. (1359)

This is not necessarily something unique to Jane, but it forms an important part of her identity. Reaction against constraint has long been considered an important quality of successful artists (though not for those where artistry is seen more in terms of a trade, e.g., the pre-Romantics; Brown, 1991). Shulman (1984), for instance, discussed the nineteenth century writers Hawthorne, Melville and Poe in context of their metaphorical prisons, where the prison is formed from artistic heritage: like prisoners, they feel a sense of enclosure; they work out ways of defying authority; they attempt to communicate with those on the outside. This is taken further, of course, with Postmodernism, where there is arguably even greater reaction against constraint, particularly those associated with Enlightenment rationality (Butler, 2002).

Subordinate Category: Playfulness

Closely related to this constraint—cf. Jane's use of "enjoy" in the above quote (1359)—is playfulness. This, for Jane, is about taking the everyday (and occasionally the less obvious) and giving herself the freedom to play with it, much as a child might play with a cup or a word. Convention-breaking features prominently in this, as does repurposing; putting something to a use that strains against the intention of the creator, or at least the normative use.

So what are the systems at work? What are the conventional ways in which things are being done? How I can use my practise to kind of intervene and who will understand that? Maybe play with it, and transform it, or subvert it somehow. (218)

In the above quote, Jane gives two elaborations of her "play" concept. The first is "transformation." This seems to hark back to an important tenet of what creativity means for most people: that is, creativity takes the raw materials of skill and experience to produce something new in the sense of being recombined or mixed. This clearly important for Jane. The artist is a lens between her audience and what she sees. The second term is "subvert": to undermine, destabilise, or unsettle. Jane seems to be using the term here in the broader sense of repurposing something for a use that is not intended. This, of course, is a shortcut to defamiliarisation, which allows the audience and the artist to go beyond the superficial, everyday conception of thing to a deeper understanding, or at least a reacquaintance with its nature. This reminds us of the classic Alternative Uses Task (Guilford, 1967), where participants must come up with different ways to use everyday objects, such as a brick, paperclip or newspaper. Alternatives to everyday or mundane function relates to avoiding cliché in fine art and fiction: a cliché like 'it was a dark and stormy night' is so common it will be hardly read; subverting the phrase to something like "It was neither dark, nor stormy, but night all the same" will cause the reader to reengage, and perhaps consider cliché in general.

The playfulness is an important part of collaboration, too. It's related to testing and trying out ideas.

Any opportunity to work collaboratively, that sort of playfulness, reels me in, it's fun and awful [Jane laughs], definitely interesting. (784)

This use of playfulness is subtly different. It's playing in the sense of bouncing ideas off people, of being surprised by them, and allowing a collaborator to introduce the unexpected, in a kind of third-party incubation (Wallas, 1926). This ties in with studies of visual artists, which show general agreement that such artists have no final image in mind before they start to sketch or paint (see Locher, 2010). There is an emotional, fun component to Jane's playfulness. Again:

...working on lots of field recordings, and I did some documenting. That was the first time that I collaborated and it was just really fun. (1004)

The playfulness also ties in with representation. Jane is clear that what she does as an artist goes beyond what might, at first blush, be termed simple representation. A photograph, for instance, is a relatively faithful representation of the physical characteristics of something external to the camera on that occasion, skewed somewhat by the camera's physical properties, post production, as well as the choice of the photographer to capture and present that particular moment. A photograph is not truly a simple representation, but it is comparatively simpler than the kind of contemporary art that features in Jane's portfolio. She talks about this in the quote below.

So, what's the difference between an anthropologist and an everyday artist, and the conclusion I came up with, was that it's something to do with representation, whereas an artist you can be very very playful, and your goal isn't necessarily to represent, or my goal is not to represent, it might be more to play with misrepresentation or sort of like tease, play more of a trickster role in a sense. So, it's not always really serious. ...It's saying that I don't really agree with that, and although some of the methods might be similar in just participating in the situation, being a participant, by making something, or turning something into some kind of knowledge, there is more scope to be playful or misrepresentative. (826)

Playfulness can be seen as a delay of closure; the idea that one is putting off the serious work of completion, after which there is no further opportunity for play. Getzels and Csikszentmihalyi's (1976) longitudinal study of artists suggested that artistic success was related to "delay in closure," that is, putting off the inevitable moment when an artist must commit. This point is reiterated by Mace (1997), who states that the artists in her study had an excitement with, and preference for, the experimental stage of the artistic process. This might be related to a lack of concern with goal-focused behaviour, or a wish to dwell within the part of the process that is most flexible and unset. There is also a sense in which it is hard to identify when an artwork is finished, perhaps due to a difficulty in objectively evaluating the artwork while still retaining an emotional connexion to it (Mace and Ward, 2002, p. 191). This also touches upon what (Nelson and Rawlings, 2007) have called the "freedom-constraint dynamic," which is about having enough freedom to

be creative, but not too much—as mentioned earlier in this paper, constraints are important in providing a path, even if they turn out to be not directly important in their own terms, like a “soup stone.”

Subordinate Category: Movement

Jane often spoke about the role of movement, which can be metaphorical as well as literal. It is linked to coming at something with fresh eyes but goes beyond newness for its own sake. Context is important in her work and, with changing context, comes changing ideas. In his interviews with older American artists, Santlofer (1993) makes the point that these artists are always on the look-out for discovery, including self-discovery: “constant struggle and reevaluation [is] inherent in the creative process” (p. 87).

Jane is interested in exploring, so novelty¹ is an important aspect of her work. She says, for instance,

I kind of feel that there is a role in being able to move around and I suppose, to offer a different perspective—questioning. This kind of “questioning” function—a challenge—a question, which is definitely a challenge function. (618)

This attempt to see things from a different angle, and distance, is similar to the theme “distant-engagement” identified by Nelson and Rawlings (2007) as “an alternation between immersion in the manipulation of material and distancing oneself” (p. 221).

Here is an example of movement in a metaphorical sense, connected to Jane’s common practise of creating works that are very much “new”² in the sense that, for her, she is not repeating herself:

I don’t like remaking, previous work—or restating previous work. I always like doing something different and moving on. (1413)

Though the movement often involves geographical travel, I felt that movement as a metaphor for travel and change is most attractive to Jane. When I asked her outright about the importance of travel, she was quick to identify its limits:

I don’t think it needs to be travelling to somewhere new. I don’t think it needs to be that at all. I think, often, through having done lots of residencies, for example, and having produced work through that, you are somewhere different. (201)

Turning to an artwork that explored the concept of risk assessment, she goes on to say:

I was in an art school and I just became very interested in this kind of form of procedure. (201)

So movement can be an artistic driver for Jane, but the connexion to movement is not a simple one. It involves the notion of “edge walking,” or the artist being an outsider making discoveries with fresh eyes, and producing an artwork

that sparks off this unfamiliarity³. The idea that unfamiliarity is consistent with examining more closely is echoed in Freeman (2014), who looked at artists drawing inverted faces, which are not organised according to well-known principles that shape the drawing of upright faces. Freeman goes on to write that trained artists gain familiarity with both their medium and their subject; this can lead to a kind of abstraction, or overview perspective, where the structure of the whole subject becomes as important as the details found within the structure. The movement can also be a form of escapism in its non-perjorative sense; Fisher and Specht’s (2000) study of older artists called this “escaping the mundanity of life,” as well as its aches and pains. There is an obvious connexion to Csikszentmihalyi’s (1997) concept of flow, and recent work by Zimbardo and Boyd (2008) on time perception: for them, individuals involved in creative endeavours (particularly when it provides immediate feedback) are likely to be more present-focused and hedonistic.

SUPERORDINATE CATEGORY: IDENTITY

Subordinate Category: The Ill-Defined Artist

Her identity as an artist is something that Jane has thought about and perhaps struggled with, though I think I might have been making more of this than she did; I was particularly interested how she saw herself, both as an artist and creatively.

I remember the moment when you think, “Am I an artist? Am I not an artist?” What do you have to be doing to be an artist? Do you just have to, like, say I’m an artist? What qualifies you to be an artist? It’s quite interesting. (458)

Later in this response, Jane talks about the artistic identity as being related to what is *done*. That is, action is a crucial component. One is, therefore, an artist because one attempts at art. This contrasts with the position of Clark/Keefe (2014), in which an artist is something that one becomes, rather than is or is not.

For Jane, art does not have to take up a majority of one’s time.

Lots of artists, because of the fact that you can’t make much money through being an artist [...] you have to find different jobs. You might work in construction, you might do teaching, and do all sorts of things. (458)

This comes back to a method of going about things. The artistic identity is one of method.

³Jane writes: “I think this is more about the artist adopting different roles or positions to explore how knowledge is constructed. I think “the artist being an outsider making discoveries with fresh eyes” sounds a bit clichéd. It’s not that simple. Being an outsider making claims on behalf of others is something I would never want to be associated with doing. Thinking critically about what it means to be an outsider, yes. Exploring power and positionality, yes. Juxtaposing different ways of knowing and doing, yes. Exploring the feeling of unfamiliarity that comes with being in a new situation, and using this as a trigger for work, yes. I’d just be very wary of ever claiming to have fresh or privileged eyes.”

¹Jane writes: “Not novelty. I’m not actively in search of the “new.” It’s more that I end up in situations I’m not an expert on, so I often assume the role of the novice, and end up having to familiarise myself with new situations.”

²Jane writes: “Though of course new works will build on previous works. The ideas I’m interested in don’t change.”

So, for me, my view on my art is kind of a way finding some sort of liminal way of operating. (882)

This touches upon a point made by Fisher and Specht (2000), who studied older artists. They seemed to have their identity as artists shored up by their sense of self in terms of competency and efficacy, what Herzog and House (1991) have referred to as the “agent self.” For Jane, there seems to be a sense in which the performance of art is “liminal,” operating on the threshold of art and not-art, or the familiar and unfamiliar.

Subordinate Category: Becoming

For Jane, making the transition between a less fulfilling professional career to the more interesting, but risky, career of artist was, obviously, important. She told me that she was interested in art from a young age, but was influenced to follow an academic career. At school, she had an art scholarship, and she considers art one of her best subjects. In terms of “becoming,” this, presumably, is the same struggle that afflicts all those who make the transition from the more orthodox, salaried track to the arts sector. Throughout the interview, I got the sense that she saw her life—her professional life, at least—as dividing very much into two. Indeed, when she first became active in the artistic community as a practitioner, she was not keen to disclose her former profession. This aspect chimes with the position of Clark/Keefe (2014); an artist is something one becomes.

And to start with, when I became an artist, I just didn’t talk about that period of my life at all. I didn’t want to identify with it, didn’t want to bring it up, because how can those two things. . . Those things feels diametrically opposed, like the [previous employer] and the artist. (466)

This sense of becoming is linked to expertise. This might be a categorical distinction for a third party observer, but for the artist the concept is more nuanced. An artistic expert, after all, is no more than a person with more mature creative processes (Mace and Ward, 2002). Additionally, artists with experience are more likely to know what can lead to success and failure. The kind of knowledge built up, according to Mace and Ward (2002), is “explicit and implicit understanding of techniques, skills, art genre, art theory, aesthetics, emotion, values, personal theories, personal interests and experience, previous work, and historical and contemporary art knowledge” (p. 183).

Subordinate Category: Mixing Identities

Immediately after saying the above about the separation between two professional identities, she adds:

But not necessarily. (466)

This reflects her belief that two identities might have appeared separate at the time, but from her present day perspective, they are less far apart. The separation seems less obvious given her experience now of being an artist.

Subordinate Category: Choosing an Identity

For Jane, one difference between being in her previous employment and being an artist seems related to self control. As an artist, you are, for better or for worse, your own person. Whereas, her previous employment had characteristics that led to situation where. . .

. . . you have to sacrifice your own identity, really. Completely. You have to. . . Yes, you’re doing a lot of problem solving, but you have to conform to the system, and I don’t like conforming to systems, have never. (482)

For Jane, then, this sense of ownership, or personal sovereignty, is an important part of being an artist. It also fits with what she sees as her non-conformist, iconoclastic attitude.

Subordinate Category: Calling

There is a sense in which the artistic identity is all-encompassing. Because it runs like a thread through everything, Jane rarely “switches off.” This is exacerbated by the sporadic nature of the freelance work, as well as its intrinsically enjoyable nature (see *Playfulness*), and “flow.”

It’s not your job, it’s a whole identity. . . you don’t quite know when the next pay cheque will come up, so you end up actually, well I do anyway, having a lot on and not eating until nine o’ clock or eleven o’ clock (642)

The idea of calling has been linked to an individual’s search for meaning in life and, for some researchers, this search for meaning is our primary drive (Frankl, 1985). Dobrow (2013) followed musicians over 7 years to identify factors related to their calling. She found that, far from being a stable construct, calling changes over time, which is consistent with Jane’s change from a person who is interested in art to a person who actively produces art and is part of the local and wider artistic community.

Subordinate Category: Collaboration

From the perspective of collaboration, Jane sees her identity as changeable and responsive to context. Collaboration also raises the issue of authorship; during collaboration, there is sense in which authorship is challenged. In the quote below, Jane refers to a previous collaboration in which an artistic colleague was offered, and used, Jane’s hard drive in an art installation. The collaboration raised issues of control and boundaries.

I’d like to think that authorship – I’m very lazy [Jane laughs] about it. But it’s interesting because that experiment [involving the handover of the hard drive] proved that there are certain things, which I feel a part of my identity as an artist, which is sort of the way in which I do things and it felt uncomfortable having someone replicate that so precisely. (1072)

The use of “lazy” above is interesting; it seems to be more about being patient and able to delegate, both of which are parts of her strategy to avoiding repeating herself in art. Indeed, collaboration, far from being an unusual part of the creative process, is fundamental to Jane.

I am producing some knowledge about something which takes a form of art but often it involves engaging with other people, so I think its more about that engagement with others and how do you represent that in the artistic process. (132)

And:

I tend to just find it really interesting listening to other people, I have to say. (164)

There are downsides to this collaboration. In our collaboration, she sometimes felt she was being measured and judged by me. Here is an excerpt including us both:

Interviewer: So [the middle of the collaboration] was that a kind of . . . that was an anxious time. (1680)

Jane: Yeah. Because I think, I initially had this sort of anxiety around [the question], “Would I be negatively impacted from through just being a participant?” And through having my process observed and particularly I was worried about [our private, collaborative blog] because . . . and [. . .] like when I was putting this blog post up today. . . I was like, “Do I really want to post that? Do I want to send that?”

Outsider

The concept of the outsider was raised repeatedly throughout the interview, mostly by me, reflecting my own notion of what the “typical” artist does. Jane is well aware of the literature on the role of a certain type of artist, that the idea of what she calls an “edge walker” is connected to the notion that she, as an artist, works best on the periphery or borderline. In this view, the artist is a person who takes a different perspective for a viewpoint advantage, just as a person walking a ridge can see down both sides of it⁴. Jane can be physically outside, or displaced, too.

I always end up immersing myself in different situations where I’m quite. . . I don’t know much about psychology, but here I am, so I’m like, “Oh that sounds interesting, I’d love to do that.” (132)

Jane goes on to say that this movement to the “outer” or “outside” realm is an important part of the fluidity of the creative process, which connects to my own experience of the artwork changing over the 9 months of our collaboration.

I . . . often [feel] like an outsider in different places from having moved around a lot, and not having a sense of, well, this is my home, this is my culture, but seeing that there is something more fluid. (164)

The notion of being outside feeds somewhat into her identity of *becoming*; because she didn’t become a professional artist as early as some of her artist friends, this helps avoid what she terms an “artist” bubble.

And I think that because I’ve moved around a lot, been heavily involved in lots of different professional different systems . . . I’m interested in the fact that things operate really differently outside to the art world, than inside the art world. So, I think you can get a bit of a—I’m making huge generalisations here [Jane laughs]—that you can get into a lot of art bubble I suppose, if you go through

art school, all your friends are artists who are going to art school, you carry on working, or you know, a lot of my friends aren’t artists, they’re not in that bubble, they struggle to make sense of contemporary art. So, yeah, it is sort of a different perspective on things. I have friends who are bankers.

CONCLUDING COMMENTS

The present study looked at the way in which one contemporary artist sees herself and her work, taken from the perspective of her collaborator.

The model outlined above—Process and Identity—suggests a separation between the components, but they are, unsurprisingly, well connected. Process, with its ideas of *constraint*, *playfulness*, and *movement* in all forms, are in many ways a reflection of Identity. Here, I’ve broken down identity into several elements, the first of which is the *ill-defined artist*: what we mean by art, artist, and creativity are questions that Jane touched upon throughout her interviews and our collaboration. There is a sense in which keeping these ill-defined allows for a protean, shifting and flexible self-characterisation that keeps avenues of expression open. The second, closely related concept is *becoming*: I use this in the sense of making the transition from the amateur to the professional, or in, another sense, reaching the point where Jane felt comfortable self-identifying as an artist. It involves expertise, commitment, and sacrifice. Third is *mixing identities*: Jane does not see her life as divided into sections where she is totally one thing or the other; as well as being an artist, she is a mother, friend, academic, and so on. *Choosing an identity*, the fourth strand, is about taking ownership of one’s identity, particularly when pitched against jobs or situations (such as her previous employer) in which conforming to a system can involve a “sacrifice [of] your own identity” (Jane: 482). For Jane, an important part of being an artist is regaining, and maintaining, sovereignty over one’s identity. *Calling*, the fifth strand, emphasises the all-encompassing nature of being an artist; meals, and much else, might be skipped in the service of art. This art-first approach was evident during our collaboration, but, as suggested by Dobrow (2013), the vocational sense is likely to change over time. Certainly, it would have been strong at the point Jane chose to pursue a path that took her away from a well-paid job with clear progression. The sixth part is *collaboration*, which brings with it issues of authorship and ownership; these can sometimes overshadow the art, but Jane sees the art that she produces as essentially collaborative, particularly in understanding the potential of the final artwork from the perspective of her “official” collaborator—me—and others (such as, in the case of our artwork, telephone psychics). Lastly, there is the concept of the *outsider*; Jane was wary of facile perceptions of the artist as an outsider. For her, this was rather more staying outside the “art bubble” (Jane: 585) than taking an “objective” stance toward her art. There is a sense in which being within this “art bubble” can lead to a parochial or less interesting approach.

Thus, at the end of this process, and though a psychologist/artist collaboration of the kind called for by Freeman (2014), I was able to identify superordinate themes

⁴Jane writes: “I’m just interested in an artistic practise at the juncture of social encounters, and what happens when different world views come together.”

of *Process* (Constraint, Playfulness, Movement) and *Identity* (The Ill-Defined Artist, Becoming, Mixing Identities, Choosing an Identity, Calling, Collaboration and Outsider). These are broadly similar to themes found in previous research cited in the Introduction and throughout this paper, which often draws from clinical, older, vulnerable or otherwise special participants, suggesting a commonality between these and the professional artist described here, though some qualities, such as self reflection (cf. Howie et al., 2004), the need to become an “insider” (Spence and Gwinner, 2014) were less important for Jane.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

- Allen, A. P., and Thomas, K. E. (2011). A dual process account of creative thinking. *Creat. Res. J.* 23, 109–118. doi: 10.1002/cad.20153
- Amabile, T. M. (1996). *Creativity in Context: Update to the Social Psychology of Creativity*. Boulder, CO: Westview Press.
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422
- Berger, P. L., and Luckmann, T. (1991). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. London: Penguin.
- Brown, M. (1991). *Preromanticism*. Stanford: Stanford University Press.
- Butler, C. (2002). *Postmodernism: A Very Short Introduction*. Oxford: Oxford University Press.
- Clark/Keeffe, K. (2014). Becoming artist, becoming educated, becoming undone: toward a nomadic perspective of college student identity development. *Int. J. Qual. Stud. Educ.* 27, 110–134. doi: 10.1080/09518398.2012.737048
- Colman, A. M. (2015). *A Dictionary of Psychology*. Oxford: Oxford University Press.
- Csikszentmihalyi, M. (1997). *Creativity: Flow and the Psychology of Discovery and Invention*. New York, NY: HarperCollins Publishers.
- Dobrow, S. R. (2013). Dynamics of calling: a longitudinal study of musicians. *J. Organ. Behav.* 34, 431–452. doi: 10.1002/job.1808
- Finlay, L., and Gough, B. (2008). *Reflexivity: A Practical Guide for Researchers in Health and Social Sciences*. Hoboken, NJ: John Wiley & Sons.
- Fisher, B. J., and Specht, D. K. (2000). Successful aging and creativity in later life. *J. Aging Stud.* 13, 457–472. doi: 10.1016/S0890-4065(99)00021-3
- Frankl, V. E. (1985). *Man's Search for Meaning*. New York, NY: Simon and Schuster.
- Freeman, T. E. (2014). The reconciliation of art and psychology. *Int. J. Art Design Educ.* 33, 184–194. doi: 10.1111/j.1476-8070.2014.01714.x
- Getzels, J. W., and Csikszentmihalyi, M. (1976). *The Creative Vision: A Longitudinal Study of Problem Finding in Art*. Hoboken, NJ: John Wiley & Sons.
- Gruber, H. E. (1980). “The evolving systems approach to creativity “And the bush was not consumed,” in *Toward a Theory of Psychological Development*, eds S. Modgil and C. Modgil (Windsor: NFER Publishing), 209–299.
- Guilford, J. P. (1967). Creativity: yesterday, today and tomorrow. *J. Creat. Behav.* 1, 3–14. doi: 10.1002/j.2162-6057.1967.tb00002.x
- Herzog, A., and House, J. S. (1991). Productive activities and aging well. *Generations* 15, 49–54.
- Howie, L., Coulter, M., and Feldman, S. (2004). Crafting the self: older persons' narratives of occupational identity. *Am. J. Occup. Ther.* 58, 446–454. doi: 10.5014/ajot.58.4.446
- Husserl, E. (1931). *Ideas (WR Boyce Gibson, Trans.)*. London: George Allen & Unwin Ltd.
- Jarzombek, M. (2000). *The Psychologizing of Modernity*. Cambridge: Cambridge University Press.

FUNDING

The collaboration was supported by the Arts and Culture group at Canterbury Christ Church University, as well as the CCCU PPS Incentive Fund.

ACKNOWLEDGMENTS

I would like to thank “Jane” for her work on our artistic collaboration, participating in the interviews, and her comments on the manuscript. Thanks also to my transcribers Josie Hutchins and Marisa Kolovos. Kate Gee, Joe Hinds, and my reviewers provided valuable comments on the manuscript.

- Johnson, J. S., and Wilson, L. E. (2005). “It says you really care”: motivational factors of contemporary female handcrafters. *Cloth. Textiles Res. J.* 23, 115–130. doi: 10.1177/0887302X0502300205
- Lindell, A. K., and Mueller, J. (2011). Can science account for taste? Psychological insights into art appreciation. *J. Cogn. Psychol.* 23, 453–475. doi: 10.1080/20445911.2011.539556
- Locher, P. (2010). “How does a visual artist create an artwork,” in *The Cambridge Handbook of Creativity*, eds J. C. Kaufman and R. J. Sternberg (Cambridge, MA: Cambridge University Press), 131–144.
- Mace, M.-A. (1997). Toward an understanding of creativity through a qualitative appraisal of contemporary art making. *Creat. Res. J.* 10, 265–278. doi: 10.1080/10400419.1997.9651225
- Mace, M.-A., and Ward, T. (2002). Modeling the creative process: a grounded theory analysis of creativity in the domain of art making. *Creat. Res. J.* 14, 179–192. doi: 10.1207/S15326934CRJ1402_5
- Mayer, R. E. (1999). “22 Fifty Years of Creativity Research,” in *Handbook of Creativity*, ed. R. J. Sternberg (New York, NY: Cambridge University Press), 449–460.
- Miall, R. C., and Tchalenko, J. (2001). A painter's eye movements: a study of eye and hand movement during portrait drawing. *Leonardo* 34, 35–40. doi: 10.1162/002409401300052488
- Nelson, B., and Rawlings, D. (2007). Its own reward: a phenomenological study of artistic creativity. *J. Phenomenol. Psychol.* 38, 217–255. doi: 10.1163/156916207X234284
- Perls, F., Hefferline, G., and Goodman, P. (1951). *Gestalt Therapy: Excitement and Growth in the Human Personality*. New York, NY: Dell.
- Perruzza, N., and Kinsella, E. A. (2010). Creative arts occupations in therapeutic practice: a review of the literature. *Br. J. Occup. Ther.* 73, 261–268. doi: 10.4276/030802210X12759925468943
- Reid, K., Flowers, P., and Larkin, M. (2005). Exploring lived experience. *Psychologist* 18, 20–23.
- Reynolds, F., and Prior, S. (2006). The role of art-making in identity maintenance: case studies of people living with cancer. *Eur. J. Cancer Care* 15, 333–341. doi: 10.1111/j.1365-2354.2006.00663.x
- Reynolds, F., and Vivat, B. (2010). Art-making and identity work: a qualitative study of women living with chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *Arts Health* 2, 67–80. doi: 10.1080/17533010903495306
- Reynolds, F., Vivat, B., and Prior, S. (2011). Visual art-making as a resource for living positively with arthritis: an interpretative phenomenological analysis of older women's accounts. *J. Aging Stud.* 25, 328–337. doi: 10.1016/j.jaging.2010.12.001
- Ricoeur, P. (1990). *Oneself as Another. (K. Blamey, Trans.)*. Chicago: University of Chicago Press.
- Santlofer, J. (1993). Lions in winter. *Art News* 92, 86–91.
- Savin-Baden, M., and Fisher, A. (2002). Negotiating “honesties” in the research process. *Br. J. Occup. Ther.* 65, 191–193. doi: 10.1177/030802260206500407

- Shulman, R. (1984). The artist in the slammer: hawthorne, melville, poe and the prison of their times. *Modern Lang. Stud.* 14, 79–88. doi: 10.2307/3194509
- Smith, J. A. (2004). Reflecting on the development of interpretative phenomenological analysis and its contribution to qualitative research in psychology. *Qual. Res. Psychol.* 1, 39–54.
- Smith, J. A., and Osborn, M. (2015). “Interpretative phenomenological analysis,” in *Qualitative Psychology: A Practical Guide to Research Methods*, ed. J. A. Smith (London: SAGE), 25–52.
- Spence, R., and Gwinner, K. (2014). Insider comes out: an artist’s inquiry and narrative about the relationship of art and mental health. *Arts Health* 6, 254–265. doi: 10.1080/17533015.2014.897959
- Tzanidaki, D., and Reynolds, F. (2011). Exploring the meanings of making traditional arts and crafts among older women in Crete, using interpretative phenomenological analysis. *Br. J. Occup. Ther.* 74, 375–382. doi: 10.4276/030802211X13125646370852
- Valle, R. S., King, M., and Halling, S. (1989). “An introduction to existential-phenomenological thought in psychology,” in *Existential-Phenomenological Perspectives in Psychology*, eds R. S. Valle and M. King (New York, NY: Oxford University Press), 3–16. doi: 10.1007/978-1-4615-6989-3_1
- Wallas, G. (1926). *The Art of Thought*. New York, NY: Harcourt Brace and World.
- Weisberg, R. (2004). On structure in the creative process: a quantitative case-study of the creation of Picasso’s Guernica. *Empir. Stud. Arts* 22, 23–54. doi: 10.2190/EH48-K59C-DFRB-LXE7
- Wertz, F. (1983). From everyday to psychological description: analyzing the moments of a qualitative data analysis. *J. Phenomenol. Psychol.* 14, 197–241. doi: 10.1163/156916283X00108
- Worringer, W. (1911). “Entwicklungsgeschichtliches zur modernsten Kunst,” in *Im Kampf um die Kunst: Die Antwort auf den Protest Deutscher Künstler* (Munich: Piper Verlag), 92–99.
- Zimbardo, P., and Boyd, J. (2008). *The Time Paradox: The New Psychology of Time that will Change your Life*. New York, NY: Simon and Schuster.

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 Hocking. 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.



Looking at the Process: Examining Creative and Artistic Thinking in Fashion Designers on a Reality Television Show

Jillian Hogan^{1*}, Kara Murdock¹, Morgan Hamill², Anastasia Lanzara³ and Ellen Winner¹

¹ Department of Psychology, Boston College, Chestnut Hill, MA, United States, ² Department of Psychology, Northeastern University, Boston, MA, United States, ³ Department of Psychology, University of Bath, Bath, United Kingdom

OPEN ACCESS

Edited by:

Kathryn Friedlander,
University of Buckingham,
United Kingdom

Reviewed by:

Jenni Barrett,
University of Central Lancashire,
United Kingdom
Wim Van Den Noortgate,
KU Leuven Kulak, Belgium

*Correspondence:

Jillian Hogan
jillian.hogan@bc.edu

Specialty section:

This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: 15 May 2018

Accepted: 01 October 2018

Published: 23 October 2018

Citation:

Hogan J, Murdock K, Hamill M,
Lanzara A and Winner E (2018)
Looking at the Process: Examining
Creative and Artistic Thinking
in Fashion Designers on a Reality
Television Show.
Front. Psychol. 9:2008.
doi: 10.3389/fpsyg.2018.02008

We examine creativity from a qualitative process rather than a quantitative product perspective. Our focus is on “habits of mind” (thinking dispositions) used during the creative process, and the categories we used were those of the eight Studio Habits of Mind observed in visual arts classrooms (Hetland et al., 2007, 2013). Our source of data was footage from a popular reality television show, *Project Runway*, in which nascent fashion designers are given garment design challenges. An entire season of the show (14 episodes) was transcribed and coded for the presence of eight Studio Habits of Mind. We found abundant evidence of all eight of these thinking dispositions in all portions of the show. We argue that the creative thinking occurring during fashion design bears strong resemblances to that which occurs in the art studio-classroom. Qualitatively created frameworks, like those of the Studio Habits of Mind, can be used to inform our understanding of creative behavior in various disciplines.

Keywords: creative thinking, habits of mind, fashion design, reality television, Project Runway

INTRODUCTION

The current emphasis in creativity research is on what Glăveanu (2014) calls the “quantification of creativity” (p. 22) – the overwhelming proportion of creativity assessment is measured through quantitative psychometrics. But creativity is a complex, multi-sensory, and situation-dependent phenomenon, not easily captured in a numerical value. Here, we argue that creative behavior can and should be examined through a rigorous and systematic qualitative lens during the act of authentic creation. In short, we should be analyzing *processes* of creative thinking and activity, alongside ongoing work in assessing created products. Our view is shaped by conceptions developed by researchers in the field of education, and specifically in the field of primary and secondary school visual art education.

Concepts of Process and Product

The dichotomy between process and product is a familiar one in the field of education (Bruner, 1960; Lachman, 1997; Runco, 2003), and particularly in visual art education (Sullivan, 2001; Gude, 2010; McLennan, 2010). Educators must balance teaching and assessing concrete technical skills, which often lead to polished products, versus teaching and assessing creative thinking potentials, which are often exhibited through exploratory, messy processes, as discussed by Sawyer (2017).

We focus here on the discipline of arts education and argue for a process-based rather than product-based approach to examining creative thinking in the arts. In a product-based view, the artwork is paramount in assessing a student. These works may be assessed on various dimensions – e.g., technique, expression, realism, composition, etc. While this lens offers some information about the student's skills and interests, arts educators have countered that a process-based view is one that provides an alternative lens that is informative in ways that final products cannot capture.

In a process-based view, the final “product” is the artistic mind of the student (Hetland et al., 2007, 2013). The authentic behaviors, motivations, and awareness of various thinking dispositions that are useful in the domain are only accessible through close observation of students at work, or through evidence of their reflection on their making process (through conversation, critique, and written artist statements.) In short, a process-based view is not one that depends solely on any particular tangible artifact that can be ranked, counted, or numerically measured. Rather, it is one that requires attention to the ways a student thinks and how those thoughts form habits of cognition and behavior. These observations and reflections form evidence of thinking in the act of making (or the student's artistic mind).

Like art educators, psychologists have also categorized creative thinking in terms of both product and process. Additional categories include personality and press [or environment], constituting the 4 Ps (Fishkin and Johnson, 1998; Barbot et al., 2011; Said-Metwaly et al., 2017). But in psychology, even those approaches to creativity assessment that are “process” based are essentially dependent upon what educators would think of as products. The most processual approaches are those that aim to measure the cognitive aspects that can eventually lead to creative behavior – most namely, divergent thinking. Process-based approaches include tests like the Remote Associations Test (Mednick and Mednick, 1967), the Structure of the Intellect divergent production tests (Guilford, 1967), the Wallach-Kogan Creativity Tests (Wallach and Kogan, 1965), and the Torrance Tests of Creative Thinking (Torrance et al., 1966). These measures of creativity examine characteristics such as number of ideas, uniqueness, or level of detail in generated drawings, writings, and verbalizations. It is argued that the divergent thinking captured in these tasks is one aspect of the process that can lead to the creation of creative products. However, the quantitative paradigm of psychology's process approach is very different from art education's depictions of process, which focuses more heavily on qualitative data collection, analysis, synthesis, and assessment of individual growth. When we discuss process here, we refer to understandings from the discipline of art education, which we believe can be applied to creativity research at large as a complement to existing approaches.

Current Creativity Approaches

There are no perfect measurements of creativity. However, when process and product, qualitative and quantitative, or subjective and objective measures are combined, each approach complements the other. This is especially true for a construct like

creativity, which is complex (Cropley, 2000; Barbot et al., 2011), ill-defined (Plucker et al., 2004), and changes with historical and/or discipline-based lenses (Hennessey and Amabile, 2010; Barbot et al., 2011). Given the relatively slow progress in the area of creativity assessment in comparison to other areas (Plucker and Makel, 2010), we support the view that varied approaches to assessment allow for methods to be more widely tested and help advance the field (Silvia et al., 2012).

Within the 4 Ps of approaches in psychology (Fishkin and Johnson, 1998; Barbot et al., 2011; Said-Metwaly et al., 2017), each category has benefits and drawbacks.

Psychology's Process Approach

While process-focused psychological approaches described above are generally accepted as reliable (Cropley, 2000), their validity is debated (for discussions of validity, see Hocevar and Bachelor, 1989; Cropley, 2000; Simonton, 2003; Clapham, 2004; Said-Metwaly et al., 2017). This issue is put plainly by Glăveanu (2014, p. 16), who writes: “How is [the] experiential and ontological richness of creativity as a phenomenon ever contained in tasks like ‘please generate as many uses as possible for a brick?’” As Said-Metwaly et al. (2017) note, process-focused approaches (and all other currently accepted approaches) suffer from a limited scope in what they measure; therefore the use of only one approach will fail to capture the complexity of creative behavior.

Psychology's Product Approach

Product-focused approaches are those in which products of a task are assessed using the Consensual Assessment Technique (CAT; Amabile, 1982). In CAT, a social psychological perspective is taken – a team of judges who are experts within the domain independently determine whether and to what degree a product is “creative.” This approach is generally highly reliable and valid (Baer et al., 2004; Kaufman et al., 2007; Said-Metwaly et al., 2017). However, this approach can be time-consuming and expensive, requiring skilled judges. Teams of non-experts do not produce consistent or reliable ratings (Kaufman et al., 2008), and thus findings from CAT depend upon the opinion of experts in the field, which may or may not align with perceptions of the general public or experts from other domains. Because this is a subjective approach, results are limited to the historical and socio-cultural contexts at the time of judging (Amabile, 1982).

Psychology's Personality Approach

Personality-focused approaches constitute the third P. These consist primarily of self-report questionnaires about qualities associated with creative people (i.e., attraction to complexity, high energy, behavioral flexibility, non-conformity, self-esteem, self-acceptance, risk taking, perseverance, introversion, the inclination to connect abstract ideas, and tolerance for ambiguity [Barron and Harrington, 1981; Feist, 1998; Selby et al., 2005; Barbot et al., 2011]) or self-reports of creative accomplishments. Examples of these types of measures include the Creative Personality Adjective Checklist (Gough, 1979), the Creative Perception Inventory (Khatena and Morse, 1994), the Creative Achievement Questionnaire (Carson et al., 2005), and the Runco Ideational Behavior Scale (Runco et al., 2001).

Personality-focused approaches are usually standardized and objectively scored and are accepted as highly reliable (Gough, 1979; Said-Metwaly et al., 2017). Like all self-reports, however, findings are biased by participants' views. Some studies have also shown these measures to lack construct validity (Said-Metwaly et al., 2017). These measures are argued to assess stable traits, which means that this approach does not capture the notion that creativity is something that can be developed (Fishkin and Johnson, 1998). Additionally, Silvia et al. (2012) report that many of these measures result in skewed scores and therefore require careful analysis.

Psychology's Press Approach

The press approach focuses on the environmental factors that come into play when creative behavior is enacted. This is the most historically recent approach to examining creativity assessment and relies on research linking aspects of environmental situations to increased or decreased creativity (Hunter et al., 2007; Hennessey and Amabile, 2010). Like the approach we suggest here, much of the research in this area focuses less on a spirit of *assessment* (connoting ranking, sorting, or other categorizations) and more on *examination* (looking for characteristics), though measures have been created that look for how or less creativity-conducive an environment is or is perceived to be (e.g., KEYS: Assessing the Climate for Creativity, Amabile et al., 1996; the Situation Outlook Questionnaire, Isaksen et al., 2001; and the Virtual Team Creative Climate Instrument, Nemiro, 2001). This approach and these instruments call out for more research, particularly because many are dependent upon subjective judgment.

Here, we take a different approach to what has been discussed. We ask, *how do creative people act and think while engaged in creative behavior?* And *can we systematically capture the thinking dispositions of creative people as we observe them at work?* We believe our method falls outside the scope of the 4 Ps, and acknowledge that, like all current approaches to assessing creativity, this method contains both strengths and limitations. We consider these matters in the Section "Discussion." In arguing that there are observable behaviors that govern creative behavior, we rely on concepts of disciplinary thinking, or habits of mind, which have been developed within the field of education.

Disciplinary Thinking

Teachers who assess children, both summatively (as on a report card) and formatively (as part of ongoing feedback during classroom conversations or contained in notes written on an essay or exam), face the same challenges that psychologists do when evaluating skills (both in creativity and other areas). What precisely should be assessed? A final, tangible product like an artwork, essay, or problem set? Effort, participation, and attitude? The intention behind the work? Or technical skills, like how well one shades color values, recites times tables, or constructs clear prose? If a combination, in what proportion?

Some address these matters by choosing to teach and assess habits of mind within general education (Costa and Kallick, 2008; Ritchhart et al., 2011; Root-Bernstein and Root-Bernstein, 2013), discipline-specific education (Hetland et al., 2007, 2013 [art];

Cuoco et al., 1996 [math], Çalik and Coll, 2012 [science]; Epstein, 2003; Lunney, 2003 [medicine]) and in creativity education (Lucas and Spencer, 2017). In this way, the thinking process (traditionally viewed simply as a means to an end product) becomes the primary evidence of learning (in other words, the product of education). The *process* and *product* become blurred: evidence of the thinking process is used to determine what and how a student has learned or grown. Additionally, teachers consider a student's *personality* or proclivities as part of assessments – if students are naturally inclined to explore, to draw realistically, or to reflect thoughtfully on their process, then teachers may push them harder than they push others in the effort to enhance these inclinations, or to use those strengths as leverage for areas of weakness (Hogan et al., 2018, p. 108–133). These are context dependent judgments, similar to *press* approaches. A teacher knows the time spent creating on a hot Friday afternoon in June will likely yield inferior work to that created on a crisp Tuesday morning in October. The life cycle of the school year, the weather, and special events all play into the ways teachers approach the examination of their students and their thinking and growth. Considered this way, teachers seem to use pieces of each of the 4 Ps, but their wholistic approach cannot easily be captured by the use of any one of these. The approach we describe here is a systematic example of some of the pieces of the assessment process that teachers use in the visual art classroom every day (for examples, see Hogan et al., 2018).

Students can be encouraged to develop thinking dispositions that form part of the creative artistic process. Developing disciplinary thinking in education has been emphasized and described by many (e.g., Gardner, 1999; Lévesque, 2008; Rantala, 2012) and focuses on the processes of thinking authentically in a particular discipline (often through inculcating habits of mind or thinking dispositions). For instance, history teachers can strive to teach students to think like historians and to consider how to make arguments from historical data; and science teachers can encourage students to form testable hypotheses as do scientists.

Studio Thinking

The approach we use here is based on a framework developed by educational and developmental psychologists studying the kind of creative disciplinary thinking developed in studio art classrooms at the high school level (Hetland et al., 2007, 2013). The Studio Thinking framework identified eight habits of mind – broad types of disciplinary thinking – taught in the studio art classroom, as shown in **Table 1** (Hetland et al., 2007, 2013). This framework was developed from the ground up: researchers videotaped, transcribed, and thematically coded utterances of five high school art teachers during many class periods. These teachers were also practicing artists, and taught in arts-centered high schools. The Studio Thinking framework has been adopted by visual arts teachers all over the world, at all levels of primary and secondary school education. Teachers use this framework to teach and assess the thinking processes that students use in their artmaking (Hogan et al., 2018).

The eight Studio Habits of Mind are forms of disciplinary thinking in the visual arts. Habits of mind, or thinking dispositions, encompass not just the skill to complete a task

TABLE 1 | Studio Habits of Mind in the visual art classroom (in alphabetical order; Hogan et al., 2018).

Studio Habit	Sub-habit	Definitions
Develop craft	Technique	Learning to use tools, materials, and artistic conventions
	Studio practice	Taking care of tools, materials, works, and workspace
Engage and Persist	n/a	Finding personally meaningful projects and sticking to them
Envision	n/a	Imagining what cannot be seen and a plan to create artwork of these imagined ideas
Express	n/a	Making works that convey personal meaning
Observe	n/a	Looking closely and noticing what might not ordinarily be seen
Reflect	Question and Explain	Talking about work and working processes
	Evaluate	Talking about what works well, what does not, and why, in works by self and others
Stretch and Explore	n/a	Trying new things, making mistakes, and learning from them
Understand Art Worlds	Domain	Learning what artists have made
	Communities	Learning to collaborate and understanding that artists often work in teams

(*Can* the student do it?), but also the attitudes that interact with those skills (*Will* the student do it? Does the student know *when* and *why* to do it? Perkins et al., 1993; Hetland et al., 2007, 2013; Hogan et al., 2018). If a person uses a habit of mind, this can best be seen through authentic observation of the person working naturally. Only through making artistic decisions independently can a person's motivation, awareness, and other attitudes be observed. In many testing situations, and some teacher-centered environments in education, students are not given the opportunity to make decisions or exhibit the attitudinal aspects of a thinking disposition. Instead, they simply follow directions. Through observation of habits of mind, we look not just for discrete skills but also the attitudes that allow those behaviors to be enacted into the practice of creative work. We consider this to be more ecologically valid – if skills, behaviors, or attitudes are only exhibited at the request of a teacher or tester, they are unlikely to appear organically in another situation. In classroom settings, habits of mind are observable when students are given opportunity to make independent decisions about their work processes and products.

There are eight Studio Habits of Mind: Develop Craft, Engage and Persist, Envision, Express, Observe, Reflect, Stretch and Explore, and Understand Art Worlds. When students *Develop Craft*, they learn techniques, artistic knowledge, and proper tool usage. This Studio Habit also includes setting up one's workspace, caring for materials, and cleaning the studio to be shared by all. *Engage and Persist* can be seen when teachers make sure to allow student interest to play a part in the class, and actively help students recognize what engages them. When students are authentically engaged, persistence through challenges that arise in the artmaking process happens naturally. *Envision* is a synonym for imagine – in art, students use their imagination to create a plan, a vision for their work, to manage their time and predict how long processes will take, and see various possibilities for making changes to their work. Art teachers encourage use of subject matter and media choices, as well the artistic elements and principles to help students *Express* meaning and feeling in their creations. When making art, teachers and students also *Observe* closely – they don't superficially glance at their or others' artworks or at their environment – they notice and look with

sensitivity. *Reflect* most often happens in one of two forms – one is *Evaluate*, in which students comment on their own and others' artworks in terms of what pleases them and what bothers them; the second is *Question and Explain*, which is how teachers encourage metacognition, as students talk about their process, what worked, what didn't, and how they were inspired to make the artwork. Teachers encourage students to *Stretch & Explore* by allowing time for play, discovery, and “mucking around” – sometimes through center-based activities, media explorations, or simply by encouraging a student to go forward with a risky decision about modifying an artwork. The final Studio Habit, *Understand Art Worlds*, is seen when teachers help students to recognize that what they are working on in school connects to what professional artists work on, and to recognize that there is an art world out there in which collaborations of artists, curators, art historians, media, and critics have together shaped the rules and guidelines and canon of the visual art domain.

The Studio Habits of Mind emerged from naturalistic observation of authentic processes of creative making in the classroom. While widely used by arts educators (Hogan et al., 2018), this use of this framework has never been empirically investigated in professional artists. We chose to study this by analyzing the behavior and talk of fledgling fashion designers on the television show, *Project Runway*. This allowed us to capture artists in a naturalistic, creative work environment. This footage was ideal because contestants are constantly required to speak with producers in “confessionals” on camera, and to interact verbally with other contestants, their mentor, and the show's judges. Given our focus on artistic process, influenced from art education, which depends on listening to creators reflect on their work, the reality show setting allowed us to look at patterns of thinking.

The aim of the study reported here is to demonstrate how the Studio Thinking framework can be used as a way to illuminate habits of mind, or thinking dispositions, during creative acts. Unlike any currently accepted approaches, the process identified here has applicability to other domains to help researchers examine what it means to be creative through a lens that is not dependent upon numbers, ranking, or other quantitative paradigms.

MATERIALS AND METHODS

Dataset

Project Runway is an American reality television show that premiered in 2004. The show serves as both a platform to showcase talented up-and-coming fashion designers and as a way to illuminate the intricacies of the design process for viewers. In the words of Heidi Klum, renowned supermodel and the show's host, "we knew that designing is a really creative, interesting, inspiring process, and that it wouldn't be a boring hour of watching people sew" (Mell, 2012).

The show has run for 16 seasons (186 episodes), and six spin-offs have been created, including *Project Runway: All Stars* for returning designers and *Project Runway: Junior* for teen designers. Additionally, 28 international versions exist including *Project Runway Middle East*, *Mission Catwalk* (Jamaica), *Project Runway Philippines*, and *Project Catwalk* (Netherlands; "Project Runway," n.d.). The show's popularity has resulted in 81 Emmy nominations and six wins, including a nomination for Outstanding Reality-Competition Program every year since 2005. The show is immensely popular and reaches viewers not only in the United States, but around the world.

In each episode, designer-contestants compete against one another to create garments for the given challenge of the week. One of the lowest scoring designers is eliminated each week, as determined by three permanent judges from the fashion industry (Klum, fashion magazine *Elle's* editor in chief, Nina Garcia, and American fashion designers Michael Kors [seasons 1–10] and Zac Posen [seasons 11–16]) and one rotating guest celebrity judge. The last remaining three (or sometimes four) designer-contestants are given time and financial resources to design a complete collection to be premiered at Fashion Week in New York City. One final season winner is chosen from these finalists.

Each episode follows a prescribed format: a preparation period (contestants are first assigned a challenge and given time to prepare, sketch, and shop), worktime (contestants spend time constructing in the workroom, seeking feedback from fellow-contestants and mentor and show co-host Tim Gunn), and finally the runway (a presentation and judging of garments on the runway.) Each episode features a unique challenge. Sometimes contestants must collaborate in groups. Other times, challenges constrain the designers, for example to avoid textiles and instead use materials from unexpected locations, such as a flower shop (Season 2), a candy store (Season 4), or a pet store (Season 9; Heching, 2017). *Project Runway* challenges have included avant-garde fashion, toddler wear, dog clothes, outfits for stiltwalkers, professional wrestling outfits, drag costumes, and "everyday woman" challenges which include average people of all shapes and sizes as models.

Coding Manual

We selected Seasons 8 and 9 of *Project Runway* for the development of a coding method and coding manual. These seasons were chosen because they fall at the mid-point of the show's 16 season run. All 28 episodes were transcribed and verbal statements by all persons on the show were coded

using the Studio Habits of Mind framework. Four researchers coded these two seasons using the online coding platform Dedoose.

During coding, a deductive process was used (Crabtree and Miller, 1999), with eight codes reflecting the Studio Habits of Mind (develop craft, express, envision, engage and persist, observe, stretch and explore, reflect, and understand art worlds; Hetland et al., 2007, 2013). Our manual included example behaviors and statements sorted into the appropriate Studio Habit of Mind. The manual included three levels of information: the code label (the Studio Habit of Mind), what the code concerns (a sub-grouping/short definition, based primarily on Hetland et al., 2007, 2013), and a description of what the code sounds like within the context of a *Project Runway* episode (including guidelines for using or not using the code; Boyatzis, 1998; MacQueen et al., 2008). Coding was an iterative process: the P.I. and three coders independently coded transcripts and returned to the group to discuss decisions and the fundamental characteristics of each Studio Habit of Mind as outlined in *Studio Thinking 2: The Real Benefits of Visual Art Education*. This process underwent several rounds of individual coding, followed by group meetings to compare observed behaviors to definitions from Hetland et al. (2007, 2013). While exemplars of behaviors differed between those identified in Hetland et al. (2007, 2013) and what was observed on *Project Runway*, all examples retained the fundamental definitions of each Studio Habit of Mind as defined by Hetland et al. (2007, 2013). Researchers engaged in a process of constant comparison (Glaser, 1965) throughout Seasons 8 and 9 in order to make sure various manifestations of each code were included in the example section of the manual. This process also included periodic checks for inter-rater reliability across coders, discussion of discrepancies, and clarifications to the manual. Additionally, those researchers creating the coding manual engaged in periodic peer debriefing (Lincoln and Guba, 1985) with the fifth research team member.

Data Coding

All 14 episodes of Season 10 of *Project Runway* were selected for analysis using the coding manual developed with Seasons 8 and 9. Episodes each averaged 63 min of content. Three research team members participated in coding of Season 10. These were also transcribed and coded in the online coding platform Dedoose. Nine of the 14 episodes were coded individually by one of three coders (each coder independently coded three episodes). Three were coded by two independent coders (each person in the pair coded separately in order to calculate inter-rater reliability). The pooled Cohen's kappa of these episodes averaged 0.84 which is considered good to excellent agreement (Fleiss, 1971; Cicchetti, 1994; Miles and Huberman, 1994). The last two episodes were coded consensually by the three-person data coding team (these are finale episodes that include an unusual format – visits to the designer-contestants' homes by Tim Gunn and the preparation for and presentation at New York Fashion Week). The decision to code these two episodes consensually was made prior to beginning data coding.

The show's structure switches frequently between two formats: the primary action of the show and confessional-style reflective

interviews with individual contestants. With each switch, a new unit of analysis began. Each particular code could be assigned only once per unit of analysis, but unlimited types of codes could be assigned per unit of analysis. Some units of analysis received no codes because no Studio Habits of Mind were exhibited. During portions of the show that were on the runway, this scheme created units significantly longer than the other two sections. Therefore, for this part of the show, we switched units when a new judge began critiquing a designer. When the designers left the runway, we switched units when the judges began a conversation about a new designer.

RESULTS

To reiterate, our goal was to answer the following two questions: *How do creative people act and think while engaged in creative behavior?* and *Can we systematically capture the thinking dispositions of creative people as we observe them at work?* These questions are not answerable by current approaches to creativity research. We used the Studio Thinking framework, shown to be useful in visual art education, as the framework for systematizing collected data.

The most important finding is that we saw abundant instances of each of the eight Studio Habits in the *Project Runway* episodes. These did not stray from the original definitions and descriptions as put forth in *Studio Thinking* (Hetland et al., 2007, 2013), but examples specific to this fashion design setting do of course differ from those seen in the high school classroom art studio (the ways in which this happened were uncovered and notated within the creation of the coding manual). This translation of the framework to another setting shows that the framework can be used as a lens for looking at creative and artistic behavior outside of the art studio-classroom. In this section, we describe examples of each Studio Habit of Mind displayed on *Project Runway*, in alphabetical order.

Studio Habits of Mind in Fashion Design

Because the habits work in conjunction with one another (Hogan et al., 2018, p. 44), examples described below may demonstrate more than one Studio Habit of Mind. During the coding process, all appropriate codes would have been applied.

Develop Craft

Designers and judges regularly discussed technical abilities of garment construction, and the effect these had on other Studio Habits—like the impact construction mistakes had on being able to express the appropriate feel of the garment, or a mistake being very obvious to an observer. These are the skills of being a fashion designer – choosing fabric, budgeting, constructing and fitting a garment, styling and editing, adding make-up and hair style, and presenting on the runway. Codes for Develop Craft often reflect how designers use these technical skills to make other informed decisions about what their garment will look like, or how they will change it. Without technical skills, a creative vision cannot be achieved. Develop Craft was seen during judging, as shown in this critique of technical skill and styling from judge Michael

Kors in Episode 3: “The skirt was a piece of fabric. It literally, just gathered at the waist. Crooked hem, with that ugly red belt in the wrong place.” Develop Craft was also seen in this critique of fabric selection and compliment of silhouette design in Episode 7:

I think that when we look at, you know, [the garment of designers] Gunnar and Kooan, it could have been a really fabulous gown, but I think they picked the wrong fabric. But do I think it's a great silhouette? Do I think the back of it was really pretty? I like the chiffon. She looked gorgeous. The silver at the neck was fabulous. But I think there were some fabric issues.

In addition to discussion of technical skills during judging, during their worktime the designers discussed the importance of technical skills and the consequence of not having them, as in this excerpt from Episode 10:

[Designer] Melissa: Fabio, my zipper fell off!

[Designer] Fabio: Hold on. Don't—hold on to it. Did you sew the top of it?

Melissa: No, I forgot. This is not good.

Engage and Persist

Designers showed signs of Engage and Persist when they found personal engagement in the work process, became immersed in garment making, buckled down to find solutions to problems, and made compromises for the sake of time management. The most simplistic form of Engage and Persist was when designers displayed satisfaction and focus in their work. In Episode 10, designer Sonjia declares, “I love making over-the-top kind of pieces, so for me, this—this challenge is exciting,” showing her engagement in the work process. On the other hand, in Episode 14, designer Christopher explains his lack of engagement, which affects his work process, “It's so emotional and physically draining. It's just too much to deal with at once.”

This code was also used for instances in which designers specifically mentioned their inspirations, or sources of engagement. For instance, as designer Dmitri introduces his collection at New York Fashion Week in Episode 14, he says, “My inspiration for this collection was organic architecture. I'm proud of what I did. I hope you guys like it.” Other times, this Studio Habit of Mind appears when problems need to be solved, and focus was required, which was often due to the time constraints (most garments must be completed in one day). This is exemplified in mentor and co-host Tim Gunn's signature phrase, “Make it work!”, which refers to making the best of situations, and persisting to complete one's look, even if not to the standard of the designer's original goals. As he tells the designers before departing the work room in Episode 14, “This is about making it work. If there ever were a make it work moment, it is this one. Off we go!” Other times, this is a message specific to design issues, as in this Episode 7 moment as the group departs the work room and heads to the runway:

Tim: Sonjia, why are you freaking out?

Sonjia: I ran out of time, and it's just—

Tim: She looks good.

Sonjia: The hem's not done, and I didn't put enough room for the zippers. She couldn't get into it, and then I had to hand-sew the zipper, and it's just not what I would do, like I-

Tim: That's all right. As long as she can— as long as you fake it on the runway, it's gonna be fine, okay?

Sonjia: Thank you.

Tim: Remember, channel your inner winner, okay?

Envision

Instances in which designers used their imaginations were coded with Envision. These included considering ideas for one's work, or making a plan for reaching those imagined visions. In Episode 10, designer Ven discusses what he imagined for his model's eyeshadow with the makeup designer.

[Designer] Ven: So, this is the fabric [shows the pattern of dress fabric], and I really want the focus to be the eyes.

Make up artist: Start with a highlight, right in the center.

Ven: And then fade it out to a color. Oh, that's perfect.

Ven's conversation shows how the designers often have very precise visions for their work. When working with other designers, hairstylists, or makeup artists, they try to articulate this vision and know whether or not it's been achieved. As Ven comments in Episode 6, "[My model] Terri comes in for the fitting and her hair looks beautiful. It's exactly the direction that I was going for."

These visions also affect the plans that designers must make in order to achieve them. In Episode 11, designer Melissa has to rethink her plans as the challenge includes a last-minute "twist" in which designers must create garments for not only for a child, but also a complementary adult outfit. "I really have to change my course of action. I am going to...cut the white denim into a dress, and do a drape kind of shift dress for the little girl."

Express

Designers regularly used their garments to convey a meaning, feeling, or message. They also used them to express their own personality, style, and individual signature as a designer. This is often articulated in the detailed descriptions of the woman they are theoretically designing for. Sonjia speaks about her muse in Episode 4:

I wanted to create a look for a woman who has a lot going on during the day so she's probably running errands in the morning, in the office during the day and basically something that can take her from wearing her hair up to down to, you know, flats to pumps to basically anything she wants to wear.

Conveying associated moods and feelings are not only part of the designer's process, but also part of the experience of viewing a garment, as often articulated by the judges. In Episode 4, the judges respond to the work of Fabio and Ven.

Michael Kors: The mohair coat's a full flop.

Fabio: Oh.

Michael: I mean, to me, it's a Grandma housecoat. She should have Kleenex in her pocket. I mean, it's just—

Heidi: It just hangs.

Michael: It's sad. What I'm mystified is, where are you in all of this? None of this looks like anything that you would ever touch.

Michael's comment, "where are you in all of this?" refers to the signature styles each designer expresses through their work – so much so, that when something is out of character, like the grandmotherly feel of Fabio's jacket – it is notable.

Observe

When attention was called to something that wouldn't ordinarily be seen, Observe was coded. This sometimes was an observation that came from close inspection (like comparing a garment to something else), or from a critique of something that required careful looking to see. When someone asked to see a garment in a different way (from the back or side, with a jacket removed), this also revealed careful looking and was coded as such. These types of codes appear in this excerpt from a judging session in Episode 4:

Michael Kors: It looks like a hairdressing smock. Like she was cutting her hair, she—you know, there was a fire in the beauty salon, she belted it, and she ran out in her zebra dress, and the whole thing is just weird.

[Guest Judge] Hayden Panettiere: Can you lift up the coral [part of the dress]?

Designer Buffi: Yeah.

Michael: Well, the hem is cuckoo, too.

Reflect

This is the only Studio Habit of Mind which we treated specially due to the fact that our context was reality television footage. Because of the nature of the program, all cast members were constantly put in situations in which they were asked to recall for the camera what had just happened, or the steps of their work process. Therefore, Question and Explain (one portion of Reflect) happened frequently, but artificially due to the nature of the reality television situation. For this reason, we limited Reflect codes to those of the other Reflect sub-habit, Evaluate.

Reflect codes were given for any assessment or critical analysis of one's own or another's work. These occurred in all possible pairings of cast members – designers evaluated each other's work and work process, Tim Gunn and the judges evaluated designer work, and even designers evaluated the judging competency of the judges. In Episode 7, Fabio reflects positively on Dmitry's design, "I like Dmitry's dress because the fit, that is, like, so form-fitting, but at the same time, so effortless" while designer Christopher evaluates the datedness of Sonjia's work negatively, "Sonjia, the 80's called and they want everything back. Cyndi Lauper is missing a dress and a clutch."

Sometimes reflections were given more generally about a designer's relative strengths and weaknesses, or about his or her broad trends in working. In Episode 7, designer Alicia is both complimentary and critical of Christopher's technique use: "Chris does a lot of the same stuff. He does a flowy gown; he does this textile thing, raw-edged silk, and it's cool, but when you keep doing it over and over again, I don't want to see it anymore."

Stretch and Explore

This code was most commonly given when taking risks or breaking out of one's comfort zone was discussed. For instance, in Episode 7, Christopher addresses the critique from Alicia above

about using the same technique multiple times, “Yes, I’ve done this technique for the first challenge, and for the skirt in the *Marie Claire* challenge. It’s kind of getting, you know, old, and it’s a huge risk that I’m taking.” In this statement, he acknowledges that trying new things is part of what is expected of him as a fashion designer. These are expectations that the designers have internalized – Fabio says in Episode 13, “I just hope that [the judges see] that I am pushing myself as a designer, but I’m also pushing the boundaries on design.” The judges and mentors were often coded for encouraging these types of behaviors. In Episode 2, Tim Gunn reiterates this to the designers as he leaves them to work, “I just want to encourage everybody to really push at the boundaries. Wow the judges.” While discussing Fabio’s work in the avant-garde challenge of Episode 12, Michael Kors jokes, “Out of all of our designers, you don’t have to ask him to be avant-garde. He’s playing with proportion. He’s playing with gender roles. I mean, this guy is thinking outside the box.”

Understand Art Worlds

Working with others and having an understanding of the larger domain in which one is working are the two primary tenets of Understand Art Worlds, and both were present in *Project Runway*. This code was given when designers talked about both positive and negative aspects of the unavoidable collaborative process in the real world of fashion design and clothing production. In Episode 3, designer Elena talks about the challenges of working with someone not as skilled: “I’m realizing now that [my partner] is not going to be able to help me with the construction of the dress. She’s moving at a snail’s speed. I’m handling this by working even faster.” Melissa reflects on the help she receives when her zipper unexpectedly breaks in episode 10: “So, Fabio tries helping me; Sonja tries helping me; Christopher tries helping me get this little freaking zipper back on. It’s not happening.”

Understand Art Worlds also encompasses the additional understandings needed as a member of the fashion community. As Elena says in Episode 4, “The fashion industry is a shark. If you can’t handle it, then maybe you shouldn’t be in this industry, because that’s the way it is.” In coding, this included concepts like whether garments are sellable, whether they are constructed properly for their purpose (like toddler-proof child clothing, or bold designs that can be seen from afar on a pop star’s stage costume), and the referencing of famous fashion designers’ previous designs. These codes often appeared in challenges that included prizes that brought designs out to the community – like the department store Lord and Taylor challenge, for which the prize was a contract to have the created garment reproduced and sold in stores. Judge Heidi Klum critiques designer Elena on her garment in Episode 7:

You have to think that you want to sell. I think that this is a very sellable dress. I think that a lot of women are attracted to this kind of silhouette. . . I think it’s a very flirty and fun kind of a dress.

Later in the episode, the judges discuss Melissa’s knowledge of marketability:

Heidi Klum: Melissa did a good job today, you know, which is nice. She’s really cool and edgy. It was nice to see something different.

[Guest Judge and Lord and Taylor representative] Bonnie Brooks: I think it would look great in the window.

Michael Kors: Hers is the most dramatic.

Nina Garcia: It felt very modern. It was dramatic. Yet it’s wearable.

Michael: Listen, this is the most dramatic–Melissa’s– but it’s the toughest, probably, of our favorites to sell.

Bonnie: I think so.

Habit Frequencies

We have shown here ample examples of each of the Studio Habits of Mind in the behavior and talk of fledgling fashion designers. In **Table 2**, we include tallies of each Studio Habit of Mind to show how prolifically each was included in our analysis. Because reality television shows undergo considerable editing, we avoid claims about the proportions of certain Studio Habits or in which sections particular habits appear. We include these numbers simply to show that the instances of codable Studio Habits of Mind talk and behavior were not in any way rare.

DISCUSSION

We have proposed two not commonly used methods for the study of creativity: examining the broad thinking dispositions, or habits of mind, that govern the act of artistic creation (instead of quantitative, product-based measures), and using footage from a reality television show as a source of data.

The Studio Habits of Mind (Hetland et al., 2007, 2013) are widely used in primary and secondary school visual art education. Administrators use them to identify quality arts education, teachers use them to assess their students’ thinking, and students use them as a way to practice metacognition during artmaking (Hogan et al., 2018). We argue here that a habits of mind framework can be used to investigate creative behavior in a variety of settings, and the applicability of the Studio Habits of Mind to the design process illustrated on *Project Runway* is an example of how this can happen.

Through deductive qualitative analysis, we have answered our research questions: *How do creative people act and think while engaged in creative behavior?* and *Can we systematically capture the thinking dispositions of creative people as we observe them at work?* In these examples of fashion designers, we find ample evidence of all Studio Habits of Mind during the work process. The Studio Habits of Mind provide a systematic lens for capturing the thinking behaviors (as evidenced through the spoken words of fashion designers) during the act of garment design. We view this as initial evidence of the validity of this framework for looking at creative behavior, and hope it serves as a catalyst for other creativity researchers to think more deeply about the examination of creators as they work.

Purposes of Assessment Tools

It is important to note that assessments of creativity needn’t always be high-stakes, and we do not suggest that the approach

TABLE 2 | Studio Habits of Mind frequencies.

	Average length	Develop craft	Engage and Persist	Envision	Express	Observe	Reflect	Stretch and Explore	Understand Art Worlds
Total	63 min.	410	747	343	354	283	1137	74	903
Preparation	12 min.	50	134	80	51	9	63	5	190
Worktime	23 min.	149	335	181	111	102	400	24	300
Runway	27 min.	165	156	47	147	151	543	34	253

Preparation, Worktime, and Runway figures reflect episodes 1–12 only; Total figures reflect all 14 episodes of season 10.

articulated here be used alone in high-stakes assessment situations. Some situations require ranking, cut-off scores, or other means of quantitative sorting. But many do not. The approach described here provides an alternate lens for looking at creative behavior – one already shown to be useful for teachers who think about the work processes of their students, and one which could be adopted by creativity researchers as a way to illuminate other parts of the creative process not captured by current measures. For instance, Engage and Persist is not a habit of mind we see encapsulated within current approaches (though these constructs may appear in personality measures, they do not exist in measures within the context of creative behavior), yet anecdotal and historical evidence of highly creative people shows that many creators are extremely persistent and deeply engaged in their processes (Gardner, 1993). Without looking systematically at the behaviors of those who participate in creative acts, how can we know which aspects of creative behavior to choose to measure quantitatively?

Qualitative investigations can help researchers as they develop new objective and quantitative measures more suitable for traditional psychological means. For instance, Hogan et al. (unpublished) have created quantitative measures of some of the Studio Habits of Mind for primary school aged students. And an international research project by the OECD in assessing creative habits of mind (Lucas and Spencer, 2017) has led to the development of a creativity section on the PISA (the international assessment used to compare educational systems) to be administered in 2021 (Lucas, 2017). In both of these examples, qualitative, habit of mind-based approaches have helped to inform and inspire the creation of new quantitative measures.

We believe the adoption of new approaches is particularly important as the ways in which we look at creative behavior continue to expand. For many years, investigations of creativity were grouped into one of two groups: Big C (eminent, domain-changing creativity) and little c (every day acts of creativity). But as Kaufman and Beghetto (2009) suggest, our understandings of creativity can be broadened to include not just famous Big-C creators like Einstein or Picasso, but also categories like Pro-C (professional expertise, like that found in the average office or in the workroom of *Project Runway*) and mini-C (transformative learning, as is found in art classrooms like those in which the Studio Thinking framework was developed). As our classifications of “creativity” continue to expand, the ways in which we examine these behaviors should, as well.

Our Approach and the 4 Ps

We don’t see any of the current approaches described earlier in the paper as ones that can answer our research questions: *how do creative people act and think while engaged in creative behavior?* and *can we systematically capture the thinking dispositions of creative people as we observe them at work?* We do, however, see similarities and differences between our approach and those of some of the 4 Ps. Distinctions of “process” and “product” are blurred when using a disciplinary thinking or habit of mind approach. So while the spirit of looking at the procedures that lead to creative artifacts (or products) is shared between our visual arts education-influenced approach to process and process approaches of psychology, these differ in their qualitative and quantitative approaches.

Our goal to create an ecologically valid, discipline and situation dependent approach shares similarities with ideas put forth by Amabile. The product-based approach of the Consensual Assessment Technique (Amabile, 1982) acknowledges the contextual distinctions of what can be considered creative. Perhaps most similarly, environment (or press) approaches often use frameworks for looking at characteristics of workplaces (Amabile often looks at indicators of sources of motivation by workers, e.g., Hennessey and Amabile, 1988; Amabile et al., 1996; Amabile, 1997), and how those may influence creative behavior. We see our subjective approach as similar, but rather than looking at environment, we focus on evidence of thinking by the creator.

It is possible that some creators would report personality characteristics related to some of the Studio Habits of Mind – like persistence (Engage and Persist), free-thinking (Stretch and Explore), or imagination (Envision). But rather than rely on self-report of general personality characteristics, we think a third-party observer of these thinking dispositions *during* the act of creating is more useful and potentially more reliable.

Limitations

There are several considerations that future researchers should review when applying similar methodologies.

The Relationship Between Artmaking and Creativity

It is important to note that the Studio Habits of Mind emerged in the process of studying artmaking, without specific regard for creativity. Artmaking is not always creative (as in paint-by-number activities, or step-by-step art class activities sometimes used by art educators), and creativity can be found in many domains besides artmaking.

However, the Studio Habits of Mind are related to what is required in creative behaviors, and there is a natural connection

between art and creativity. As Hetland and Winner (2011) point out, creative and artistic thinking dispositions share several qualities: they tie subjects together interpretively (Perkins, 1994; Efland, 2004), allow for adaptive novelty (Perkins, 1981), and are situations in which an individual interacts with a field and a domain (Csikszentmihalyi and Csikszentmihalyi, 1992). Additionally, even a superficial glance at the Studio Habits of Mind suggests connections to lay understandings of creativity. Stretch and Explore includes taking risks and learning from mistakes; committing to solving a problem is exemplified by Engage and Persist; Understand Art Worlds and Observe call for a critical awareness of what's going on around you.

More systematically, we can map aspects of the Studio Habits of Mind onto more formalized definitions of creativity. Consider Guilford's (1967) view of convergent and divergent thinking. Thinking divergently requires a willingness to Stretch and Explore and Envision new possibilities, while convergent thinking requires Understanding Art Worlds (to understand conventions), Develop Craft (to be able to execute those conventions), and Observe (to have an awareness of what's going on around you.) While not precisely the same, we see a clear resemblance between strong artistic thinking and creative thinking.

The Nature of Reality Television and Bias

Reality television footage is not untouched reality – it has gone through many hands in an editing process. The editors of reality shows have considerations beyond showing an authentic work process: they must create enough “drama” to maintain viewership and to properly include reference to sponsoring products or organizations. Each reality show has its own aims, and not all are appropriate as a source of data. However, we believe that in many cases of reality television, the viewer is a witness to the creative process.

Project Runway is particularly notable for minimizing “drama” and keeping the work process at the center. In fact, the show was praised for the authentic way that it uses the television reality contest genre to “engage, enlighten, and inform,” when given a Peabody Award in 2007 (Project Runway, n.d.). Hendershot (2009) also notes this in her analysis:

This is not a series driven first and foremost by character conflict. [*Project Runway*] producers choose to show long sewing sequences in the Parsons School of Design workrooms rather than focusing on personality issues back at the apartments that the designers share. In fact, contestants are only occasionally pictured there. . . Here, if people's issues do come up, it is only a distraction from the work that must be done.

Producers also emphasize that the creative process is at the heart of the show. After noticing that full open calls to find designers meant “too many people were coming in who were clearly less interested in design than they were interested in being on TV,” (Mell, 2012), they cut back to only one or 2 days of open calls in New York and Los Angeles, and now use casting directors across the country to find twenty to thirty contestants for the casting judge panel. When asked if she thought the designers are

their true authentic selves on the show, Desiree Gruber, one of the show's producers, responded:

I wouldn't believe if somebody said they were able to hide their true personality throughout the whole season. It's too stressful. I think one of the reasons the show is so popular is that viewers get into the act of creating along with the designers. We're following people who are authentically very creative; it's not manufactured. They're trying to bring out their best, which is hard to do in a timed experience. Being creative under pressure is not easy (Mell, 2012).

Even in instances when former designer-contestants have complained to popular media outlets about possible injustices regarding predetermined winners or unfavorable editing, they admit that the challenges and work process are very real (Wayne Hughes, 2012; Forbes, 2015; Berman, 2017). In short, while many factors go into the editing of reality television, we feel confident that for the purposes of looking at the creative process over time, footage from *Project Runway* is a useful and valid dataset.

Footage from many kinds of reality shows can provide both researchers and the general public an easily available data source for understanding the creative making process. The popularity of shows like *Project Runway* makes analysis and results accessible to a broader audience – both among researchers and with the general public. In addition, it is a way for researchers to look at creative behavior without the particular limitations of an artificial laboratory setting. Finally, this source of data is widely available and easily accessible.

We acknowledge potential issues of bias as a result of editing for television. However, we believe this issue is minimized because of the research question and the coding methodology applied here. Were this a grounded theory study, in which data was inductively analyzed for the emergence of habits of mind (such as that reported in Hetland et al., 2007, 2013), edited footage would be problematic. We would not know what habits of mind were contained in those parts not deemed worthy of television footage. But since we used a deductive approach, mapping a pre-existing framework on to footage, and since we found evidence of all Studio Habits, we argue that this dataset is useful for the purpose of answering the research questions put forth in this study.

Broader Impact and Future Directions

The ideas put forth here can be useful to two primary groups: teachers and researchers.

Teachers

Many art teachers already regularly use the Studio Habits of Mind in their classroom language, curriculum planning, and assessments. Videos of contemporary artists at work can help illustrate these concretely for students (such as videos on Art21; see Hogan et al., 2018). Excerpts from some reality shows, including *Project Runway*, can also be used to exemplify the Studio Habits of Mind at work in a way that is engaging and relevant to students. While these excerpts should be carefully chosen and screened for school appropriate themes and language, much of what we viewed in our coding procedure could be shown

to students (particularly high school students) to help illustrate Studio Habits of Mind and foster class discussion. Educators in other disciplines have used reality television in the classroom (such as connecting social studies to *The Amazing Race* [Weddell, 2011], using *Undercover Boss* and *Bar Rescue* in management classes [Quain et al., 2018] and using reality shows as a model for designing classroom activities [Bach, 2011] or as models of good and bad teacher behavior for critiques [Higdon, 2008]).

Researchers

People naturally associate artistic endeavors, like fashion design, with creativity. But of course much of creative behavior happens in realms outside of the arts. The Studio Habits of Mind are broad and have potential to be relevant to all domains. This research method need not be limited to a lens for looking at artistic endeavors but can be expanded to look at creativity in domains not traditionally associated with creativity, including cooking (as in Food Network's *Chopped*), tattoo design (exemplified by Spike TV's *Ink Master*), hair design (like Bravo's *Shear Genius*), or even dog grooming (seen on Animal Planet's *Groomer Has It*). Of course, while reality television competition shows are convenience samples, a researcher can also use unedited filmed data or in-person observations.

We propose that the Studio Habits of Mind be used, at minimum, as an initial framework for systematic qualitative analysis of creative behaviors. Use of the framework, while commonly accepted in arts education, should be replicated for

its utility and applicability in other places. We have begun this here with a look at the domain of fashion design. More work is needed to see whether and if so how these habits are used in other professional artistic areas, and/or outside of traditional arts disciplines.

It is possible that for some domains, additional or different habits of mind are more relevant to creative behaviors. In situations in which the Studio Habits of Mind seem inauthentic, we encourage researchers to use a two-part study to examine what habits of mind emerge most frequently in different domains by using the grounded theory approach of Hetland et al. (2007, 2013) and then replicate and examine those habits of mind for validity in different settings, as we have done here.

AUTHOR CONTRIBUTIONS

JH conceived of the study. MH and KM contributed to literature reviews. JH, MH, KM, and AL designed the coding manual and coded the data. JH analyzed the data and wrote the paper with input from KM and EW. EW supervised the study.

ACKNOWLEDGMENTS

We thank the research assistants of the Arts & Mind Lab for their help with transcriptions.

REFERENCES

- Amabile, T. M. (1982). Social psychology of creativity: a consensual assessment technique. *J. Pers. Soc. Psychol.* 43, 997–1013. doi: 10.1037/0022-3514.43.5.997
- Amabile, T. M. (1997). Motivating creativity in organizations: on doing what you love and loving what you do. *Calif. Manag. Rev.* 40, 39–58. doi: 10.2307/41165921
- Amabile, T. M., Conti, R., Coon, H., Lazenby, J., and Herron, M. (1996). Assessing the work environment for creativity. *Acad. Manag. J.* 39, 1154–1184.
- Bach, J. (2011). How teachers negotiate the use of reality television in their pedagogy. *Pedagogies* 6, 144–153. doi: 10.1080/1554480X.2011.554621
- Baer, J., Kaufman, J. C., and Gentile, C. A. (2004). Extension of the consensual assessment technique to nonparallel creative products. *Creat. Res. J.* 16, 113–117. doi: 10.1207/s15326934crj1601_11
- Barbot, B., Besançon, M., and Lubart, T. (2011). Assessing creativity in the classroom. *Open Educ. J.* 4, 58–66. doi: 10.2174/1874920801104010058
- Barron, F., and Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annu. Rev. Psychol.* 32, 439–476. doi: 10.1146/annurev.ps.32.020181.002255
- Berman, N. (2017). *Is it Possible that Project Runway is fake?* Available at: <http://www.tvovermind.com/project-runway/possible-project-runway-fake> [accessed February 22, 2018].
- Boyatzis, R. E. (1998). *Transforming Qualitative Information: Thematic Analysis and Code Development*. Thousand Oaks, CA: Sage.
- Bruner, J. (1960). *The Process of Education: Towards a theory of Instruction*. Cambridge, MA: Harvard University Press.
- Çalik, M., and Coll, R. K. (2012). Investigating socioscientific issues via scientific habits of mind: development and validation of the scientific habits of mind survey. *Int. J. Sci. Educ.* 34, 1909–1930. doi: 10.1080/09500693.2012.685197
- Carson, S. H., Peterson, J. B., and Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creat. Res. J.* 17, 37–50. doi: 10.1207/s15326934crj1701_4
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6, 284–290. doi: 10.1037/1040-3590.6.4.284
- Clapham, M. M. (2004). The convergent validity of the Torrance tests of creative thinking and creativity interest inventories. *Educ. Psychol. Meas.* 64, 828–841. doi: 10.1177/0013164404263883
- Costa, A. L., and Kallick, B. (eds). (2008). *Learning and Leading with Habits of Mind: 16 Essential Characteristics for Success*. Alexandria, VA: ASCD.
- Crabtree, B. F., and Miller, W. L. (eds). (1999). *Doing Qualitative Research*, 2nd Edn. Thousand Oaks, CA: Sage.
- Cropley, A. J. (2000). Defining and measuring creativity: are creativity tests worth using? *Roeper Rev.* 23, 72–79. doi: 10.1080/02783190009554069
- Csikszentmihalyi, M., and Csikszentmihalyi, I. S. (eds). (1992). *Optimal Experience: Psychological Studies of Flow in Consciousness*. New York, NY: Cambridge University Press.
- Cuoco, A., Goldenberg, E. P., and Mark, J. (1996). Habits of mind: an organizing principle for mathematics curricula. *J. Math. Behav.* 15, 375–402. doi: 10.1016/S0732-3123(96)90023-1
- Efland, A. (2004). “Emerging visions of art education,” in *Handbook of Research and Policy in Art Education*, eds E. W. Eisner and M. D. Day (Mahwah, NJ: Lawrence Erlbaum Associates), 691–700.
- Epstein, R. M. (2003). Mindful practice in action (II): cultivating habits of mind. *Fam. Syst. Health* 21, 11–17. doi: 10.1037/h0089495
- Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Pers. Soc. Psychol. Rev.* 2, 290–309. doi: 10.1207/s15327957pspr0204_5
- Fishkin, A. S., and Johnson, A. S. (1998). Who is creative? Identifying children's creative abilities. *Roeper Rev.* 21, 40–46. doi: 10.1080/02783199809553925
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychol. Bull.* 76, 378–382. doi: 10.1037/h0031619
- Forbes, J. (2015). *Former Project Runway Contestant Says the Show is B.S.* Available at: <http://www.thefashionspot.com/buzz-news/latest-news/572481-former-project-runway-contestant-says-show-b-s/> [accessed February 22, 2018].
- Gardner, H. (1993). *Creating Minds: An Anatomy of Creativity Seen through the Lives of Freud, Einstein, Picasso, Stravinsky, Eliot, Graham, and Gandhi*. New York, NY: Basic Books.

- Gardner, H. (1999). *The Disciplined Mind*. New York, NY: Simon & Schuster.
- Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Soc. Prob.* 12, 436–445. doi: 10.2307/798843
- Glăveanu, V. P. (2014). The psychology of creativity: a critical reading. *Creat. Theor. Res. Appl.* 1, 10–21. doi: 10.15290/ctra.2014.01.01.02
- Gough, H. G. (1979). A creative personality scale for the adjective check list. *J. Pers. Soc. Psychol.* 37, 1398–1405. doi: 10.1037/0022-3514.37.8.1398
- Gude, O. (2010). Playing, creativity, possibility. *Art Educ.* 63, 31–37. doi: 10.1080/00043125.2010.11519059
- Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York, NY: McGraw-Hill.
- Heching, D. (2017). *Project Runway Unconventional Challenges, From Worst to Best*. Available at: <http://ew.com/tv/2017/09/15/project-runway-unconventional-challenges/> [accessed February 22, 2018].
- Hendershot, H. (2009). “Belabored reality: making it work on the simple life and project runway,” in *Reality TV: Remaking Television Culture*, eds S. Murray and L. Ouellette (New York, NY: New York University Press), 243–259.
- Hennessey, B. A., and Amabile, T. (2010). Creativity. *Annu. Rev. Psychol.* 61, 569–598. doi: 10.1146/annurev.psych.093008.100416
- Hennessey, B. A., and Amabile, T. M. (1988). The conditions of creativity. *Nat. Creat.* 22, 11–38.
- Hetland, L., and Winner, E. (2011). “The relationship between creativity and studio thinking,” in *The Routledge International Handbook of Creative Learning*, eds J. Sefton-Green, P. Thomson, K. Jones, and L. Bresler (London: Routledge), 226–233.
- Hetland, L., Winner, E., Veenema, S., and Sheridan, K. (2007). *Studio Thinking: The Real Benefits of Visual Arts Education*. New York, NY: Teachers College Press.
- Hetland, L., Winner, E., Veenema, S., and Sheridan, K. (2013). *Studio Thinking 2: The Real Benefits of Visual Arts Education*. New York, NY: Teachers College Press.
- Higdon, M. (2008). From Simon Cowell to Tim Gunn: what reality television can teach us about how to critique our students’ work effectively. *Perspect. Teach. Leg. Res. Writ.* 15, 168–173.
- Hocevar, D., and Bachelor, P. (1989). “A taxonomy and critique of measurements used in the study of creativity,” in *Handbook of Creativity*, eds J. A. Glover, R. R. Ronning, and C. R. Reynolds (New York, NY: Plenum Press), 53–75.
- Hogan, J., Hetland, L., Jaquith, D., and Winner, E. (2018). *Studio Thinking from the Start: The K-8 Art Educator’s Handbook*. New York, NY: Teachers College Press.
- Hunter, S. T., Bedell, K. E., and Mumford, M. D. (2007). Climate for creativity: a quantitative review. *Creat. Res. J.* 19, 69–90. doi: 10.1080/10400410709336883
- Isaksen, S. G., Lauer, K. J., Ekvall, G., and Britz, A. (2001). Perceptions of the best and worst climates for creativity: preliminary validation evidence for the situational outlook questionnaire. *Creat. Res. J.* 13, 171–184. doi: 10.1207/S15326934CRJ1302_5
- Kaufman, J. C., Baer, J., Cole, J. C., and Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creat. Res. J.* 20, 171–178. doi: 10.1080/10400410802059929
- Kaufman, J. C., and Beghetto, R. A. (2009). Beyond big and little: the four c model of creativity. *Rev. Gen. Psychol.* 13, 1–12. doi: 10.1037/a0013688
- Kaufman, J. C., Lee, J., Baer, J., and Lee, S. (2007). Captions, consistency, creativity, and the consensual assessment technique: new evidence of reliability. *Think. Skills Creat.* 2, 96–106. doi: 10.1016/j.tsc.2007.04.002
- Khatena, J., and Morse, D. T. (1994). *Khatena-Morse Multitalent Perception Inventory: Scoring and Administration Guide: Norms-Technical Manual*. Bensenville, IL: Scholastic Testing Service.
- Lachman, S. J. (1997). Learning is a process: toward an improved definition of learning. *J. Psychol.* 131, 477–480. doi: 10.1080/00223989709603535
- Lévesque, S. (2008). *Thinking Historically: Educating Students for the Twenty-First Century*. Toronto: University of Toronto Press.
- Lincoln, Y. S., and Guba, E. G. (1985). *Naturalistic Inquiry*. Thousand Oaks, CA: Sage.
- Lucas, B. (2017). *Why PISA is Moving Towards Creativity*. Available at: <https://www.educationreview.com.au/2017/11/why-pisa-is-moving-towards-creativity/> [accessed September 11, 2018].
- Lucas, B., and Spencer, E. (2017). *Teaching Creative Thinking: Developing Learners who Generate Ideas and Can think Critically*. Carmarthen: Crown House Publishing.
- Lunney, M. (2003). Critical thinking and accuracy of nurses’ diagnoses. *Int. J. Nurs. Knowl.* 14, 96–107.
- MacQueen, K. M., McLellan-Lemal, E., Bartholow, K., and Milstein, B. (2008). “Team-based codebook development: structure, process, and agreement,” in *Handbook for Team-Based Qualitative Research*, eds G. Guest and K. M. MacQueen (Lanham, MD: AltaMira Press), 119–135.
- McLennan, D. M. P. (2010). Process or product? The argument for aesthetic exploration in the early years. *Early Child. Educ. J.* 38, 81–85. doi: 10.1007/s10643-010-0411-3
- Mednick, S. A., and Mednick, M. T. (1967). *Examiner’s Manual, Remote Associates Test: College and Adult forms 1 and 2*. Boston, MA: Houghton Mifflin.
- Mell, E. (2012). *Project Runway: The show that Changed Fashion*. New York, NY: Weinstein Books.
- Miles, M. B., and Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: Sage.
- Nemiro, J. (2001). “Assessing the climate for creativity in virtual teams,” in *Virtual Teams (Advances in Interdisciplinary Studies of Work Teams, 8)*, eds M. Beyerlein, D. Johnson, and S. Beyerlein (Bingley: Emerald Group Publishing Limited).
- Perkins, D. N. (1981). *The Mind’s Best Work*. Cambridge, MA: Harvard University Press.
- Perkins, D. N. (1994). *The Intelligent Eye: Learning by Looking at Art*. Santa Monica, CA: Getty Center for Education in the Arts.
- Perkins, D. N., Jay, E., and Tishman, S. (1993). Beyond abilities: a dispositional theory of thinking. *Merrill Palmer Q.* 39, 1–21.
- Plucker, J., Beghetto, R. A., and Dow, G. (2004). Why isn’t creativity more important to educational psychologists? Potential, pitfalls, and future directions in creativity research. *Educ. Psychol.* 39, 83–96. doi: 10.1207/s15326985sep3902_1
- Plucker, J. A., and Makel, M. C. (2010). “Assessment of creativity,” in *The Cambridge Handbook of Creativity*, eds J. C. Kaufman and R. J. Sternberg (New York, NY: Cambridge University Press), 48–73. doi: 10.1017/CBO9780511763205.005
- Project Runway (n.d.). In *Wikipedia*. Available at: https://en.wikipedia.org/wiki/Project_Runway [accessed February 22, 2018].
- Quain, B., Bokuniewicz, J. F., and Criscione-Naylor, N. M. (2018). The Profit: using reality TV to teach management theories and strategies. *Cogent Educ.* 5, 1–10. doi: 10.1080/2331186X.2018.1444326
- Rantala, J. (2012). How Finnish adolescents understand history: disciplinary thinking in history and its assessment among 16-year-old Finns. *Educ. Sci.* 2, 193–207. doi: 10.3390/educsci2040193
- Ritchhart, R., Church, M., and Morrison, K. (2011). *Making Thinking Visible: How to Promote Engagement, Understanding, and Independence for All Learners*. Malden, MA: John Wiley & Sons.
- Root-Bernstein, R. S., and Root-Bernstein, M. M. (2013). *Sparks of Genius: The Thirteen Thinking Tools of the World’s Most Creative People*. New York, NY: Houghton Mifflin Harcourt.
- Runco, M. A. (2003). *Critical Creative Processes*. Cresskill, NJ: Hampton Press.
- Runco, M. A., Plucker, J. A., and Lim, W. (2001). Development and psychometric integrity of a measure of ideational behavior. *Creat. Res. J.* 13, 393–400. doi: 10.1207/S15326934CRJ1334_16
- Said-Metwaly, S., Van den Noortgate, W., and Kyndt, E. (2017). Approaches to measuring creativity: a systematic literature review. *Creat. Theor. Res. Appl.* 4, 238–275. doi: 10.1515/ctra-2017-0013
- Sawyer, R. K. (2017). Teaching creativity in art and design studio classes: a systematic literature review. *Educ. Res. Rev.* 22, 99–113. doi: 10.1016/j.edurev.2017.07.002
- Selby, E. C., Shaw, E. J., and Houtz, J. C. (2005). The creative personality. *Gift. Child Q.* 49, 300–316. doi: 10.1177/001698620504900404
- Silvia, P. J., Wigert, B., Reiter-Palmon, R., and Kaufman, J. C. (2012). Assessing creativity with self-report scales: a review and empirical evaluation. *Psychol. Aesthet. Creat. Arts* 6, 19–34. doi: 10.1037/a0024071
- Simonton, D. K. (2003). “Expertise, competence, and creative ability: the perplexing complexities,” in *The Psychology of Abilities, Competencies,*

- and Expertise, eds R. J. Sternberg and E. L. Grigorenko (New York, NY: Cambridge University Press), 213–239. doi: 10.1017/CBO9780511615801.010
- Sullivan, G. (2001). Artistic thinking as transcognitive practice: a reconciliation of the process-product dichotomy. *Vis. Arts Res.* 27, 2–12.
- Torrance, E. P., Ball, O. E., and Safer, H. T. (1966). *Torrance Tests of Creative Thinking*. Bensenville, IL: Scholastic Testing Service.
- Wallach, M. A., and Kogan, N. (1965). *Modes of Thinking in Young Children: A Study of the Creativity-Intelligence Distinction* (Oxford, England: Holt, Rinehart & Winston).
- Wayne Hughes, K. (2012). *Catwalk Exposed! Gunnar Deatherage's Project Runway Tell-All*. Available at: <http://www.kellywaynehughes.com/cause-ceacutemiddotlegravemiddotbre/catwalk-exposed-gunnar-deatherages-project-runway-tell-all> [accessed February 22, 2018].
- Weddell, M. S. (2011). Bringing reality TV into the classroom: a semester-long Amazing Race. *SCHOLE J. Leis. Stud. Recreation Educ.* 26, 76–80. doi: 10.1080/1937156X.2011.11949674

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 Hogan, Murdock, Hamill, Lanzara and Winner. 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.



A Decision Tree Based Methodology for Evaluating Creativity in Engineering Design

Trina C. Kershaw^{1*}, Sankha Bhowmick², Carolyn Conner Seepersad³ and Katja Hölttä-Otto⁴

¹ Department of Psychology, University of Massachusetts Dartmouth, Dartmouth, MA, United States, ² Department of Mechanical Engineering, University of Massachusetts Dartmouth, Dartmouth, MA, United States, ³ Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX, United States, ⁴ Design Factory, Department of Mechanical Engineering, Aalto University, Helsinki, Finland

OPEN ACCESS

Edited by:

Ian Hocking,
Canterbury Christ Church University,
United Kingdom

Reviewed by:

Todd Lubart,
Université Paris Descartes, France
Florian Goller,
Universität Wien, Austria

*Correspondence:

Trina C. Kershaw
tkershaw@umassd.edu

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 30 April 2018

Accepted: 08 January 2019

Published: 25 January 2019

Citation:

Kershaw TC, Bhowmick S,
Seepersad CC and Hölttä-Otto K
(2019) A Decision Tree Based
Methodology for Evaluating Creativity
in Engineering Design.
Front. Psychol. 10:32.
doi: 10.3389/fpsyg.2019.00032

Multiple metrics have been proposed to measure the creativity of products, yet there is still a need for effective, reliable methods to assess the originality of new product designs. In the present article we introduce a method to assess the originality of concepts that are produced during idea generation activities within engineering design. This originality scoring method uses a decision tree that is centered around distinguishing design innovations at the system level. We describe the history and the development of our originality scoring method, and provide evidence of its reliability and validity. A full protocol is provided, including training procedures for coders and multiple examples of coded concepts that received different originality scores. We summarize data from over 500 concepts for garbage collection systems that were scored by Kershaw et al. (2015). We then show how the originality scoring method can be applied to a different design problem. Our originality scoring method, the Decision Tree for Originality Assessment in Design (DLOAD), has been a useful tool to identify differences in originality between various cohorts of Mechanical Engineering students. The DLOAD reveals cross-sectional differences in creativity between beginning and advanced students, and shows longitudinal growth in creativity from the beginning to the end of the undergraduate career, thus showing how creativity can be influenced by the curriculum. The DLOAD can be applied to concepts produced using different ideation procedures, including concepts produced both with and without a baseline example product, and concepts produced when individuals are primed to think of different users for their designs. Finally, we show how our the DLOAD compares to other measurements of creativity, such as novelty, fixation, and remoteness of association.

Keywords: creativity, engineering design, decision tree, creative products, creativity measurement, creativity metrics

INTRODUCTION

There are many ways to define creativity (Batey, 2012), but a common definition is that creativity involves the production of ideas that are novel and useful (Sternberg and Lubart, 1999). There are also many ways to narrow the focus of creativity research, but a common framework involves the 4 Ps: person, process, press (environment), and product (Runco, 2004; Cropley et al., 2017; but see

Kozbelt et al., 2010 for 6 Ps, which add persuasion and potential, or Lubart, 2017, who proposes the 7 Cs of creators, creating, collaborations, contexts, creations, consumption, and curricula). Our focus is on the evaluation of the creative product; that is, the outcome of the creative process. According to Plucker and Makel (2010), evaluation of creative products is the “gold standard” of creativity assessment. Evaluation of creative products can take different forms depending on the nature of the product and the way in which creativity is defined. For example, objective scoring is common for divergent thinking test responses, while a panel of expert judges is frequently used to subjectively score artistic (c.f. Getzels and Csikzentmihalyi, 1976) or musical (Beaty et al., 2013) works.

Evaluating Creative Products in Psychology

While objective evaluations of creative products have been utilized in psychology, such as a citation index of composers' works (Hass and Weisberg, 2015), subjective evaluations are far more prevalent. A common method of evaluating creative products within the psychology literature is the consensual assessment technique (CAT; Amabile, 1982). The CAT involves subjective ratings of creative products from a particular domain by a group of people who are knowledgeable within that domain. Amabile (1982) provides specific guidelines for products to be evaluated using the CAT, such as choosing target tasks that are open-ended, allow novel responses, and result in a product that can be judged. Amabile calls for a set of judges with experience in the target domain who rate the resulting products independently, in a random order, and versus each other rather than versus a standard. She also recommends that products are rated on multiple dimensions, such as technical aspects and aesthetic appeal, rather than purely on creativity.

The CAT is a popular method for rating products that have been produced using what has often been called *little-c creativity* (cf. Kozbelt et al., 2010), such as drawings produced by children (Rostan, 2010; Storme et al., 2014) or college students (Dollinger and Shafran, 2005), collages made by children or college students (Amabile, 1982), short stories written by college students (Kaufman et al., 2013), and improvised jazz performances (Beaty et al., 2013). As noted by Baer and McKool (2009), one of the advantages of the CAT is that it can be used in multiple settings because it is not tied to a particular theory. Further, other advantages are that the CAT shows high inter-rater reliability using multiple statistics, including Chronbach's alpha, Spearman-Brown correlations, or intraclass correlations, and that the CAT does not display any differences in ratings obtained related to race, ethnicity, or gender (Baer and McKool, 2009).

Although the CAT has wide application in psychology, there are downsides to its administration. First, the high inter-rater reliabilities that are reported are in part due to using a large number of judges. For example, Amabile (1982) reports using 6–15 judges per study. Although good reliability has been found with as few as three judges (e.g., Rostan, 2010; Beaty et al., 2013), groups of 15 judges (c.f. e.g., Kaufman et al., 2013; Jeffries, 2017) are not uncommon. Inter-rater reliability statistics are influenced

by the number of raters (Gwet, 2014), so it is possible that the agreement levels reported in published research may be inflated.

Second, the CAT requires that selected judges should have experience in the domain that they are judging (Amabile, 1982). As noted by Baer and McKool (2009), this is usually interpreted as a need to have expert judges who can rely on their knowledge of the domain. Finding and compensating appropriate experts is a further strain on researchers. A recent paper questioned the need for expert judges: Kaufman et al. (2013) found that expert judges (professional writers) provided ratings of short stories that were highly correlated with quasi-expert judges (creativity researchers, advanced elementary education or English majors, and English teachers) and moderately correlated with novice judges (college students). This finding may be dependent on domain, however: in a second study, Kaufman et al. (2013) showed that quasi-expert judges (first-year engineering students) and novice judges (students in an introductory psychology courses) did not provide creativity ratings of mousetrap designs that were sufficiently correlated with expert judges' (professional engineers) ratings.

It is possible that greater agreement could be achieved if judges received training, but that would go against another requirement of the CAT. Amabile (1982) specifies that judges should not be trained by researchers to agree with each other, and that they should not receive any definition of creativity. Judges' knowledge of the respective domain should provide enough information for them to know what is creative. This tenet of the CAT has been challenged by two recent papers. Dollinger and Shafran (2005) found that providing non-expert judges (psychology research assistants) with a 4-min review of previously rated drawings boosted their inter-rater agreement with professional artists on ratings of details and overall creativity of drawings, compared to a previous study contrasting untrained non-experts to experts. Storme et al. (2014) contrasted trained novice judges and control novice judges (all students in an introductory psychology course) with expert judges (elementary school art teachers). The trained group was provided with specific definitions of creativity, rated a practice set of drawings, and compared their ratings on the practice set to experts' ratings. On a new set of drawings, the trained group showed a higher level of agreement with the expert judges than the control group.

Overall, while the CAT (Amabile, 1982) has wide use in psychology, and is a successful way to evaluate creative products (cf. Baer and McKool, 2009), there are downsides to its administration, such as the number of judges required, the expertise of the judges, and a requirement that judges should not be trained. While various researchers have developed alternative ways of using the CAT (cf. Dollinger and Shafran, 2005; Kaufman et al., 2013; Storme et al., 2014), there are applications where it has been less useful. For example, Jeffries (2017) reports varying levels of inter-rater reliability depending on the target graphic design task that is used with the CAT. While simpler tasks, such as manipulating text to creatively express one word, had high levels of inter-rater reliability, more complex tasks, such as designing a t-shirt graphic, had unacceptable levels of inter-rater reliability. Further, and more germane to our research, Kaufman et al. (2013, Study 2) questioned the use of the CAT for evaluating what

Cropley and Cropley (2010) refer to as “functional creativity” – the generation of concrete, useful products. Cropley (2015) even goes so far as to suggest that the CAT may be better used to measure creativity of people rather than products. There are, however, alternative methods for evaluating creative products within the engineering literature.

Evaluating Creative Products in Engineering

While creative products generated in engineering settings should meet the common creativity criteria of being novel and useful (Sternberg and Lubart, 1999), it is possible that engineers are generating new ideas in different ways than are typically measured within psychology studies. Cropley et al. (2017) suggest that engineering creativity involves first determining a function and then finding ways, referred to as forms, that this function could be satisfied. While all creative products research goes beyond typical divergent thinking tests, research using creative products in engineering tends to employ different kinds of samples and different modes of evaluation. While some creative product studies in psychology involve the evaluation of products generated by individuals with high domain knowledge, such as advanced students within a field or domain experts (cf. Getzels and Csikzentmihalyi, 1976; Dunbar, 1997; Beaty et al., 2013), many involve products generated by individuals with low domain knowledge, such as children or undergraduates drawn from a research pool (cf. Amabile, 1982; Dollinger and Shafran, 2005; Kaufman et al., 2013, Study 1; Rostan, 2010; Storme et al., 2014). In contrast, research in engineering creativity tends to involve the evaluation of products generated by individuals with high domain knowledge, such as engineering students at various levels (e.g., Charyton et al., 2008; Chan et al., 2011; Youmans, 2011; Oman et al., 2013; Toh and Miller, 2014) or professional engineers (e.g., Jansson and Smith, 1991, Experiment 4; Moreno et al., 2014; Yilmaz et al., 2014).

While much assessment of creative products within psychology has used the CAT, one common form of assessment of creative products within engineering uses several metrics developed by Shah et al. (2003). Shah et al. (2003) propose metrics for the evaluation of novelty (uniqueness of a single idea generated by one person among a given set of ideas generated by many people), variety (number of different ideas generated by one person), quality (feasibility of meeting design specifications by one or more ideas generated by one person), and quantity (all the ideas generated by one person). The novelty metric is similar to the CAT in that it can be used to evaluate the overall creativity of a product. This metric, however, is applied in a very different way than the CAT. The CAT requires the subjective judgment of creativity by a panel of raters, while the novelty metric is a mostly objective determination of the uniqueness of a product within a particular set of products.

The novelty metric is applied by first decomposing a given product into features based on different functions (Shah et al., 2003). For example, if the creative product were an alarm clock, then the features may include the mode of alarm, the display type, the information shown on the clock, and its energy

source (Srivathsavai et al., 2010). Second, product ideas are then described by labeling the expression of each feature. For example, an alarm clock could play a set of songs selected by the user as an alarm, incorporate an LED display, show the time, date, and weather forecast on the clock, and power itself by battery. Third, all described features of a given creative product are compared to the range of features expressed within a set of products. For example, the novelty of a product's mode of alarm is determined by comparing it to the mode of alarm of all other products within the set. If the mode of alarm is highly unique within the set (e.g., waking a user with a mist of water on the face) it receives a higher novelty score than a mode of alarm that is common with the set (e.g., waking a user with music or a beep). Shah et al.'s (2003) novelty metric can express the uniqueness of a particular feature within a set of creative products, or can combine the uniqueness of the features of a creative product to provide an overall measure of a creative product's novelty.

Shah et al.'s (2003) metrics are very popular. A recent search on Google Scholar shows over 750 citations of Shah et al.'s (2003) paper. Despite their frequent use, some limitations to Shah et al.'s (2003) metrics have been expressed. For example, Sarkar and Chakrabarti (2011) critique Shah et al.'s (2003) reliance on uniqueness to measure novelty. Srivathsavai et al. (2010) raise the same criticism, noting that creative products within a particular set are not compared to other sets, and are not compared to current products in the market. As noted by Silvia et al. (2008), this is an issue with all rarity scoring methods: creativity is dependent on sample size (the chance of a rare idea with a smaller sample size is higher). Srivathsavai et al. (2010) also found low correlations between raters, average $r = 0.24$, using Shah et al.'s (2003) novelty metric, which contrasts to Shah et al.'s (2003) reported average $r = 0.62$.

An alternative to Shah et al.'s (2003) novelty metric is Charyton et al.'s (2008) Creative Engineering Design Assessment (CEDA; also see Charyton, 2014). The CEDA is a measurement of creative product design in which participants are asked to create designs that incorporate provided three-dimensional objects, satisfy particular functions (ex. designs that produce sound), list potential users for the resulting creative product(s), and generate alternative uses for their creative product(s). Similar to the CAT, judges rate each participant's resulting creative products for their fluency, flexibility, and originality. For originality, judges are asked to view the product and generate a label that best describes the level of originality, then match that label to the descriptions provided in the CEDA originality metric, which is an 11-point scale that ranges from 0 (dull) to 10 (genius).

Charyton et al. (2008) report high inter-rater reliability, with $r = 0.84$ between two raters, one with a psychology background and one with an engineering background, on the originality scale. In later work, Charyton (2014) reported $r = 0.59$ on the originality scale between five raters, four with an engineering background and one with a psychology background. Neither paper reports the number of creative products that were evaluated to achieve these levels of inter-rater reliability, which calls into question the quality of the scale. In addition, other researchers have had trouble applying the CEDA originality metric to other design problems. As noted by Brown (2014), it may be difficult for judges

to determine which label to choose from the metric, as the labels are open to subjective interpretation. For example, Srivathsavai et al. (2010) found low inter-rater agreement between judges, with an average of $r = 0.35$ for the 11-point scale. They also created modified 3- and 4-point originality scales that kept some of the same labels used in the CEDA rubric. These modified scales did not show improved inter-rater correlations, $r = 0.21$ and $r = 0.29$, respectively, but did show an increase in simple agreement over the 11-point scale (3-point scale = 0.68, 4-point scale = 0.57, 11-point scale = 0.20; it should be noted that there was not a statistically significant difference in simple agreement between the 3- and 4-point scales). Srivathsavai et al.'s (2010) results showing better simple agreement with a smaller set of alternatives is similar to findings showing that higher inter-rater agreement is reached with scales that have fewer intervals (c.f. Friedman and Amoo, 1999).

DEVELOPMENT OF OUR DECISION TREE BASED ORIGINALITY SCORING METRIC

Refining the Originality Metric

Despite the potential risk of insufficient correlations between raters, Srivathsavai et al. (2010) argued that the simple agreement levels of their 3- and 4-point scales, as well as the ability to use the modified CEDA metric to evaluate the originality of a creative product in relation to existing products in the marketplace, justified the use of the scale in further research. In further research from the same group, Genco et al. (2011) used a 5-point version of the modified CEDA originality metric to rate the creativity of alarm clock concepts. Genco et al. (2011) reported a kappa of 0.67 between two raters for 10 concepts, which Landis and Koch (1977) called a substantial level of agreement. Kappa is also considered to be a stricter method of inter-rater agreement than correlations or simple agreement (Cohen, 1968; Gwet, 2014), thus showing the improvement of the 5-point scale over the 3- and 4-point scales used by Srivathsavai et al. (2010). Likewise, this modified 5-point originality metric was used by Johnson et al. (2014) with alarm clocks and with litter collection systems. Johnson et al. (2014) reported kappas of 0.90 and 0.70 for the alarm clock and litter collection system concepts, respectively, with two raters independently scoring approximately 45 of each creative product type. Our group also used the modified originality metric with alarm clocks (Kershaw et al., 2014). We reported a kappa of 0.70 between two raters for 20 concepts. This collection of findings (Genco et al., 2011; Johnson et al., 2014; Kershaw et al., 2014) shows that the 5-point modified CEDA originality metric was successfully used to evaluate creative products for two different design problems produced by students at different levels of the curriculum and from different institutions.

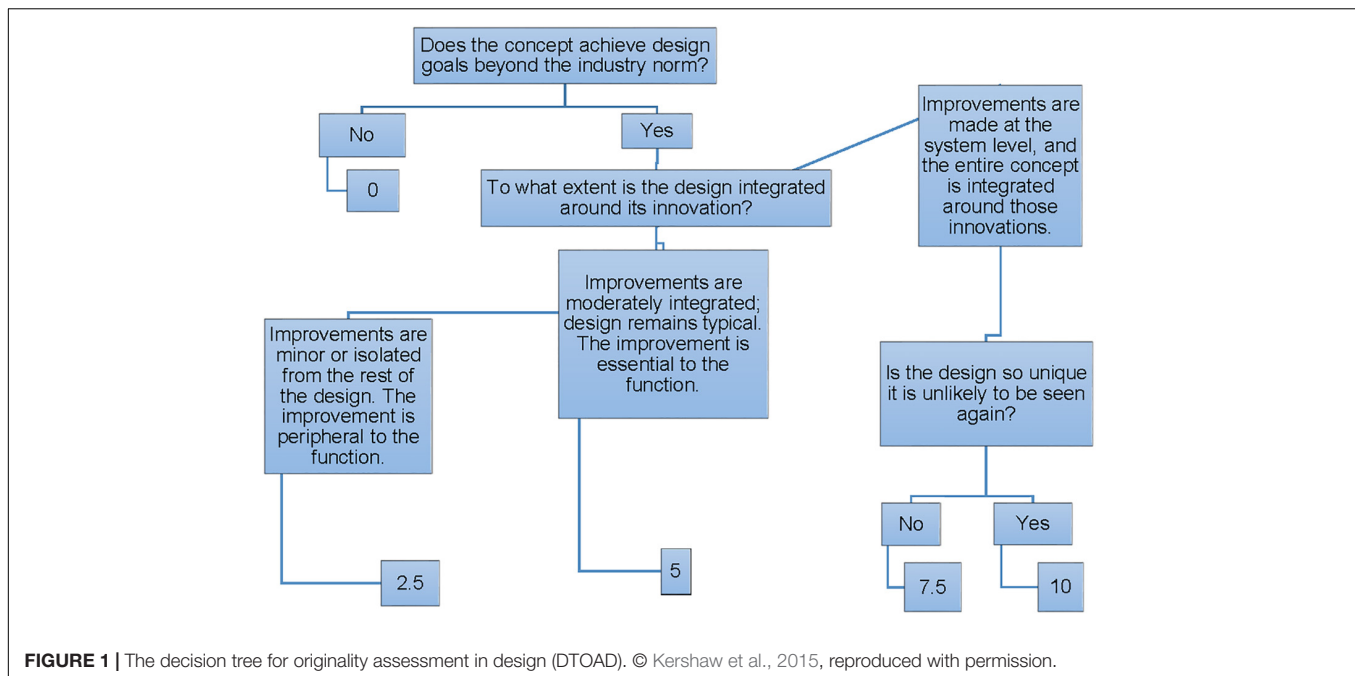
The data reported in this article focus on student-generated concepts for next-generation litter collection systems. While we had success in using the modified 5-point CEDA originality metric to evaluate alarm clock concepts (Kershaw et al., 2014),

we had difficulty in applying it to the litter collection systems. Kappas between the first and third authors, and between two research assistants, remained low ($\kappa = 0.09\text{--}0.42$) despite several rounds of training and discussion. There were several potential reasons for these low levels of agreement. One reason was differences we discovered in the instructions that were given to participants: students at one university were told that the litter collection systems were to be used by volunteers doing highway beautification projects (cf. Johnson et al., 2014), while students at another university were not provided with target users for their concepts. Not being provided with target users led to a wider variety of generated products, some of which did not align well with a previously created list of litter collector features. Another reason for the low levels of agreement could be due to this list of litter collector features and its use in evaluation of the concepts. Prior research showed that evaluating originality based on features rather than the overall concept led to better agreement (Srivathsavai et al., 2010), and thus the feature-based evaluation procedure was followed by Johnson et al. (2014) and Kershaw et al. (2014). While multiple feature categories were generated for the litter collection systems, such as how the device harvests litter (garbage interface), its mobility, how a user triggers garbage collection, its storage components, and the overall architecture of the system, most of the variability in originality scores only came from two features: garbage interface and actuation. We began to question if it was necessary to decompose a product into features and evaluate the originality of each feature, or if we could evaluate creativity more globally.

Development of the Decision Tree for Originality Assessment in Design (DTOAD)

The first two authors, along with two research assistants, made a further modification to the 5-point originality metric by developing a decision tree to aid in the originality evaluation process. Decision trees are a common tool in business, medicine, and machine learning (Goodwin and Wright, 2004) to assist in problem solving. Decision trees are effective for the reasons that diagrams in general are often effective (cf. Larkin and Simon, 1987) – they simplify cognitive operations by providing an external representation of a problem space. Cheng et al. (2001) concur with the cognitive offloading that is afforded by an external representation, and suggest that the most effective diagrams limit the size and complexity of the search that would be necessary to solve a problem or make a decision.

In developing our decision tree for originality assessment in design (DTOAD), we went through several iterations. One of the first versions of the metric focused on how concepts alleviated design flaws. We also originally developed different versions of the decision tree for different types of designs, such as personal litter collectors vs. industrial systems. As noted by Goodwin and Wright (2004), decision tree development is often iterative, just like the development of other types of coding schemes (cf. Chi, 1997). In the end, our final version of the DTOAD incorporated principles from other creative product evaluation methods. First, we kept the 5-point originality metric developed



by Genco et al. (2011) based on its past success in describing student-generated concepts. Second, we applied this metric to evaluating the overall concept, rather than its features, to be more in line with the approach taken by the CAT (Amabile, 1982) and the CEDA (Charyton, 2014), as well as other creative product evaluation metrics like Cropley and Kaufman's (2012) revised Creative Solution Diagnosis Scale. The DTOAD differs from these previous approaches by (a) using a diagram to assist with the originality evaluation process and (b) focusing on how integral design innovations are to the overall concept, rather than parsing a concept into features. The final DTOAD is shown in **Figure 1**. A description of how we train coders to use this protocol and examples of scored concepts at each level of the decision tree follow in the next section.

APPLYING THE DTOAD: FULL PROTOCOL

Training Coders to Use the DTOAD

In applying the DTOAD to the scoring of creative products, we follow several guidelines from the literature about how to train coders. First, anyone evaluating the originality of the creative products must become familiar with the coding scheme and the domain from which the products are drawn. As noted by Chi (1997), having an established scheme that is understood by the coders is necessary before coding begins. To establish this familiarity, the coders review the decision tree (see **Figure 1**) and common features of available products that solve the specified design problem (e.g., the most common features of consumer alarm clocks) and then work together to apply it to a small set of concepts (approximately 10 or so) that have already been scored for originality. The obtained originality scores are then compared

to the scores that were already established, and any discrepancies are discussed, thus following a procedure established by Storme et al. (2014) to provide feedback to coders.

Second, the coders independently rate a set of previously coded concepts for originality, blind to curriculum level or any other conditions. The coders' scores are compared to each other and to established scores. By again providing coders with a schema of how to judge the creativity of the target creative products (cf. Dollinger and Shafran, 2005; Storme et al., 2014). If the coders have reached a sufficient level of inter-rater reliability with the established codes, they are ready to move onto the next step. If not, this process is repeated until a sufficient level of inter-rater reliability is achieved (cf. Chi, 1997). It usually takes coders 2–3 rounds to reach a sufficient level of inter-rater reliability (e.g., Kershaw et al., 2015; Simmons et al., 2018).

At this point, the reader may be asking what a sufficient level of inter-rater reliability is, and how large a sample size must be to reach a sufficient level. Several researchers (e.g., Landis and Koch, 1977; Fleiss, 1981) have published benchmarks for appropriate levels of kappa. Fleiss (1981) called a kappa of above 0.75 "excellent," and Landis and Koch (1977) noted that a kappa between 0.61 and 0.80 was "substantial." Neither Fleiss nor Landis and Koch, however, provide guidelines for the sample size needed to establish a reliable level of kappa. Cantor (1996) suggested a well-known set of guidelines for the necessary sample size, but unfortunately his guidelines (as well as those of Gwet, 2014), are based on having only two coding categories (such as deciding that a product is creative or not). As noted above, we are using a 5-point scale. Thus, to determine a sufficient level of inter-rater reliability, we rely on two guidelines: we make sure to reach a kappa of at least 0.7 to meet Fleiss' (1981) benchmark and we make sure that this kappa is reached through scoring

at least 20% of the sample, a common practice in cognitive psychology (Goldman and Murray, 1992; Nye et al., 1997; Chi et al., 2008, 2018; Braasch et al., 2013; Muldner et al., 2014; Kershaw et al., 2018). In our newest work, we also make sure to report the standard error and the 95% confidence interval so that the precision of our kappa values are known (cf. Gwet, 2014).

Once a sufficient level of inter-rater reliability is achieved between the coders and the established scores, we move to the third step of our training procedure. The coders each independently code a subset of the target creative products, i.e., those products that do not already have established originality ratings. As in the second step, coders are blind to condition when rating the creative products. Like in the second step, we again compare the coders' ratings to see if a sufficient level of inter-rater reliability has been reached. If we have a kappa of at least 0.7, with a low standard error and a confidence interval that only contains acceptable kappa levels, and this level is achieved for at least 20% of the target creative products, then we know that one coder can proceed to code the rest of the set. This coder remains blind to condition as s/he rates the concepts. If we do not have a sufficient level of kappa or have not coded enough creative products, then this process is repeated until we have established inter-rater reliability (cf. Chi, 1997).

Examples of Coded Data at Each Level of Originality

The DTOAD is shown in **Figure 1**. First, a coder must decide if the concept achieves design goals that are beyond the industry norm. That is, does the creative product embody any features or solutions that are different from current market products? Recall that the coders were originally exposed to the basic litter collection products available in the market. If it does not, then the product receives a 0 for originality and the coder stops. This category included two main types of designs: designs that were almost identical to the example provided (for cases where the example was provided) and designs that resembled a product used in the market. For example, **Figure 2** shows an example of a backpack vacuum system. Based on **Figure 2**, it is clear that the student essentially chose a leaf-blower system with a vacuum pump replacing a blower. However, this is an existing product, and thus this concept does not differ significantly from current market products. The student is essentially reproducing prior knowledge.

If the creative product embodies features or solutions that extend beyond current products, then the coder must decide the extent to which the concept is integrated around those innovations. If the nature of the new feature is minor, isolated from the rest of the design, or peripheral to the function of the product, then it would receive a 2.5. For example, **Figure 3** shows a personal litter picker that can extend. The litter picker is identical to the design of a standard picker, except the flexibility to extend or contract the length of the shaft to desired length. This telescoping modification allows for a longer reach when using the product, but otherwise the concept is equivalent to market products. This is not a fundamental design alteration that would be a new mode of litter collection.

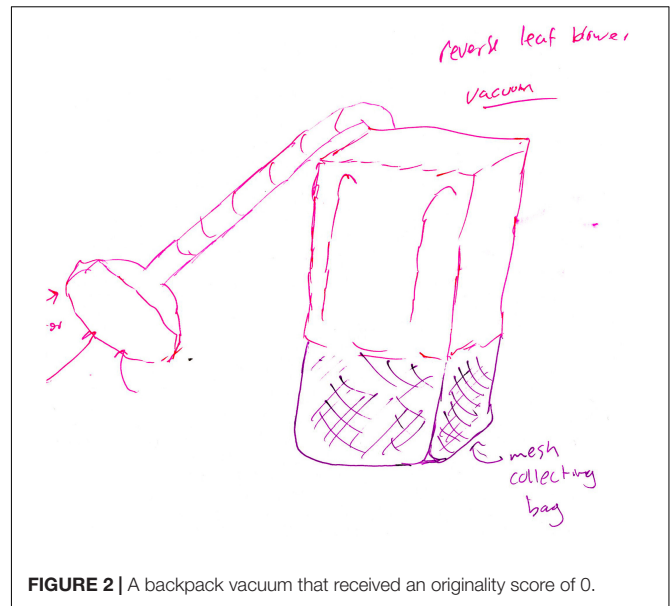


FIGURE 2 | A backpack vacuum that received an originality score of 0.

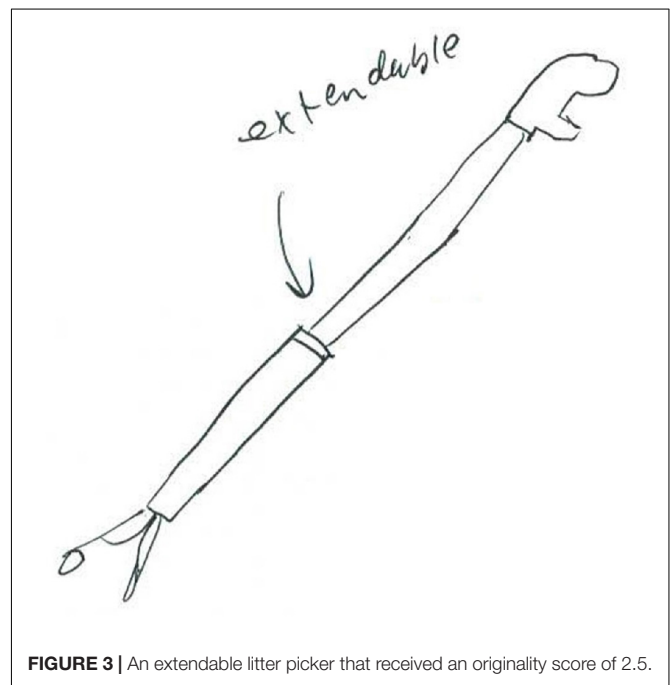


FIGURE 3 | An extendable litter picker that received an originality score of 2.5.

If the new feature entails a moderate level of integration and is essential to the function of the product, yet much of the product's design remains typical, it would receive a 5. For example, **Figure 4** shows several new features that have been incorporated into a garbage truck: it has a vacuum hose that extends from the back, and a means to sort rocks and debris from the trash inside of the truck. The overall architecture of the design is a garbage truck, which is a typical design for a large mobile garbage collection system, however, the atypical placement of the vacuum hose and the internal filtering system show moderate integration with the overall product and are essential to its function, which therefore enhance the overall design.

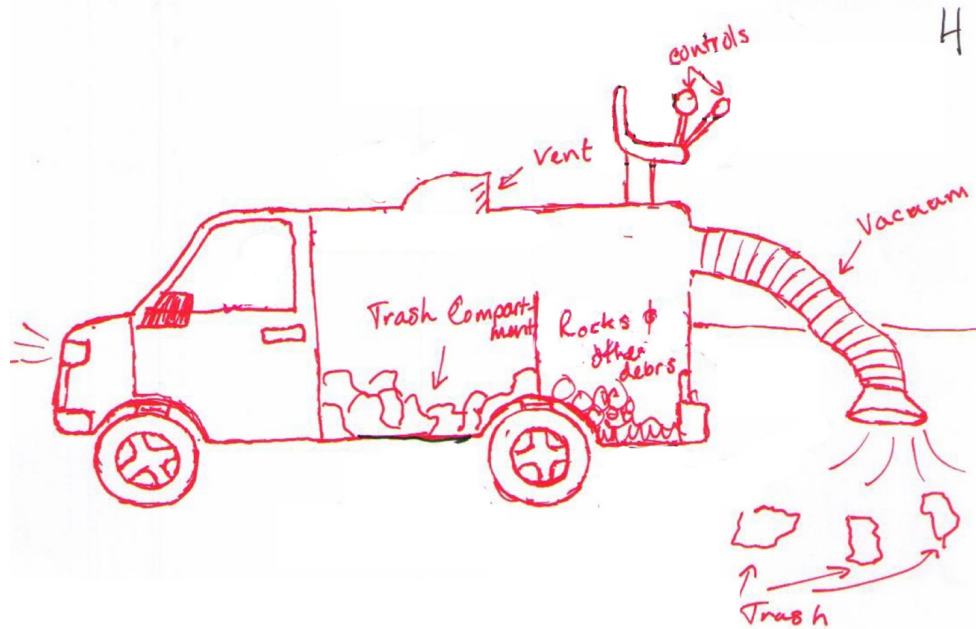


FIGURE 4 | A modified trash truck that received an originality score of 5.

When the new features are at a system-level, and the entire concept is integrated around those innovations, then the creative product can receive a 7.5 or a 10, depending on the likelihood of seeing the product again. For example, **Figure 5** shows a trash collection system that could be used in a neighborhood. Underground tubes carry trash from each home on a street directly to a landfill. This concept displays a unique way of collecting garbage that could be integrated into other infrastructure within a town, such as existing underground water or sewer systems. While this concept shows unique system-level innovations relative to typical litter collection systems, it has appeared several times within our data sets. In contrast, **Figure 6** shows a unique device that collects litter from bodies of water, such as a harbor. This floating drone skims trash from the water and compacts it, and then returns to a docking station to deposit the trash and recharge. This concept requires multiple system-level innovations that are not present in current litter collection systems. While autonomous robotic vacuum cleaners are available on the market, they are generally for in-home use and do not contain a compactor. The device in **Figure 6** is designed specifically for water use, filters trash rather than vacuuming it, and uses geolocation to return to its “home.” We have not seen a comparable design concept among all the concepts we have coded so far.

VALIDATING AND USING THE DTOAD

As described above, and shown through **Figures 2–6**, the DTOAD has primarily been used to evaluate the creativity

of litter collection system concepts. We have also used Shah et al.’s (2003) technical feasibility metric to rate each concept. Technical feasibility has been generally high across concepts (e.g., 9.67 out of 10 for 569 concepts; Kershaw et al., 2015) and we have not found any differences in technical feasibility based on curriculum level (Kershaw et al., 2015) or experimental manipulation (Johnson et al., 2014). Thus, our focus in this paper is on the originality of produced concepts. In applying the DTOAD, we have evaluated undergraduate students across all levels of the mechanical engineering curriculum at the University of Massachusetts Dartmouth and have compared their originality at an overall concept level and at the level of individual contributions to concepts, as well as making cross-sectional comparisons across the curriculum and tracking longitudinal changes in creativity (Kershaw et al., 2015). Much of the previous research summarized in this section was collected using the modified 6-3-5 procedure (Otto and Wood, 2001), in which students are placed in non-interacting groups of approximately 6 individuals. Each student interacts with a sample product (e.g., a personal litter picker) and is asked to generate three ideas. These ideas then circulate through the group so that each student can comment on and modify the ideas of other group members. The ideas circulate through the group five times, or until they come back to the concept originator. While our preliminary work was done following the 6-3-5 procedure, some of our more recent work involved individuals designing on their own with no inputs from others after the preliminary design (cf. LeGendre et al., 2017). The reason for this change in procedure is that Kershaw et al. (2016) found that the concept originator contributed most to the overall originality of

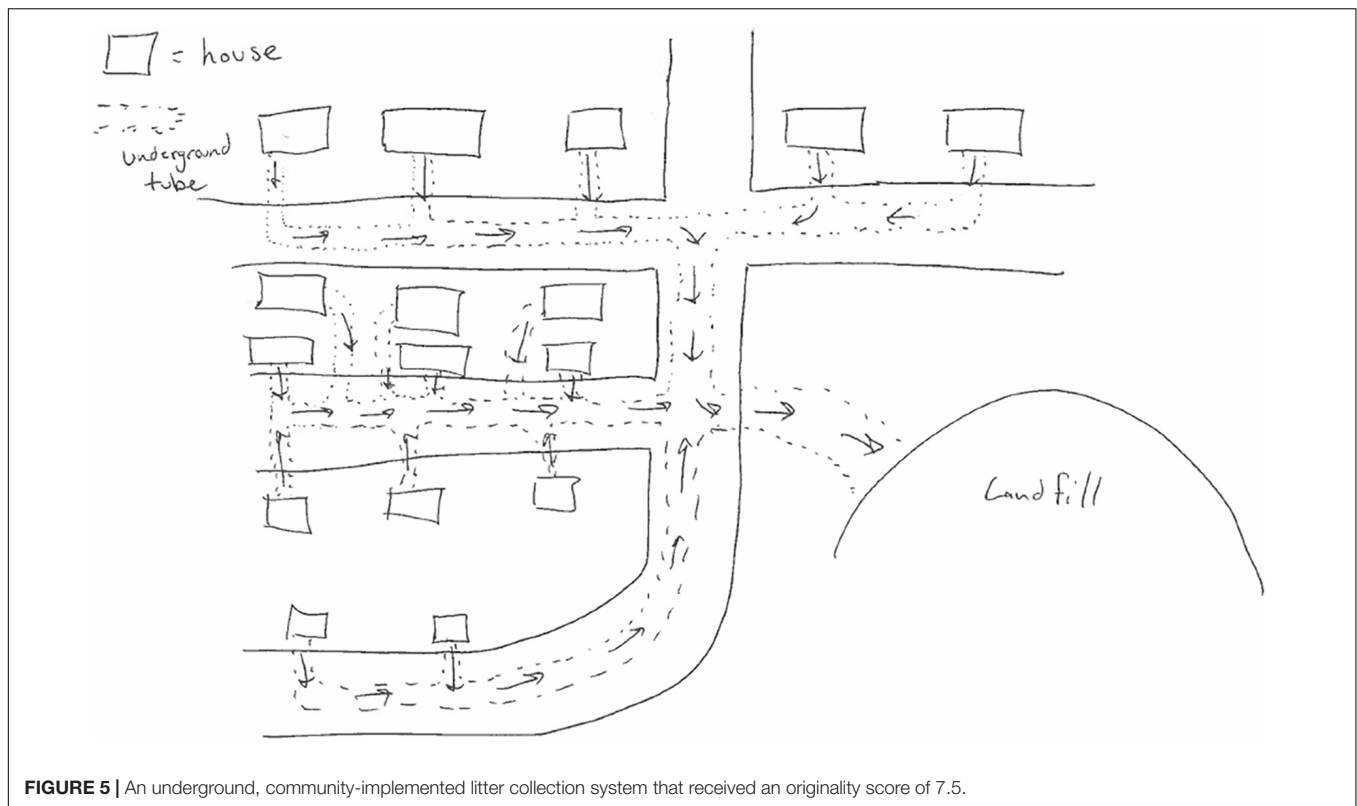


FIGURE 5 | An underground, community-implemented litter collection system that received an originality score of 7.5.

a concept. Further, Simmons et al. (2018) showed that there was no difference in the originality level of concepts produced via the 6-3-5 method and those produced using individual ideation. In the following sections, we summarize the results of our previous work (Kershaw et al., 2015; Kershaw et al., 2016; LeGendre et al., 2017; Simmons et al., 2018), then re-analyze a number of litter collection system concepts to reflect what we have learned. We then apply the DTOAD to a different design problem.

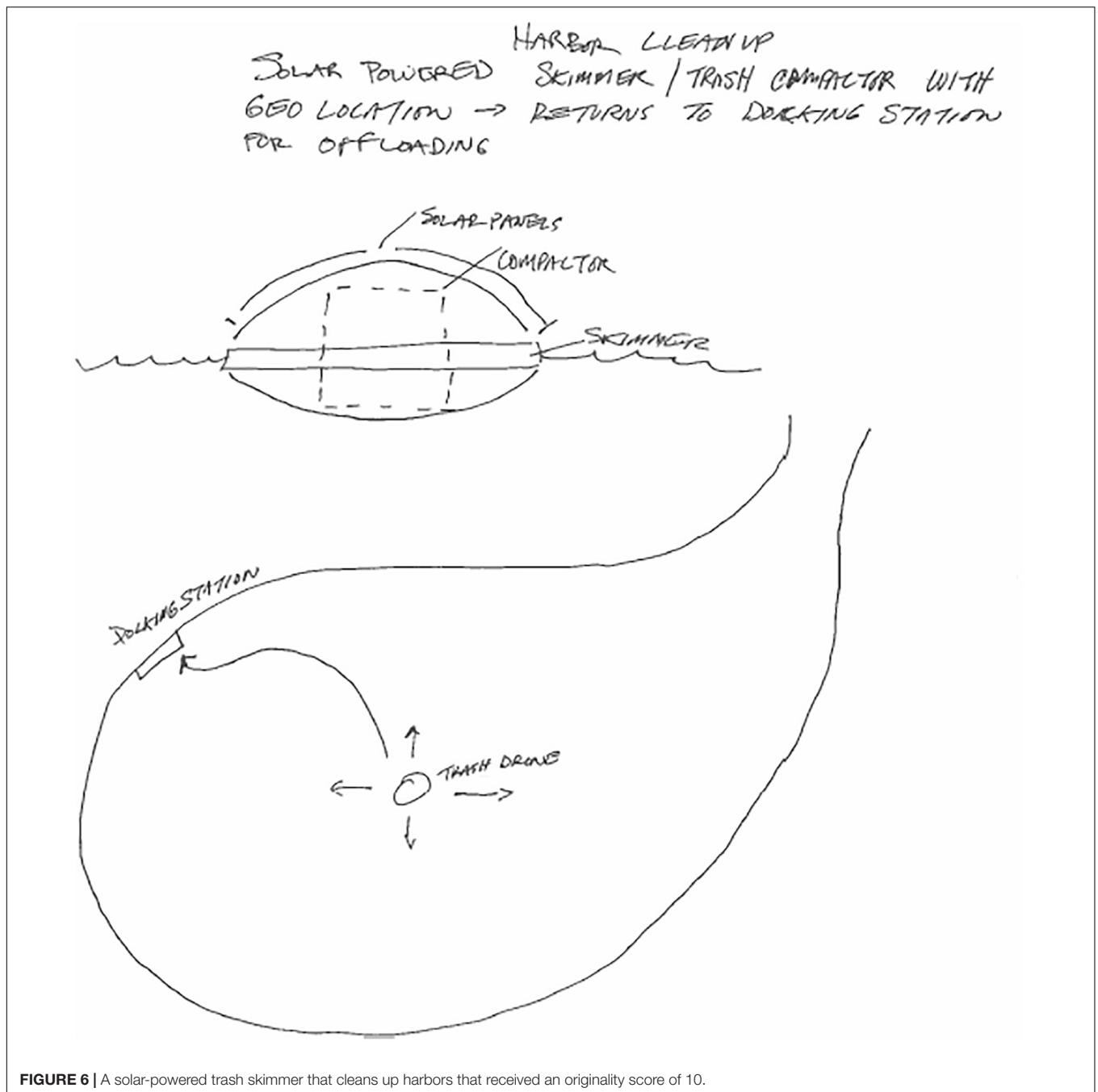
Summary of Previously Published Results

A big part of developing the DTOAD as a firm basis for engineering design creativity coding was to establish inter-rater reliability between coders. As described in Section “Training Coders to Use the DTOAD,” the protocol followed by Kershaw et al. (2015) involved coding of concepts using the DTOAD by multiple coders (three in this case), followed by discussion and clarification to reach convergence. As mentioned above, all coders were blind to condition during the coding process. After each round, the reliability between raters was evaluated using Cohen’s (1968) weighted kappa. Once we achieved a kappa above 0.7, the remaining concepts could be coded reliably (cf. Fleiss, 1981). In Kershaw et al. (2015), 90 of the 569 concepts produced by the participants were coded by three raters, yielding a kappa of 0.73. After this training round, the remaining concepts were coded by a research assistant. This process was then repeated at the individual level for each concept. Each individual’s contribution to each concept, both those that s/he originated and those

that s/he modified through the 6-3-5 process, was scored for originality using the DTOAD. To establish inter-rater reliability, three raters coded the contributions of 35 individuals to 90 concepts, yielding $\kappa = 0.85$. A research assistant then coded the remaining individual contributions.

Our first work using the DTOAD explored engineering creativity across the curriculum (Kershaw et al., 2015). Cross-sectional analysis of results was performed at both the overall and individual level, examining 569 concepts produced by 242 individuals. Our first goal was to ascertain whether we could find any difference in creativity across the curriculum in Mechanical Engineering cross-sectionally. We did not find a significant difference between the 4 years (freshmen, sophomores, juniors, and seniors), either at the concept or individual contribution level. A follow-up analysis comparing extreme groups (freshmen vs. seniors) showed no significant difference between these groups, but some significant differences within the groups, such that seniors tested at the end of the school year had higher originality scores than those tested at the beginning of the school year. There was not a significant change in freshmen’s originality. This pattern in the extreme groups was shown at both the concept and individual contribution level.

In the same paper, another set of analyses assessed longitudinal differences with students who were tested multiple times during the undergraduate curriculum (Kershaw et al., 2015). Specifically, the concept-level and individual-level litter collection system originality scores were compared within a small group ($n = 7$) of students who were juniors during the Fall, 2012 semester and seniors during the 2014 semester. We found that originality



significantly increased from the beginning of the junior year to the end of the senior year without any changes in self-reported GPA or self-reported engineering design self-efficacy. In summary, an improvement in design creativity was observed from the junior to the senior year, with seniors showing some of the highest originality. Although we did not see the cross-sectional results shown in other design problems collected at the same institution (cf. Genco et al., 2011; Kershaw et al., 2014), we were able to establish inter-rater reliability, thus providing some confidence in using the DTOAD for evaluation of engineering creativity.

As mentioned above, the concepts in Kershaw et al. (2015) were collected using the modified 6-3-5 method (Otto and Wood, 2001). In Kershaw et al. (2016), we examined the effects of within-group processes on originality. We used 290 freshman and senior concepts from Kershaw et al. (2015) that received originality scores of 2.5 or higher to examine the weight of each group member's contribution to the originality of a concept. We classified the top scoring contributor of each concept as the originator of the concept, a different group member of the same group, or multiple members of the same group with the same originality score (Kershaw et al., 2016). We found that the

majority of concepts produced (73%) had the concept originator as the top contributor. Further, a comparison of originality scores between these three types of top contributors indicated that groups in which the concept originator was the top contributor had higher originality scores than groups in which a different group member was the top contributor. There were no other significant differences between the contributor types.

While the only significant difference in concept originality in Kershaw et al. (2016) was between the concept originator and a different group member as the top contributor, the large percentage of concepts in which the originator was the top contributor pointed to a possible limitation of group design exercises like the 6-3-5 method. Since the majority of the creativity came from design originators, it is possible to argue that the subsequent contributors fixated on the originator's design and did not contribute anything new. Thus, for our next paper (LeGendre et al., 2017), students generated litter collection system concepts individually, i.e., they completed individual ideation but did not work in groups nor make contributions to other concepts within a group. Further, unlike our previous work, students in LeGendre et al. (2017) did not receive a sample product with which to interact prior to the ideation phase.

Simmons et al. (2018) compared these individually generated concepts from LeGendre et al. (2017) to concepts that were collected using the 6-3-5 method. The first and second author, along with two research assistants, established inter-rater reliability by first reviewing sets of concepts (34 from the group-ideation set and 35 from the individual-ideation set) and then independently coding additional concepts. Thirty (18%) of the group-ideation concepts and 39 (21%) of the individual-ideation concepts were coded by the first and second authors and the research assistants, yielding kappas of 0.79 for the group-ideation concepts and 0.84 for the individual-ideation concepts. The research assistants then coded the remaining concepts. For analysis purposes, only the concept originator scores were used for those concepts collected using the 6-3-5 method. Simmons et al. (2018) found a difference between concepts generated by seniors and freshmen, such that seniors had higher originality scores. They did not, however, find any difference in originality scores between concepts generated through the individual-ideation and group-ideation methods. This result shows us that the DTOAD can be used when concepts are produced by individuals or by groups, and when students are provided with an example product or not prior to ideation. Further, it shows us that similar levels of originality are reached whether an example product is provided or not.

Re-analysis of Litter Collection System Concepts

Over the course of multiple years, we have collected data from over 450 students who have produced over 1000 litter collection system concepts. Our original aim with collecting these concepts was to assess differences in creativity between students at different points in the undergraduate mechanical engineering curriculum. In the following re-analysis, we focus on groups of students who are at opposite ends of their undergraduate careers, and for whom

we have the most data: freshmen and seniors. In selecting the concepts for this re-analysis, we chose concepts produced by students in the spring of their respective year. For the freshmen, this would be the first course that focused on their specific sub-field of mechanical engineering. For the seniors, this would be the last course that is the culmination of their undergraduate training: senior design. Because Kershaw et al. (2016) found that the bulk of the originality score of a given concept came from the concept's originator, we only used concept originator scores for this analysis. Likewise, because Simmons et al. (2018) found no difference in originality between concepts that were produced via the modified 6-3-5 procedure (Otto and Wood, 2001) and concepts that were produced via individual ideation, we include concepts that were produced via both methods.

Based on the above criteria, we selected 420 concepts that were produced by 216 freshmen and 318 concepts that were produced by 141 seniors. These concepts had already been scored for originality and had been part of the analyses in their respective publications (Kershaw et al., 2015; Kershaw et al., 2016; LeGendre et al., 2017; Simmons et al., 2018). An examination of distribution of originality scores led to the removal of two outlying scores, one from the freshman concepts and one from the senior concepts, which were more than three standard deviations above the mean. Thus, 419 freshman concepts and 317 concepts were analyzed. An independent-samples *t*-test indicated that seniors ($M = 2.47$, $SD = 2.42$) produced concepts that were more original than freshmen ($M = 1.77$, $SD = 2.04$), $t(734) = -4.27$, $p < 0.001$, $d = 0.31$. These results support the findings of several other studies that have shown that advanced students display higher levels of creativity than beginning students (Cross et al., 1994; Ball et al., 1997; Atman et al., 1999; Kershaw et al., 2014).

Applying the DTOAD to a New Design Problem

Our summary of previous data and re-analysis of the litter collection system concepts show how the DTOAD metric can be applied to concepts that are produced by individuals with different levels of engineering knowledge (freshmen vs. seniors). We have also shown how the DTOAD can be applied at both the concept and individual level (Kershaw et al., 2015) for group-produced concepts. Further, we have shown how the DTOAD can be applied for concepts produced both within a group setting and individually, with and without an example product (Simmons et al., 2018). All of these applications, however, have been with the litter collection system design problem. In this section, we apply the DTOAD to a different design problem, in which students were asked to generate ideas for next-generation thermometers.

The data in this section were collected as part of a master's thesis (Genco, 2012). Participants, all senior mechanical engineering students, experienced the modified 6-3-5 procedure (Otto and Wood, 2001). All students began by interacting with two sample thermometers, one that measured temperature under the tongue and one that measured temperature by holding it to a person's forehead. All students were given up to 30 min to interact with the thermometers to understand their function. Students in an experimental group interacted with the thermometers while

using devices that were meant to mimic sensory impairments, such as limited vision, hearing, and dexterity. To mimic limited vision, participants wore blindfolds while interacting with the thermometers. To mimic limited hearing, they wore headphones while interacting with the thermometers, and to mimic limited dexterity, they wore oven mitts. Students in a control group simply interacted with the thermometers without using the disabling devices. The experimental conditions were designed to engage the participants in empathic experience design, a structured conceptual design method focusing on stimulating user-centered concept generation by engaging designers in empathic experiences. Empathic experiences are demanding product interaction tasks that simulate actual or situational disabilities experienced by lead users of a product. The goal is to help the designer empathize with these lead users and design products that better meet their needs and requirements. This study focused on evaluating the effectiveness of empathic experience design. Genco et al. (2011) had previously shown that empathic experience design increased the novelty of alarm clock concepts, and Johnson et al. (2014) showed a similar finding for litter collection system concepts.

The first and second authors, along with a research assistant, used the DTOAD metric to code the thermometer concepts, blind to condition. Due to the small number of concepts ($n = 41$), we did not follow our usual procedure of establishing inter-rater reliability and then having one coder complete the originality scoring. Instead, each coder scored all the concepts. Disagreements were resolved and the team decided on a final originality score for each concept.

To make the analysis of the thermometer concepts similar to that of the re-analyzed litter collection system concepts, we used the concept originator's scores in the following analysis. Of the 41 concepts that were coded for originality, only two included group contributions that would have been scored as original beyond the concept originator's idea. An initial examination of the distribution of originality scores indicated one score within the control group that was more than three standard deviations above the mean originality score for this group. After the outlying score was removed, an independent samples *t*-test was conducted, $t(38) = -2.06$, $p < 0.05$, $d = 0.66$. Concepts produced in the empathic experience design groups had higher originality scores ($M = 2.26$, $SD = 1.92$, $n = 21$) than concepts produced in the control groups ($M = 1.18$, $SD = 1.28$, $n = 19$). The results for the thermometer concepts using the DTOAD replicate other creativity results using Shah et al.'s (2003) metric to analyze concepts produced through the empathic experience design procedure (Genco et al., 2011).

ADVANTAGES AND LIMITATIONS OF THE DTOAD

Advantages

There are several advantages to the DTOAD. First, it is a reliable instrument for the measurement of creativity, as shown through the high levels of inter-rater agreement reached between coders (see Summary of Previously Published Results). The training

process we follow with our coders (see Examples of Coded Data at Each Level of Originality) allows them to recognize original creative products. Second, the DTOAD shows a high degree of construct validity. It shows convergent validity (cf. Cronbach and Meehl, 1955) with other evaluation instruments of creative products: fixation scores (Jansson and Smith, 1991; Vasconcelos and Crilly, 2016) and Shah et al.'s (2003) novelty metric. The DTOAD shows discriminant validity with other measures of creativity, such as the Remote Associates Test (RAT) (Mednick, 1962; Smith and Blankenship, 1991).

Relationship to Fixation

LeGendre et al. (2017) examined the relationship between fixation and originality within the litter collection system concepts that were reported in Kershaw et al. (2015). The fixation metric measured the presence or absence of each repeated feature of the example product (see **Table 1**), following the procedure of Jansson and Smith (1991). This replicated features measure of fixation is common in the literature: over half of the studies included in Vasconcelos and Crilly's (2016) meta-analysis used a replicated features measure of fixation. LeGendre et al. (2017) found a significant relationship between fixation and originality, $r(729) = -0.21$, $p < 0.001$. Using a new set of litter collection system concepts, Simmons et al. (2018) found a similar negative relationship between fixation and originality, $r(243) = -0.32$, $p < 0.001$. For example, **Figure 3** shows a litter picker that replicates four features of the provided example: a pistol trigger, an unbroken long rod, a prong quantity of two, and a prong end. As noted in Section "Examples of Coded Data at Each Level of Originality," this concept received an originality score of 2.5, thus illustrating the negative relationship that LeGendre et al. (2017) and Simmons et al. (2018) found between fixation and originality. It is important to note, however, that Simmons et al.'s (2008) results were only shown when participants were provided with an example litter collector to interact with prior to ideation. When no example litter collector was provided, there was no longer a significant relationship between fixation and originality ($r[154] = -0.01$, $p = 0.44$).

The negative relationship between fixation and originality found by LeGendre et al. (2017) and Simmons et al. (2018) is expected given the nature of these measures. Creative products that are deemed original should not show a high degree of design

TABLE 1 | Feature descriptions for fixation coding.

Design feature	Description
Pistol Trigger	Any trigger with a handle and pull mechanism resembling a square and a line.
Unbroken Long Rod	A fixed length that cannot be changed which connects the trigger to actuator.
Prong Quantity	Any two or three component grabbing structure acting as the picker, i.e., claws, cups, or plates.
Prong End	Any shape or line at the end of a prong.
Hand Support	Any small shape that is connected to the hand grip in order to add ergonomic support.
Locking Mechanism	Any indication of a locking mechanism, i.e., text or shape similar to the lock on the example.

fixation. These fixation results show convergent validity between the DTOAD and common ways (cf. Vasconcelos and Crilly, 2016) of measuring fixation.

Relationship to Novelty

Additional support for the construct validity of our conception of originality is shown through convergent validity with Shah et al.'s (2003) novelty metric. The first and fourth author chose a subset of 185 freshman, junior, and senior concepts from the set coded for originality by Kershaw et al. (2015). Following Shah et al.'s (2003) guidelines and the procedures used by Srivathsava et al. (2010), we first decomposed the litter collection systems into features based on functions including the means of collecting trash (garbage interface), mobility of the system, and its actuation (trigger to collect garbage; see **Table 2** for all features). Next, we developed labels of the expression of each feature based upon what was present in the dataset. For example, the possibilities for trash treatment within our sample included that garbage was stored within the system, that there was separate storage, or that garbage was burned, compacted, recycled, or ground. We also included a label of “none” for when a means of trash treatment was not included within a concept, and a label of “not clear” for

when it was impossible to determine the means of trash treatment for a given concept (see **Table 2** for all expressions within each feature). After developing a final set of features and expressions within those features, the first and fourth author coded the chosen concepts by describing the expression of each feature within each concept. For example, **Figure 7** shows a litter collection system that is a litter picker (architecture) with a claw that collects trash (garbage interface) when a human (control) squeezes its handles (actuation) via manual power. This litter collector is carried (mobility) by a person but no modifications were made to this design in consideration of its intended user. This concept does not include any means for trash treatment or removal within it.

After all the selected litter collection system concepts were coded, we compared all described features of a given creative product to the range of features expressed within a set of products to determine its novelty score. Shah et al.'s (2003) novelty metric can be used to measure the uniqueness of a particular feature within a set of creative products, or can provide overall measures of the novelty of a creative product by averaging the uniqueness of all features (average novelty) or choosing the highest novelty score of a feature from each concept (maximum novelty). For the purposes of comparing novelty to originality,

TABLE 2 | Litter collection system features and their expressions for novelty coding.

Feature	Garbage interface	Mobility	Actuation	Trash treatment	Trash removal	Power source	Control	User considerations	Architecture
Expressions	Vacuum	Carried	Button	Stored (in device itself)	Removable bag	Manual	Human	Personal item storage	Standard design
	Suction	Worn	Switch	Separate storage	Door access	Human, power stored	Automated	Safety cabin	Novelty shape
	Shovel	Rolling robot	Lever(s)	Burned	Zipper to access	Hydropower	None	Body strap	Extends
	Claw	Vehicle	Squeeze handle(s)	Compacted	Vibrate to shake trash off picker	Pneumatic	Not clear	Easier handles	Body extension
	Reverse claw	Pulled by vehicle	Trigger	Recycled	Release button	Battery (cordless)		Light weight	Folds for storage
	Brushes rolling	Push cart	Continuous	Grind	Push trash off collector	Electric (corded)		One-handed squeeze	Multiple effectors
	Water flow	Pushed or pulled sweeper	Manual motion	None	Dump	Gasoline		Mechanical assist	Modular system
	Spear	Fly	Senses trash	Not clear	Not clear	Solar		Padding	Distributed system
	Conveyor belt	Stationary	None		None	Wind		Grip	Integrated system
	Sticky pad	Not clear	Not clear			Fuel cell		Vision	Infrastructure
	Magnet					Biofuel		Weight distribution	
	Laser					Nuclear		Customized	
	Flap					Not clear		None	
	Net								
	Human hand								
	Robot hand								
	Animal								
	Not clear								

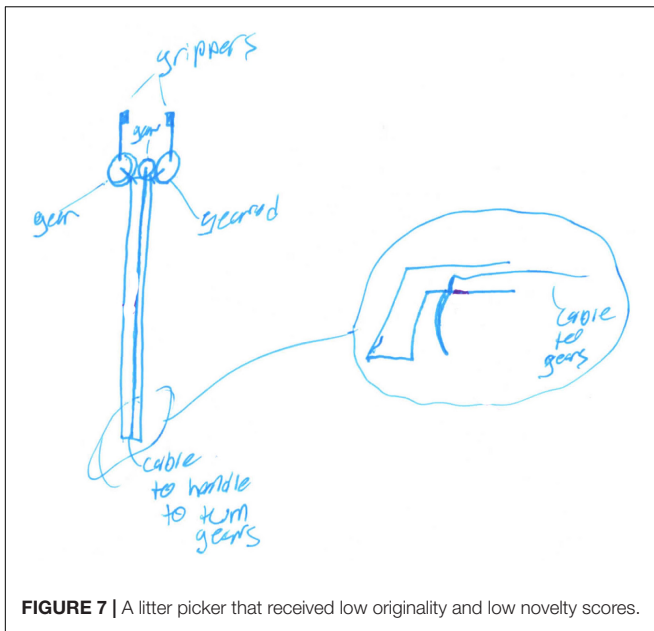


FIGURE 7 | A litter picker that received low originality and low novelty scores.

we chose the maximum novelty measurement to ensure that creative designs were not stifled by containing standard features, and because the DTOAD considers the integration of innovative features beyond the industry norm. We found a significant positive correlation between originality and maximum novelty, $r(185) = 0.35$, $p < 0.001$. This positive relationship would be expected given that both the DTOAD and Shah et al.'s (2003) novelty metric are designed to assess the creativity of ideas. At the same time, it is important to note that this correlation is moderate – the two metrics are not measuring creativity in the same way.

Figure 7 shows an instance when the originality and novelty metrics agree – this concept has low maximum novelty (6.47 out of 10), and a score of 2.5 on the DTOAD for minor improvements to a function of a typical litter picker. **Figure 8** shows another instance of agreement between the metrics. This concept received an originality score of 7.5 using the DTOAD metric because it shows system-level integration of multiple features, including the use of water currents for powering the device and enabling filtration. Within the set of concepts chosen to measure novelty, the water wheel filtration system shown in **Figure 8** contained five rare features, including the use of hydropower, being a stationary system, and its overall atypical architecture, thus boosting its novelty score to 9.88. In contrast, **Figure 9** shows an instance of disagreement between the two metrics. Any disagreements we found between the metrics occurred when the novelty metric indicated that a concept was unique and the DTOAD did not. The reverse circumstance did not occur. For example, the concept shown in **Figure 9** received a 2.5 using the DTOAD because it only displays a small modification of using suction instead of a claw to collect trash within a typical litter picker design. In contrast, the concept shown in **Figure 9** had a high novelty score (9.88) because the use of suction within the garbage interface feature was rare. As noted in Section

“Evaluating Creative Products in Engineering,” Shah et al.'s (2003) novelty metric relies on novelty within a set of creative products (Sarkar and Chakrabarti, 2011) and does not compare creative products within a set to other sets or to current market products (Srivathsavai et al., 2010). Thus, the DTOAD may provide a truer evaluation of the creativity of a design by comparing it to a large set of related designs that are not present within a given set of ideas.

Relationship to Remote Association

The DTOAD shows convergent validity with fixation and novelty. These are common ways to assess creativity, but another way to conceive of creativity is through the lens of convergent and divergent thinking (cf. Guilford, 1956; Cropley, 2006). Convergent thinking can be defined as using conventional and logical search strategies to arrive at solutions. While an individual may consider many options, a single solution is usually chosen. In contrast, divergent thinking can be defined as using unconventional and flexible thinking to arrive at solutions. Divergent thinking frequently leads to the production of multiple solutions, or multiple perspectives on a situation or problem. One aspect of divergent thinking is how well individuals can make connections between disparate ideas. It is hypothesized that individuals who are more creative have less steep association hierarchies – that is, concepts in long-term memory are less strongly related than for individuals who are less creative (cf. Mednick, 1962). Having weaker association hierarchies increases the likelihood that individuals will generate novel responses. This aspect of divergent thinking could be useful in creative design because it would allow a person to be more flexible when generating ideas for a creative product. Nijstad et al. (2010) argued that the originality of ideas is highly related to the level of flexibility a person shows by exploring many options during the ideation process. It is important to note, however, that both convergent and divergent thinking contribute to the production of original ideas. Variability alone is necessary, but not sufficient for creativity – convergent thinking is needed to evaluate generated ideas (cf. Cropley, 2006). Well-known models of creativity account for both convergent and divergent thinking, such as Campbell's (1960) blind variation and selective retention model of creativity (see also Simonton, 2011) or the creative problem-solving framework that is used in educational settings (Treffinger et al., 2006).

To understand the relationship between originality and remoteness of association, we used the RAT, a traditional psychometric creativity instrument. The RAT utilizes both divergent thinking (to explore connections between concepts) and convergent thinking (to choose the most appropriate connection, or answer to the RAT problem). Individuals commonly generate multiple possible connections between the words in a given RAT problem (divergent thinking) before choosing the best solution to the problem (convergent thinking; Wieth and Francis, 2018). Twelve RAT items were chosen from Mednick (1962) and Smith and Blankenship (1991). The RAT asks participants to generate a fourth word that forms a phrase with each of three provided words. For example, if the provided words were *blue*, *cake*, and *cottage*, a correct generated

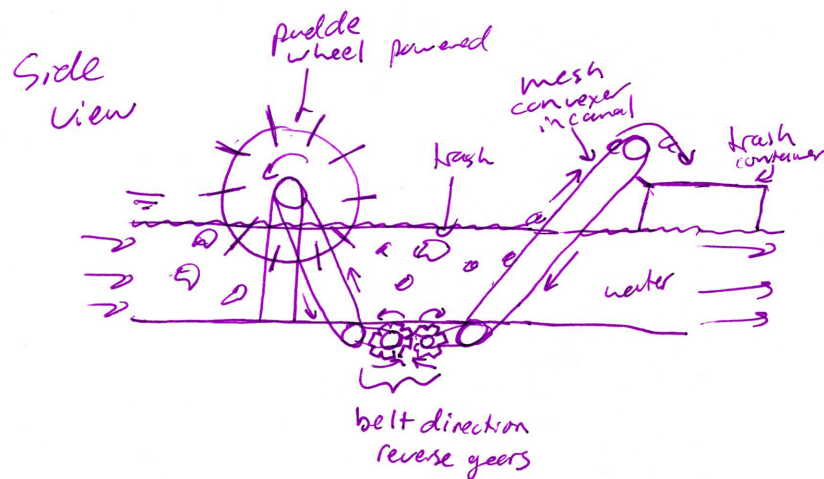


FIGURE 8 | A paddle wheel and conveyor water-based trash collection system that received high originality and high novelty scores.

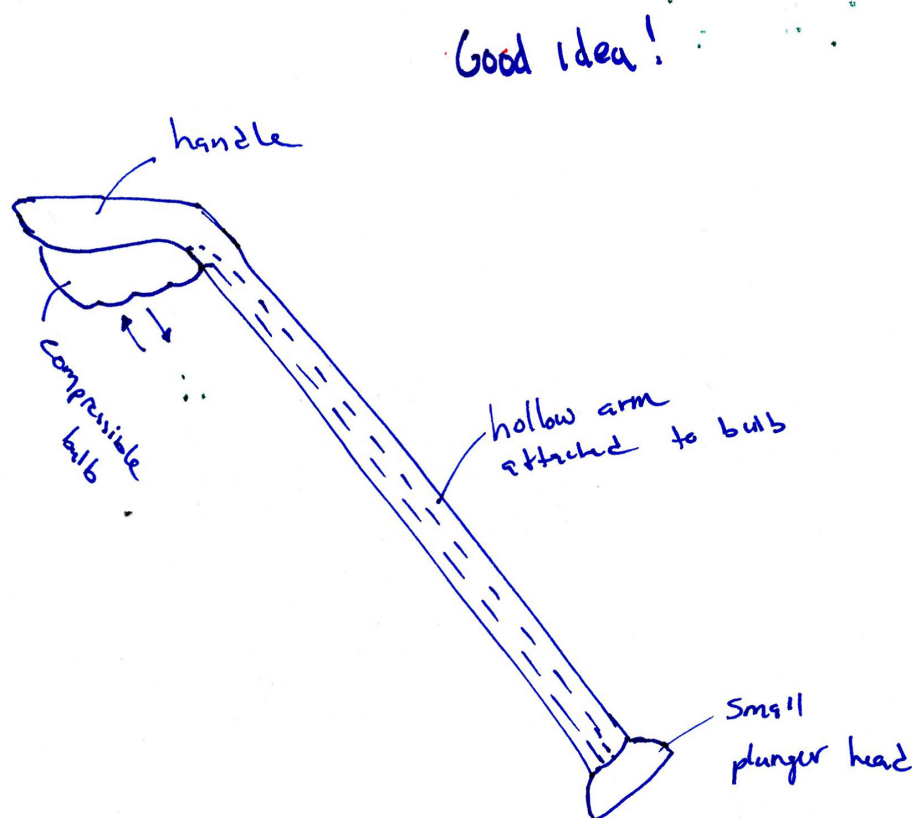


FIGURE 9 | A picker with a suction cup end that received a low originality score but a high novelty score.

answer would be *cheese* (*blue cheese, cheesecake, cottage cheese*). The RAT was administered to a subset of senior mechanical engineering students who generated the concepts that were part

of Kershaw et al. (2015). We correlated RAT scores with the average originality across the litter collection concepts these students produced. There was not a significant relationship,

$r(23) = -0.13, p = 0.55$. This result within the litter collection system concepts replicated our earlier findings that RAT scores did not predict originality for alarm clock concepts (Kershaw et al., 2014). This result also supports the findings of Kudrowitz et al. (2016), who found no relationship between the RAT and performance on creative design tasks. Remoteness of association and ideation in engineering design involve both convergent and divergent thinking ability (cf. Jaarsveld and Lachmann, 2017), but the RAT appears to be evaluating a different aspect of creativity than the ability to generate original ideas via creative products, as measured through the DTOAD (cf. Cropley et al., 2017, for a similar discussion).

Limitations

There are several limitations of the DTOAD as presented in this paper. First, the DTOAD is most appropriate for design problems or applications for which closely related products or solutions are available in the marketplace. These existing solutions serve as benchmarks or anchors for determining whether the newly proposed solutions are different in some way from those benchmark products. On the positive side, it is difficult to identify a design problem for which no benchmark solutions exist. Even products considered revolutionary upon introduction to the marketplace, such as the first smartphones, replaced or augmented existing products performing similar functions, e.g., larger laptop computers. The difficulty is that a thorough and relevant set of benchmark products must be collected prior to application of the decision tree because an incomplete set of benchmark solutions may lead to artificially high ratings for solutions that already exist in the marketplace. Moreover, with a rapidly changing technology landscape, a definition of dynamic creativity that accounts for “potential” originality rather than a fixed scale should be accounted for, as described by Corrazza (2016). The challenge remains to design a coding scheme that accommodates creative inconclusiveness in the context of the existing literature and provides insights for future scientific questions in the field of creativity.

Second, the DTOAD requires a lengthy training procedure. New raters are required to evaluate subsets of concepts and compare their results with expert ratings, and to repeat the procedure until sufficient inter-rater reliability is achieved. Our experience using the DTOAD, and training our research assistants, was that it was easier to identify highly creative products (typically assigned a 7.5 or 10 in our decision tree), but challenges were presented in the lower end of the scale (0–5). Understanding whether a product is radically different and therefore creative (e.g., receiving an originality score of 7.5 or 10) is not difficult. This is possibly why the CAT (Amabile, 1982) has been such a successful tool in non-engineering fields (cf. Baer and McKool, 2009; Beaty et al., 2013; Kaufman et al., 2013, Study 1). If one analyzes creativity in literature or art, a novice is usually able to identify a high degree of creativity without understanding all the details of the process. Similarly, a coder with no prior engineering knowledge would be able to identify highly creative engineering design for a common product like alarm clock or litter collection system (as opposed to a guided missile system). However, the disagreement in originality scoring that occurred

during the training process usually was at the lower end of the spectrum. Although there was broad agreement in scores of 0 where no novel feature was identified, we experienced challenges in separating designs in the 2.5–5 range. While this may not be an issue in business and industry, where the goal is to identify breakthrough levels of novelty, in research settings it is important that we can distinguish between all levels of originality. There was some disagreement about what constituted a ‘novel’ feature deserving a positive score. Coders also sometimes disagreed on what constituted an ‘isolated’ feature vs. a ‘moderately integrated’ feature. While some engineering knowledge may be helpful (cf. Kaufman et al., 2013, Study 2), perhaps clearer instructions are required to understand integration of features at the system level. To address these issues, we have started tabulating a database for novel features to help future coders.

Similar challenges regarding training coders apply to other creativity metrics, including the 5-point scale utilized in our previous research (cf. Genco et al., 2011; Johnson et al., 2014). Recent work in crowdsourcing, however, suggests that extensive training may not be required. For example, Green et al. (2014) spent 20 min training a large group of novice raters to evaluate concepts for the alarm clock problem discussed previously. Even with such a short training session, they found that novice raters with high inter-rater reliability amongst themselves served as a very good proxy for an expert rater. Large numbers of raters (on the order of 40) are needed, however, to identify raters with excellent inter-rater reliability amongst themselves. Perhaps these raters could be recruited via Mechanical Turk or other similar mechanisms, and the training could be conducted online. The success of this type of crowdsourcing effort may also depend on the raters’ familiarity with the design problem and the raters’ incentives for rating the concepts carefully and thoughtfully.

CONCLUSION

The DTOAD metric was an evolution from existing techniques reported in the literature, such as the CAT (Amabile, 1982), Shah et al.’s (2003) novelty metric, and the CEDA (Charyton et al., 2008; Charyton, 2014). The DTOAD also derives from previous modifications to the CEDA by Srivathsavai et al. (2010). It evolved as a part of an interdisciplinary study of engineering creativity conducted by several faculty and students at various universities. A variation of the CEDA (Charyton et al., 2008; Charyton, 2014) was previously used for analyzing alarm clocks (Genco et al., 2011; Johnson et al., 2014). However, we faced considerable challenges with low inter-rater reliability as we tried to use the method for a more complex engineering product like a litter collection system. While not every litter collection system is complex, this design problem requires individuals to generate ideas for creative products for which they have less familiarity as consumers. Other than trash cans and litter pickers, most litter collection systems are not meant for an individual consumer. Due to these challenges, we developed the DTOAD using a five point scale. The specific decision tree method described here for analyzing creativity was developed for analyzing concepts

for “next generation litter collection systems” generated by undergraduate engineering students.

The evolution of the DTOAD is an attempt to measure creativity in complex engineering designs that go beyond simple “features” or “variety” or “novelty.” It is an attempt to develop an algorithm for analyzing creativity in complex engineering designs for the future. The former creativity evaluations are useful indicators of creativity, but were not always geared toward evaluation of features as well as the system level integration. An important challenge that exists in analyzing ‘complex’ system level designs for creativity is to have a knowledge of how features are integrated at the system level. It also requires a working knowledge of the product to assess what the industry standards are, not just at the feature level, but at the system level as well. Therefore, considerable effort was expended during the training process of the coders to establish an understanding of the state of the art of the product, its features, and their integration. While analyzing the litter collection system, we did evaluate ‘features’ that were considered novel. However, to get a score of 5, a designer had to demonstrate integration of the ‘feature’ within the existing architecture. Higher scores were typically assigned to novel architectures that went beyond the existing industrial norms.

Despite some of the challenges encountered during the development and implementation of the DTOAD, we have been able to obtain meaningful and insightful trends. We have applied the decision tree to concepts generated by all 4 years of engineering students (freshmen through seniors). When we used the modified 6-3-5 method (Otto and Wood, 2001), we were able to analyze originality at the overall concept level and also at the individual level (Kershaw et al., 2015), as well as examine effects of group processes on creativity (Kershaw et al., 2016). We also have analyzed data for students who were provided an example product as well as students who were provided no examples. Further, we have also applied the DTOAD to evaluating thermometers, a product that was not considered when developing our originality metric. Overall, we have found that the senior mechanical engineering students have a higher originality score than freshmen, reinforcing studies by others reported in the literature (Cross et al., 1994; Ball et al., 1997; Atman et al., 1999). This provides some order of validity to our method of analysis. We have also found that the originator of a design typically has the highest contribution to creativity. We have also found an inverse correlation between originality and fixation (LeGendre et al., 2017; Simmons et al., 2018). Removing the example product lowered fixation drastically, but did not increase originality. In this way we have gained measurable insights into the design creativity thinking process of mechanical engineering students. These results will provide useful data points for curricular design where further creative thinking is required as a part of engineering education.

The mechanical engineering undergraduate curriculum is diverse with significant emphasis on quantitative techniques and set-piece problem solving (Accreditation Board for Engineering and Technology, 2014). Synthesis of concepts from various courses into a holistic design process is limited. Without the experience of synthesis, students are not encouraged to think

creatively or perform creative design tasks. One of the goals of creativity research in engineering is to understand how to improve the creative thinking process in the engineering curriculum (Dym et al., 2005; Phase, 2005; Duderstadt, 2010). Although a statistically significant difference was found in the originality scores between freshmen and seniors when measured longitudinally (Kershaw et al., 2015), the overall differences and trends were not drastic to indicate that students were being trained well in the creative process. It is our goal to use the data generated from the studies using the decision tree to propose active measures within the curriculum.

Measuring creativity in engineering design is an extremely important tool beyond academia as well. Establishing a creative toolbox and analyzing creativity are important for companies to develop new products for the future to stay competitive and maintain their cutting edge in an increasingly shrinking market space. In addition to a marketing survey it is becoming imperative for companies to test the “coolness” factor of many consumer products. This evaluation is often related to creative design. However, there are no standard tools available to companies that can measure creativity in engineering design. Therefore, developing creativity measuring tools for engineering design continues to be an important goal in design research (cf. Cropley and Kaufman, 2012).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

All data analyzed in this paper were retrieved from previously published or unpublished studies. Those studies were carried out in accordance with the recommendations of the Institutional Review Board (IRB) at the University of Massachusetts Dartmouth. Their respective research protocols were approved by the UMass Dartmouth IRB. All participants were provided with informed consent letters because an exemption from signed consent forms was granted by the UMass Dartmouth IRB. Participants read the consent letter, asked any questions that they had, and then consented to be part of the study by taking part in an in-class activity. Participants kept copies of the consent letter so that they could contact the researchers for additional information at a later date.

AUTHOR CONTRIBUTIONS

TK and SB developed the DTOAD based on, in part, previous originality metrics developed by KH-O and CS. TK, SB, and a research assistant completed the originality coding for the thermometer concepts. TK and KH-O completed the novelty coding. TK completed all statistical analyses. KH-O and SB assisted with the interpretation of the results of these analyses.

TK wrote the initial draft of Sections “Introduction,” “Development of our Decision Tree Based Originality Scoring Method,” “Applying the DTOAD: Full Protocol,” “Re-analysis of Litter Collection System Concepts,” “Applying the DTOAD to a New Design Problem,” and “Advantages.” SB wrote the initial drafts of Sections “Summary of Previously Published Results” and “Conclusions.” CS wrote the initial draft of Section “Limitations.” KH-O provided helpful comments on all sections. All authors edited all sections of the manuscript, responded to reviewer comments, and approved the final version.

FUNDING

The litter collection system concepts were collected under previous support from the National Science Foundation under grant no. DUE-1140424 to KH-O, TK, and SB and grant no. DUE-1140628 to CS, and from the Seed Funding program from the Office of the Provost at the University of Massachusetts

Dartmouth to TK and SB. The thermometer concepts were collected under previous support from the National Science Foundation under grant no. CMMI-0825461 to KH-O and grant no. CMMI-0825713 to CS. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

ACKNOWLEDGMENTS

We are grateful to the research assistants who have contributed to this work over the years: Adam Young, Molly McCarthy, Becky Peterson, Alex LeGendre, and Chris Simmons. We thank all the UMass Dartmouth mechanical engineering instructors who allowed us to collect data in their classes: Don Foster, Afsoon Amirzadeh, Marc Richman, Mehdi Raessi, Steve Warner, and Amit Tandon. We also thank the two reviewers for their useful feedback and suggestions for revisions.

REFERENCES

- Accreditation Board for Engineering and Technology (ed.). (2014). “Student outcomes” in *2015-2016 Criteria for Accrediting Engineering Programs*, (Baltimore, MD: ABET).
- Amabile, T. M. (1982). Social psychology of creativity: a consensual assessment technique. *J. Pers. Soc. Psychol.* 43, 997–1013. doi: 10.1037/0022-3514.43.5.997
- Atman, C. J., Chimka, J. R., Bursic, K. M., and Nachtmann, H. L. (1999). A comparison of freshmen and senior design processes. *Des. Stud.* 20, 131–152. doi: 10.1016/S0142-694X(98)00031-3
- Baer, J., and McKool, S. S. (2009). “Assessing creativity using the consensual assessment technique,” in *Handbook of Research on Assessment Technologies, Methods, and Applications in Higher Education*, ed. C. Shreiner (Hershey, PA: Information Science Reference), 65–77.
- Ball, L. J., Evans, J. S., Dennis, I., and Ormerod, T. C. (1997). Problem-solving strategies and expertise in engineering design. *Think. Reason.* 3, 247–270. doi: 10.1016/j.jbi.2017.05.007
- Batey, M. (2012). The measurement of creativity: from definitional consensus to the introduction of a new heuristic framework. *Creat. Res. J.* 24, 55–65. doi: 10.1080/10400419.2012.649181
- Beaty, R. E., Smeekens, B. A., Silvia, P. J., Hodges, D. A., and Kane, M. J. (2013). A first look at the role of domain-general cognitive and creative abilities in jazz improvisation. *Psychomusicology* 23, 262–268. doi: 10.1037/a0034968
- Braasch, J. L. G., Goldman, S. R., and Wiley, J. (2013). The influences of text and reader characteristics on learning from refutations in science texts. *J. Educ. Psychol.* 105, 561–578. doi: 10.1037/a0032627
- Brown, D. C. (2014). “Problems with the calculation of novelty metrics,” in *Paper Presented at the Design Creativity Workshop of the Sixth International Conference on Design Computer and Cognition (DCC’14)*, London.
- Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychol. Rev.* 67, 380–400. doi: 10.1037/h0040373
- Cantor, A. B. (1996). Sample-size calculations for Cohen’s kappa. *Psychol. Methods* 1, 150–153. doi: 10.1037/1082-989X.1.2.150
- Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K., and Kotovsky, K. (2011). On the benefits and pitfalls of analogies for innovative design: ideation performance based on analogical distance, commonness, and modality of examples. *J. Mech. Des.* 133, 1–11. doi: 10.1115/1.4004396
- Charyton, C. (2014). *Creative Engineering Design Assessment (CEDA): Background, Directions, Manual, Scoring guide, and Uses*. New York, NY: Springer. doi: 10.1007/978-1-4471-5379-5
- Charyton, C., Jagacinski, R. J., and Merrill, J. A. (2008). CEDA: a research instrument for creative engineering design assessment. *Psychol. Aesthet. Creat. Arts* 2, 147–154. doi: 10.1037/1931-3896.2.3.147
- Cheng, P. C.-H., Lowe, R. K., and Scaife, M. (2001). Cognitive science approaches to understanding diagrammatic representations. *Artif. Intell. Rev.* 15, 79–94. doi: 10.1023/A:1006641024593
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: a practical guide. *J. Learn. Sci.* 6, 271–315. doi: 10.1207/s15327809jls0603_1
- Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., et al. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cogn. Sci.* 42, 1777–1832. doi: 10.1111/cogs.12626
- Chi, M. T. H., Roy, M., and Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: insights about human tutoring effectiveness from vicarious learning. *Cogn. Sci.* 32, 301–341. doi: 10.1080/03640210701863396
- Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* 70, 213–220.
- Corraza, G. E. (2016). Potential originality and effectiveness: the dynamic definition of creativity. *Creat. Res. J.* 28, 258–267. doi: 10.1080/10400419.2016.1195627
- Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.* 52, 281–302. doi: 10.1037/h0040957
- Cropley, A. J. (2006). In praise of convergent thinking. *Creat. Res. J.* 18, 391–404. doi: 10.1207/s15326934crj1803_13
- Cropley, D. H. (2015). *Creativity in Engineering: Novel Solutions to Complex Problems*. Waltham, MA: Academic Press. doi: 10.1016/B978-0-12-800225-4.00002-1
- Cropley, D. H., and Cropley, A. J. (2010). “Functional creativity: “products” and the generation of effective novelty,” in *The Cambridge Handbook of Creativity*, eds J. C. Kaufman and R. J. Sternberg (New York, NY: Cambridge University Press), 301–320. doi: 10.1017/CBO9780511763205.019
- Cropley, D. H., Cropley, A. J., and Sandwith, B. L. (2017). “Creativity in the engineering domain,” in *The Cambridge Handbook of Creativity across Domains*, eds J. Kaufman, V. P. Glaveanu, and J. Baer (New York, NY: Cambridge University Press), 261–275. doi: 10.1017/9781316274385.015
- Cropley, D. H., and Kaufman, J. C. (2012). Measuring functional creativity: non-expert raters and the Creative Solution Diagnosis Scale. *J. Creat. Behav.* 46, 119–137. doi: 10.1002/jocb.9
- Cross, N., Christiaans, H., and Dorst, K. (1994). Design expertise amongst student designers. *J. Art Des. Educ.* 13, 39–56. doi: 10.1111/j.1476-8070.1994.tb00356.x
- Dollinger, S. J., and Shafran, M. (2005). Note on the consensual assessment technique in creativity research. *Percept. Mot. Skills* 100, 592–598. doi: 10.2466/pms.100.3.592-598

- Duderstadt, J. J. (2010). "Engineering for a changing world," in *Holistic Engineering Education: Beyond Technology*, eds D. Grasso and M. Brown Burkins (New York, NY: Springer), 17–36.
- Dunbar, K. (1997). "How scientists think: on-line creativity and conceptual change in science," in *Creative Thought: an Investigation of Conceptual Structures and Processes*, eds T. B. Ward, S. M. Smith, and J. Vaid (Washington, DC: American Psychological Association), 461–494.
- Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., and Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. *J. Eng. Educ.* 94, 103–120. doi: 10.1002/j.2168-9830.2005.tb00832.x
- Fleiss, J. (1981). *Statistical Methods for Rates and Proportions*. New York, NY: John Wiley and Sons.
- Friedman, H. H., and Amoo, T. (1999). Rating the rating scales. *J. Mark. Manag.* 9, 114–123.
- Genco, N. (2012). *Empathic Experience Design*. Master's thesis, University of Massachusetts Dartmouth, North Dartmouth, MA.
- Genco, N., Johnson, D., Hölttä-Otto, K., and Seepersad, C. C. (2011). "A study of the effectiveness of empathic experience design as a creativity technique," in *Proceedings of the ASME 2011 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, New York, NY. doi: 10.1115/DETC2011-48256
- Getzels, J., and Csikzentmihalyi, M. (1976). *The Creative Vision: a Longitudinal Study of Problem-Finding in Art*. New York, NY: Wiley.
- Goldman, S. R., and Murray, J. D. (1992). Knowledge of connectors as cohesion devices in text: a comparative study of Native-English and English-as-a-Second Language speakers. *J. Educ. Psychol.* 84, 504–519. doi: 10.1037/0022-0663.84.4.504
- Goodwin, P., and Wright, G. (2004). *Decision Analysis for Management Judgment*, 3rd Edn. Hoboken, NJ: John Wiley & Sons, Inc. doi: 10.1115/DETC2014-34434
- Green, M., Seepersad, C. C., and Hölttä-Otto, K. (2014). "Crowd-sourcing the evaluation of creativity in conceptual design: a pilot study," in *Proceedings of the ASME 2014 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, New York, NY.
- Guilford, J. P. (1956). The structure of intellect model. *Psychol. Bull.* 53, 267–293. doi: 10.1037/h0040755
- Gwet, K. L. (2014). *Handbook of Inter-Rater Reliability*, 4th Edn. Gaithersburg, MD: Advanced Analytics, LLC.
- Hass, R. W., and Weisberg, R. W. (2015). Revisiting the 10-year rule for composers from the *Great American Songbook*: on the validity of two measures of creative production. *Psychol. Aesthet. Creat. Arts* 9, 471–479. doi: 10.1037/aca0000021
- Jaarsveld, S., and Lachmann, T. (2017). Intelligence and creativity in problem solving: the importance of test features in cognition research. *Front. Psychol.* 8:134. doi: 10.3389/fpsyg.2017.00134
- Jansson, D. G., and Smith, S. M. (1991). Design fixation. *Des. Stud.* 12, 3–11. doi: 10.1016/0142-694X(91)90003-F
- Jeffries, K. K. (2017). A CAT with caveats: is the Consensual Assessment Technique a reliable measure of graphic design creativity? *Int. J. Des. Creat. Innov.* 5, 16–28. doi: 10.1080/21650349.2015.1084893
- Johnson, D. G., Genco, N., Saunders, M. N., Williams, P., Seepersad, C. C., and Hölttä-Otto, K. (2014). An experimental investigation of the effectiveness of empathic experience design for innovative concept generation. *J. Mech. Des.* 136, 1–12. doi: 10.1115/1.4026951
- Kaufman, J. C., Baer, J., Cropley, D. H., Reiter-Palmon, R., and Sinnett, S. (2013). Furious activity vs. understanding: how much expertise is needed to evaluate creative work? *Psychol. Aesthet. Creat. Arts* 7, 332–340. doi: 10.1037/a0034809
- Kershaw, T. C., Lippman, J. P., and Fugate, J. M. B. (2018). Practice makes proficient: teaching undergraduate students to understand published research. *Instr. Sci.* 46, 1–26. doi: 10.1007/s11251-018-9456-2
- Kershaw, T. C., Peterson, R. L., and Bhowmick, S. (2016). "The influence of group interaction on creativity in engineering design," in *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, eds A. Papafragou, D. Grodner, D. Mirman, and J. C. Trueswell (Austin, TX: Cognitive Science Society), 2135–2140.
- Kershaw, T. C., Peterson, R. L., McCarthy, M. A., Young, A. P., Seepersad, C. C., Williams, P. T., et al. (2015). "A cross-sectional and longitudinal examination of the development of innovation capability in undergraduate engineering students," in *Proceedings of the ASME 2015 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, New York, NY. doi: 10.1115/DETC2015-47650
- Kershaw, T. C., Seepersad, C. C., Hölttä-Otto, K., Williams, P. T., Young, A. P., Bhowmick, S., et al. (2014). "The effects of the undergraduate curriculum and individual differences on student innovation capabilities," in *Proceedings of the ASME 2014 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, New York, NY. doi: 10.1115/DETC2014-35540
- Kozbelt, A., Beghetto, R. A., and Runco, M. A. (2010). "Theories of creativity," in *The Cambridge Handbook of Creativity*, eds J. C. Kaufman and R. J. Sternberg (New York, NY: Cambridge University Press), 20–47. doi: 10.1017/CBO9780511763205.004
- Kudrowitz, B., Free, A., and Hough, L. (2016). "Identification of personality and cognitive indicators of creative potential," in *Proceedings of the Fourth International Conference on Design Creativity*, eds J. Linsey, M. Yang, and Y. Nagai (Atlanta, GA: Georgia Institute of Technology).
- Landis, J., and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310
- Larkin, J. H., and Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cogn. Sci.* 11, 65–100. doi: 10.1111/j.1551-6708.1987.tb00863.x
- LeGendre, A., Kershaw, T. C., Peterson, R. L., and Bhowmick, S. (2017). "The relationship between fixation and originality in undergraduate mechanical engineering students," in *Proceedings of the ASME 2017 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, New York, NY. doi: 10.1115/DETC2017-67833
- Lubart, T. (2017). The 7 Cs of creativity. *J. Creat. Behav.* 51, 293–296. doi: 10.1002/jobc.190
- Mednick, S. A. (1962). The associative basis of the cognitive process. *Psychol. Rev.* 69, 220–232. doi: 10.1037/h0048850
- Moreno, D. P., Yang, M. C., Hernandez, A. A., Linsey, J. S., and Wood, K. L. (2014). "A step beyond to overcome design fixation: a design-by-analogy approach," in *Design Computing and Cognition '14*, eds J. Gero and S. Hanna (Berlin: Springer), 607–624.
- Muldner, K., Lam, R., and Chi, M. T. H. (2014). Comparing learning from observing and from human tutoring. *J. Educ. Psychol.* 106, 69–85. doi: 10.1080/03640210701863396
- Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., and Baas, M. (2010). The dual pathway to creativity model: creative ideation as a function of flexibility and persistence. *Eur. Rev. Soc. Psychol.* 21, 34–77. doi: 10.1080/10463281003765323
- Nye, N. J., Goldman, S. R., Voss, J. F., Hmelo, C., Williams, S., and Cognition and Technology Group at Vanderbilt (1997). Complex mathematical problem solving by individuals and dyads. *Cogn. Instr.* 15, 435–484. doi: 10.1207/s1532690xcil504_1
- Oman, S. K., Tumer, I. Y., Wood, K., and Seepersad, C. C. (2013). A comparison of creativity and innovation metrics and sample validation through in-class design projects. *Res. Eng. Des.* 24, 65–92. doi: 10.1007/s00163-012-0138-9
- Otto, K. N., and Wood, K. L. (2001). *Product Design: Techniques in Reverse Engineering and New Product Development*. Upper Saddle River, NJ: Prentice Hall.
- Phase, I. I. (2005). *Educating the Engineer of 2020: Adapting Engineering Education to the New Century*. Washington, DC: The National Academies Press.
- Plucker, J. A., and Makel, M. C. (2010). "Assessment of creativity," in *The Cambridge Handbook of Creativity*, eds J. C. Kaufman and R. J. Sternberg (New York, NY: Cambridge University Press), 48–73. doi: 10.1017/CBO9780511763205.005
- Rostan, S. M. (2010). Studio learning: motivation, competence, and the development of young art students' talent and creativity. *Creat. Res. J.* 22, 261–271. doi: 10.1080/10400419.2010.503533
- Runco, M. A. (2004). "Everyone has creative potential," in *Creativity: from Potential to Realization*, eds R. J. Sternberg, E. L. Grigorenko, and J. L. Singer (Washington, DC: American Psychological Association), 21–30. doi: 10.1037/10692-002
- Sarkar, P., and Chakrabarti, A. (2011). Assessing design creativity. *Des. Stud.* 32, 348–383. doi: 10.1016/j.destud.2011.01.002
- Shah, J. J., Vargas-Hernandez, N., and Smith, S. M. (2003). Metrics for measuring ideation effectiveness. *Des. Stud.* 24, 113–134. doi: 10.1016/S0142-694X(02)00034-0

- Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., et al. (2008). Assessing creativity with divergent thinking tasks: exploring the reliability and validity of new subjective scoring methods. *Psychol. Aesthet. Creat. Arts* 2, 68–85. doi: 10.1037/1931-3896.2.2.68
- Simmons, C. C., Kershaw, T. C., LeGendre, A., and Bhowmick, S. (2018). “The influence of physical examples on originality and fixation in engineering design,” in *Proceedings of the ASME 2018 International Design Engineering Technical Conferences & Computer and Information in Engineering Conference*, New York, NY. doi: 10.1115/DETC2018-85396
- Simonton, D. K. (2011). Creativity and discovery as blind variation: Campbell's (1960) BVSR Model after the half-century mark. *Rev. Gen. Psychol.* 15, 158–174. doi: 10.1037/a0022912
- Smith, S. M., and Blankenship, S. E. (1991). Incubation and the persistence of fixation in problem solving. *Am. J. Psychol.* 104, 61–87. doi: 10.2307/1422851
- Srivathsavai, R., Genco, N., Hölttä-Otto, K., and Seepersad, C. C. (2010). “Study of existing metrics used in measurement of ideation effectiveness,” in *Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, New York, NY. doi: 10.1115/DETC2010-28802
- Sternberg, R. J., Kaufman, J. C., and Pretz, J. E. (2002). *The Creativity Conundrum: a Propulsion Model of Kinds of Creative Contributions*. New York, NY: Psychology Press.
- Sternberg, R. J., and Lubart, T. I. (1999). “The concept of creativity: prospects and paradigms,” in *Handbook of Creativity*, ed. R. J. Sternberg (New York, NY: Cambridge University Press), 3–15.
- Storme, M., Myszkowski, N., Celik, P., and Lubart, T. (2014). Learning to judge creativity: the underlying mechanisms in creativity training for non-expert judges. *Learn. Individ. Dif.* 32, 19–25. doi: 10.1016/j.lindif.2014.03.002
- Toh, C. A., and Miller, S. R. (2014). The impact of example modality and physical interactions on design creativity. *J. Mech. Des.* 136, 1–8. doi: 10.1115/1.4027639
- Treffinger, D. J., Isaksen, S. G., and Dorval, K. B. (2006). *Creative Problem Solving: an Introduction*, 4th Edn. Waco, TX: Prufrock Press.
- Vasconcelos, L. A., and Crilly, N. (2016). Inspiration and fixation: questions, methods, findings, and challenges. *Des. Stud.* 42, 1–32. doi: 10.1016/j.destud.2015.11.001
- Wieth, M. B., and Francis, A. P. (2018). Conflicts and consistencies in creativity research and teaching. *Teach. Psychol.* 45, 363–370. doi: 10.1177/0098628318796924
- Yilmaz, S., Daly, S. R., Christian, J. L., Seifert, C. M., and Gonzalez, R. (2014). Can experienced designers learn from new tools? A case study of idea generation in a professional engineering team. *Int. J. Des. Creat. Innov.* 2, 82–96. doi: 10.1080/21650349.2013.832016
- Youmans, R. J. (2011). The effects of physical prototyping and group work on the reduction of design fixation. *Des. Stud.* 32, 115–138. doi: 10.1016/j.destud.2010.08.001

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 Kershaw, Bhowmick, Seepersad and Hölttä-Otto. 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.



Spontaneous Visual Imagery During Meditation for Creating Visual Art: An EEG and Brain Stimulation Case Study

Caroline Di Bernardi Luft^{1,2*}, Ioanna Zioga¹, Michael J. Banissy² and Joydeep Bhattacharya²

¹ School of Chemical and Biological Sciences, Queen Mary, University of London, London, United Kingdom, ² Department of Psychology, Goldsmiths, University of London, London, United Kingdom

OPEN ACCESS

Edited by:

Kathryn Friedlander,
University of Buckingham,
United Kingdom

Reviewed by:

Barbara Colombo,
Champlain College, United States
Fei Luo,
Institute of Psychology (CAS), China

*Correspondence:

Caroline Di Bernardi Luft
c.luft@qmul.ac.uk

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 23 July 2018

Accepted: 22 January 2019

Published: 22 February 2019

Citation:

Luft CDB, Zioga I, Banissy MJ
and Bhattacharya J (2019)
Spontaneous Visual Imagery During
Meditation for Creating Visual Art: An
EEG and Brain Stimulation Case
Study. *Front. Psychol.* 10:210.
doi: 10.3389/fpsyg.2019.00210

Experienced meditators often report spontaneous visual imagery during deep meditation in the form of lights or other types of visual images. These experiences are usually interpreted as an “encounters with light” and gain mystical meaning. Contrary to the well-studied intentional and controlled visual imagery, spontaneous imagery is poorly understood, yet it plays an important role in creativity of visual artists. The neural correlates of such experiences are indeed hard to capture in laboratory settings. In this case study we aimed to investigate the neural correlates of spontaneous visual imagery in an artist who experiences strong visual imagery during meditation. She uses these images to create visual art. We recorded her EEG during seven meditation sessions in which she experienced visual imagery episodes (visions). To examine the functional role of the neural oscillations we also conducted three separate meditation sessions under different transcranial alternating current (tACS) brain stimulation: alpha (10 Hz), gamma (40 Hz) and sham. We observed a robust increase in occipital gamma power (30–70 Hz) during the deepest stage of meditation across all sessions. This gamma increase was consistent with the experience of spontaneous visual imagery: higher during visions compared to no visions. Alpha tACS was found to affect the contents of her visual imagery, making them sharper, shorter and causing more visions to occur; the artist reported that these sharp images were too detailed to be used in her art. Interestingly, gamma and sham stimulation had no impact on the visual imagery contents. Our findings raise the hypothesis that occipital gamma might be a neural marker of spontaneous visual imagery, which emerges in certain meditation practices of experienced meditators.

Keywords: visual arts, EEG, transcranial alternating current stimulation (tACS), gamma oscillations, meditation, spontaneous visual imagery, entrainment, alpha oscillations

INTRODUCTION

In common terms, imaginative and creative are often used interchangeably to describe ideas/objects or the individuals producing them. However, creativity is not necessarily the same as imagination (Singer, 2011), and the relationship between imagery and creative cognition is multilayered. Psychologically, imagination is a broad term representing our almost unique ability

to transcend the current constraints of space, time, and causality leading to mental simulation of future, creating fictional, unusual worlds, and experiences (Taylor, 2011); in essence, imagination includes both creative and non-creative thoughts.

Ward (1994) has argued for structured imagination, referring to a nonrandom but structured approach of generating new ideas and concepts; here, imagination is constrained by existing knowledge and categories. For example, when participants are asked to imagine animals living on a distant planet, their responses are structured by the properties of the animals living on the planet earth. Although this type of imaginative thinking has been found to be quite useful in creative idea generation (Ward, 1995), it is nonetheless a targeted method of imagination which is devoid of spontaneity.

Whereas there is a role for spontaneous thoughts in the creativity research, especially in studies related to mind-wandering and creativity (Baird et al., 2012), the spontaneous imagery, on the other hand, is much less studied, and we know very little about the role of spontaneous imagery in creativity. Interestingly, spontaneous (visual) imagery is often associated with meditation. For example, in the well-known “encounters with light” experience, meditators, primarily practicing in the Buddhist tradition, report several forms of lights or luminous experiences (Lo et al., 2003; Lindahl et al., 2013). These visual images of inner light may be a special type of visual imagery, albeit a spontaneous one, which has been overlooked in the imagery literature as most studies on visual imagery have looked into voluntary visual imagery whose content rely heavily on working memory (Albers et al., 2013; Dentico et al., 2014). The close link between spontaneous imagery and meditation was speculated by Austin (2003) almost 40 years ago, “*The ease with which meditators can learn to let go and enter a satisfying state of calm, detached awareness does correlate with their basic ability to produce spontaneous visual imagery, to free associate, and to tolerate any unreal experiences that may occur*” (Austin, 2003, p. 184). However, the underlying neuronal correlates of these types of spontaneous visual imagery are not known and nor their potential links to creative cognition.

In the present study, we aimed to address these issues by adopting a phenomenological approach based on the first person experience (Lutz et al., 2002). Lia Chavez (L.C.) is a New York (United States) based professional artist who has featured in a number of internationally renowned venues. L.C. experiences intense visual imagery generated spontaneously during her meditations, and uses the content of these visual images experienced during her deep meditative state as a source of creative inspiration for her multimedia work. We performed multisession recording of EEG signals from L.C.’s brain during meditation and the experiences of visual imagery. Subsequently, we tested whether it was possible to modulate visual imagery experiences by administering transcranial brain stimulation, in particular, transcranial alternating current brain stimulation (tACS), during meditation. tACS is a noninvasive technique that can be used to modulate endogenous brain oscillations possibly through entrainment (Antal and Paulus, 2013). It has been successfully used to probe the functional role of certain oscillations

(e.g., alpha and gamma) on perception and cognition (e.g., Laczó et al., 2012; Janik et al., 2015; Luft et al., 2018). This technique allowed us to explore the possibility of causally interfering with the visual imagery as experienced during meditation.

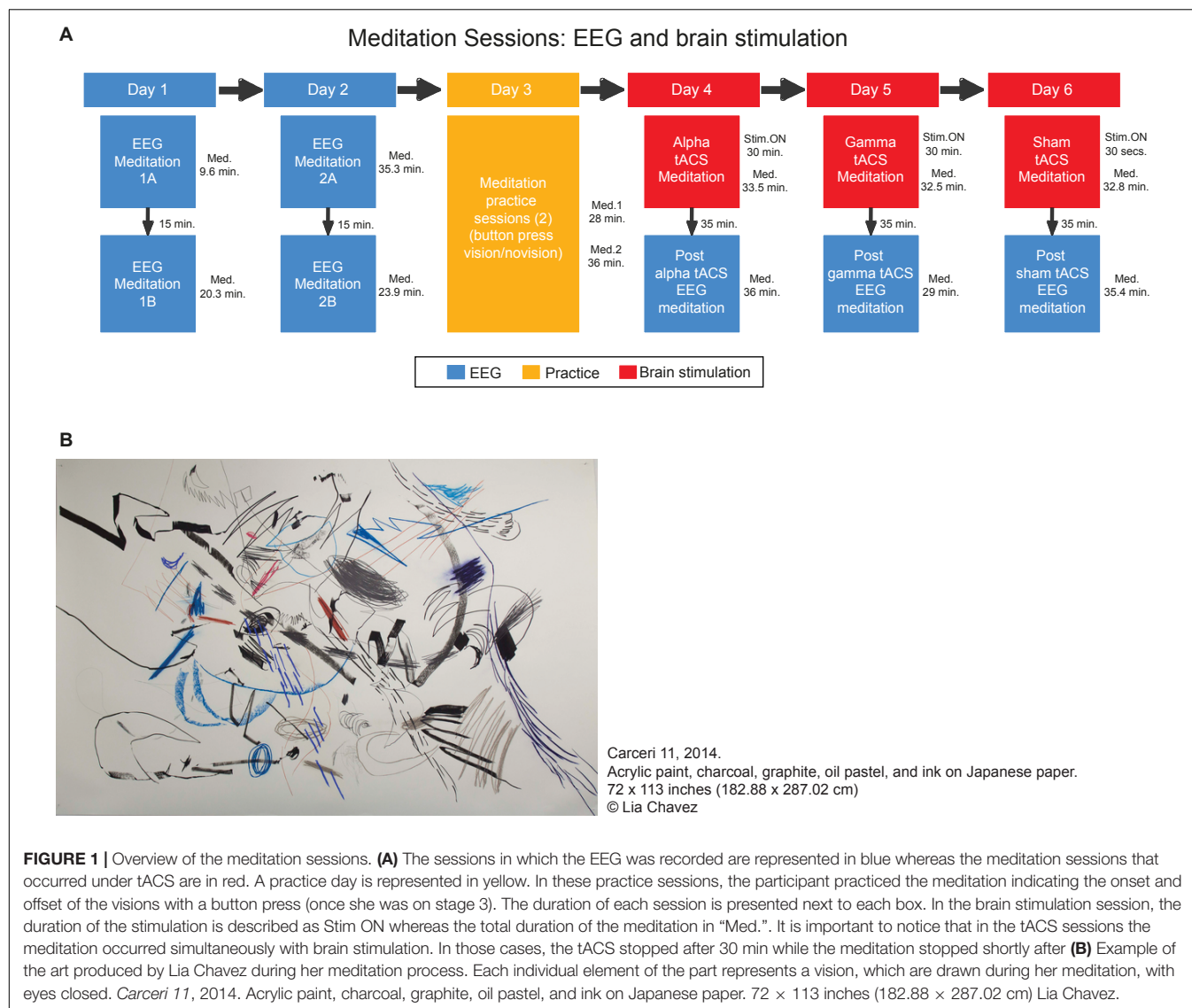
Therefore, our objectives were as follows: (1) to investigate oscillatory changes during different depth levels of meditation (three stages, see “Methods”); (2) to analyze the oscillatory correlates of each visual imagery episode during the deepest stage of meditation; (3) to explore the casual relationship of oscillatory changes during different levels by analyzing the effects of alpha, gamma and sham tACS, on her visual imagery during meditation. Based on prior neuroimaging work (e.g., Lehmann et al., 2001; Lutz et al., 2004; Cahn et al., 2010; Braboszcz et al., 2017), we predicted that occipital gamma would increase with the depth of meditation. Second, we predicted that during deep meditation, the increase in occipital gamma power would be higher during her visions rather than no-visions. Third, we predicted that gamma tACS would boost her visual imagery during deep meditation.

MATERIALS AND METHODS

In this case study, a professional artist took part in 10 meditation sessions. These sessions took place over 6 separate days spread over a period of a few months. During seven sessions EEG (electroencephalogram) signals were recorded in order to investigate the large scale neural oscillatory changes during meditation. In the three other sessions, electrical brain stimulation (transcranial alternating current stimulation, tACS) was applied to modulate cortical activity in a frequency dependent manner in order to investigate the functional role of specific neural oscillations in meditation and associated creative imagery. The overview of the experimental sessions and a sample of her work developed based on her visual imagery are shown in **Figures 1A,B**. Written informed consent was obtained from the participant at each session. Further, our participant has accepted to have her identity disclosed in the paper. The experimental protocol was approved by the local Ethics Committee of the Department of Psychology at Goldsmiths, and all procedures were conducted in accordance with the Declaration of Helsinki.

Participant

A New York-based internationally-exhibited artist, Lia Chavez (L.C. afterward), took part in this study. She is among Origin Magazine’s Top 100 Creatives Changing the World for 2015. Her work explores the phenomenology of light and the possibilities of using consciousness as an art material. She initially approached one of the authors (J.B.) detailing her experiences of spontaneous visual imagery during meditation as a source of her creative inspiration; subsequently, she expressed her willingness to participate in neuroimaging experiments to investigate the correlates between functional brain activities and “*the profound moments of interior visualizations . . . revealing the inception of the creative spark . . .*” (L.C., personal communication). L.C. has been practicing meditation intensely for over 10 years, including



periods in which she meditated for lengthy periods of time (up to 10 h a day for 2 weeks at a time). Her meditation practice includes two different types of meditation according to Tibetan Buddhism: stabilizing and analytical (Chodron, 2007). Stabilizing can be considered as a strategy for quietening the mind by focusing the attention on simple repetition of words or mantras, on the breath, or even on a symbol within the mind. This type of meditation relies on serial repetition with the purpose to prepare the mind for a deeper kind of focused contemplation. The analytical meditation, on the other hand, is that state of deeper contemplation in which the meditator experiences a quiet mind in order to obtain a conceptual understanding of how things are, to a depth that would offer enough clarity and novel insight into the true nature of that concept. Both types of meditation can usually be combined within a single meditation session. L.C. reports using a variety of stabilizing techniques such as repeating a mantra, focusing on the pause between the in and out breath, and focusing attention on different body

parts to generate heightened sensation. L.C. reports that once her mind is stable and a threshold is crossed, she goes into analytical contemplation which is the state that she experiences the spontaneous visions, which she calls “encounters with light.” During these spontaneous visions, she reports trying to remain detached from any emotions or judgments associated with the visual experience. This is how she describes her experience:

“I integrate a variety of cross-disciplinary contemplative traditions into my artistic process as a way of exploring the inception of the creative spark, how the creative artist’s own ontology of becoming incarnates into art objects, and how this process might relate to the cosmological order. Durational analytic meditation is at the core of my process. As I’ve journeyed through deep meditation into the vast unknown of my own inner landscape, I’ve discovered that the silent mind is, in fact, the seeing eye within a great storm. In deep analytical meditation, I experience cataclysmic visions of vortices, fibers of electricity, clouds of short-lived photons, cascading firebolts, and embryonic stars. It’s a process which

feels as though I am observing passionate and terrifying dances between the elements — a mental meteorology, if you will. As an artist who has always worked with light as a primary art material, you can imagine how powerful it was for me to encounter this experience for the first time in 2012. In time, I've come to discover that experiences of meditation-induced encounters with light is a widely-documented phenomenon throughout cross-cultural meditation traditions, most prominently within Buddhist meditation practice. Since first encountering these visions of luminous objects, I have cultivated durational analytic meditation as a source of inspiration for my visual and performance artwork." (L.C., personal communication). During her work, she depicts each vision on canvas with her eyes blindfolded. One example of the results can be visualized in **Figure 1B**. Her mixed media drawings are generated through several hours in a meditative state while continually blindfolded without sound. As she works, she positions herself atop the canvas surface and fashions complex gestural glyphs to depict her visions as they occur.

Meditation Sessions

In each meditation session, the subject sat in her usual meditation posture on a flat chair holding a response box with four buttons. She was instructed to rest with her eyes closed for 5 min followed by another 5 min with eyes opened, both resting periods without any meditation. This allowed us to collect resting state neural recording and to make her feel familiarized and comfortable within the laboratory setting. Following the resting periods, she was instructed to start her meditation and she pressed the button 1 (stage 0) in the response box to indicate its onset. In stage 0, she was not yet meditating but attempting to do so. Once she started reaching what she considered as an initial meditative state, she pressed button 2 (stage 1). When the meditation advanced to a deeper stage, she pressed the button 3 (stage 2). Note that this stage 2 was associated with the transition from stabilizing to analytical meditation. As soon as she entered into deep state of analytical meditation, she pressed button 4 (stage 3), which is the stage in which she experiences her visions. It is important to notice that the meditative stages she indicated were based on her individual experience of meditation depth, which cannot be compared between people. She reported that her visions usually occur in her deepest meditative states only. What defined the meditation stage for her was the depth of the state and not the presence of visions. There was no time limit or any other constraint applied to our participant in relation to her meditation practice. Once stage 3 was finished, she pressed button 1 to signal the offset of the meditation session.

In days 6, the procedures for meditation were identical except that once she was on stage 3, she indicated the onset and offset of each vision. Both the onset and the offset of individual vision episode were registered by button press. Further, any change of the content of the vision, i.e., a vision was followed by another vision rather than a no vision, was also indicated by another button press. She had two sessions in a previous day (day 3) for practicing this technique but the EEG data for these two practice sessions were not analyzed. By using this procedure, we could quantify not only the duration of each vision but the number of

different vision contents that occurred during the meditation. It is important to notice that all sessions were more than 24 h apart.

EEG Recording and Analysis

Continuous EEG signals were recorded using 64 active electrodes using a BioSemi ActiveTwo amplifier. The electrodes were placed according to the extended 10–20 system of electrode placement. Vertical and horizontal electro-oculograms were recorded using four additional external channels to monitor eye movements. The signals sampled at 512 Hz, bandpass filtered between 0.16 and 100 Hz. The instruction screen, button presses and event timings were recorded using the MATLAB based toolbox Cogent 2000¹. The EEG data was processed and analyzed by MATLAB based custom scripts and the following toolboxes: EEGLAB for preprocessing (Delorme and Makeig, 2004) and the signal processing toolbox in MATLAB. For preprocessing, we re-referenced the data to the arithmetic average of the two earlobes, and high-pass filtered at 0.5 Hz. The data was visually inspected for removal of visible artifacts such as muscle activity and eye-movements/saccades. For the vision vs. no-vision analysis, the data was also segmented into epochs of 2 s but the first epoch, immediately following the response, was excluded to avoid interference activity related to the button press.

In order to analyze the oscillatory response at each meditation stage, we estimated the power spectral density using the Welch's method (averaged periodogram), by dividing the data into 2 s windows with an overlap of 50%. We estimated the spectral power from 1 to 80 Hz in steps of 0.5 Hz. The power values at each electrode and each condition were averaged based on the standard EEG frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (13–30 Hz), gamma 1 (30–45 Hz), and gamma 2 (55–80 Hz). For the vision vs. no vision comparison we selected broad gamma band from 30 to 80 Hz since there was no difference between gamma 1 and gamma 2. The EEG data was expressed as percentage changes from the baseline power at stage 0 (non-meditative state).

For statistical comparisons, we used the spectral power values for each epoch. In order to avoid circularity, the data of days 1 and 2 were used to guide the main comparison of the vision vs. no-vision analysis. The electrodes showing peak percentage increase (analysis merging all the sessions from day 1 and 2) were selected for the contrast between vision and no vision contrasts in the sessions following brain stimulation (days 4, 5, and 6). Since we only found a robust change in gamma band, we only compared visions vs. no visions in this frequency band.

Brain Stimulation: tACS

Transcranial alternating current was delivered through a battery driven NeuroConn DC-Plus Stimulator. Two saline soaked sponged electrodes ($5 \times 5 \text{ cm} = 25 \text{ cm}^2$) were attached to conductive rubber electrodes attached to participants' scalps with rubber head straps. A sinusoidal current of 1.5 mA peak-to-peak was applied at the frequency of 10 Hz for alpha tACS, 40 Hz for gamma tACS, and 10 Hz for sham (tACS duration was only for the first 30 s for sham) with a zero-degree phase

¹<http://www.vislab.ucl.ac.uk/cogent.php>

offset and no DC offset. The electrodes were attached bilaterally on the occipital areas: electrodes PO7 and PO8 according to the extended 10–20 system. This montage was chosen for two reasons: (1) we needed to stimulate occipital areas as this was the area where we observed increased gamma oscillations in stage 3 of meditation; (2) we specifically chose PO7 and PO8 because this montage minimizes the risk of phosphenes as it reduces the current flow to the eyes (Laakso and Hirata, 2013). Additionally, we opted for traditional frequencies of stimulation (10 and 40 Hz) rather than individualized peak frequencies due to the difficulty of a clear peak in the gamma band. The stimulation lasted for 30 min and happened simultaneous to the meditation. After each session, the participant was asked to report any sensations she had experienced during the tACS session. She did not know the modality of the brain stimulation or the possibility of different frequencies. After the three sessions, she reported feeling a tickling sensation during the beginning, but she also reported that this sensation faded away shortly and she could not feel anything during the meditation (in all three sessions). After the last session, she was asked to indicate if she could guess whether the sessions were sham or stimulation. She reported having active stimulation during all sessions, which shows that she could not notice the difference between the sham and active stimulation sessions. This is expected since it was found that the cutaneous sensation persists after the stimulation is switched off on sham sessions which makes it hard for the participants to distinguish the stimulation conditions (Ambrus et al., 2012).

RESULTS

Meditation EEG Sessions

As there was no constraint or interference with the meditation practice as performed by L.C. in the lab, the duration varied across sessions (see **Figure 1**). The first session was the shortest, with a total duration of 9.65 min (4.4 min on stage 1, 2.50 on stage 2, and 2.75 on stage 3). The longest session was the third (session 2A), which lasted for 35.30 min (1.10 min on stage 1, 11.50 on stage 2, and 22.70 on stage 3). The second session on the same day (session 2B) was also long during stage three (1.65 min on stage 1, 1.70 on stage 2, and 20.55 on stage 3). The second session of the first day (session 1B) was longer than the first but still much shorter than sessions 2A and B (1.2 min on stage 1, 8.90 on stage 2, and 10.25 on stage 3).

First, we investigated the EEG oscillatory correlates of the meditation and its various stages. We analyzed the relative spectral power changes in each of the three stages of the meditation (as indicated by L.C.) in relation to baseline, which was at the start of the meditation. The results for the first session showed a robust increase in posterior gamma power (both gamma 1 and 2), which was larger toward stage 3 (**Figure 2**). There were no robust signal changes in other frequency bands. In order to test the consistency of this finding, we repeated our analysis for the second meditation session, and the results were quite similar, but with an added increase in frontal gamma power (**Figure 3**).

Next, we explored the consistency of this increase in gamma oscillations during stage 3 of meditation by analyzing gamma power changes in four meditation sessions. We extracted gamma power from occipital (O2, Oz, and O1) and frontal (AF4, AFz, and AF3) electrodes in each stage of meditation in all sessions (**Figure 4**). Occipital gamma power increased significantly in the stage 3 of meditation in all sessions regardless of the baseline level (**Figure 4A**), and associated scalp topographies of gamma power in each stage indicate that gamma oscillations increased especially in the occipital regions independent of how much it changed from the baseline. In addition, the relative change in the occipital area seems to be lower in the second session of the day (sessions B), which could be due to some carry over effects on the baseline coming from the previous meditation session.

Lia Chavez reported experiences of intense visual imagery only during stage 3 of meditation, stage that she denominated as analytical contemplation in which she reported to simply observe her inner visions. Due to our earlier findings on the gamma band, we conducted three other meditation sessions targeting the occurrence of the specific visions in relation to the oscillations. According to L.C., her visions occur as specific events and do not have a clear shape. In her words “*they are abstract, they look like volcanic explosions*”; they occur spontaneously with varying durations.

Meditation tACS Sessions

In order to understand the nature of these spontaneous creative visions and whether their content could be modulated by brain stimulation, we performed three brain stimulation conditions during the meditation of L.C. as follows: (1) Occipital gamma tACS (40 Hz at PO7 and PO8); (2) Alpha tACS (10 Hz at PO7 and PO8); and (3) Sham (30 s at 40 Hz on PO7 and PO8). These conditions were carried out in separate days and L.C. was blind to the stimulation condition. During each session, the first meditation practice was done simultaneously to the brain stimulation followed by the EEG procedures for a second session monitoring EEG without brain stimulation. During each stimulation condition, L.C. indicated the onset and the offset of each individual vision by pressing two buttons in the response box. We measured the frequency and duration of each vision during three different brain stimulation conditions.

After each session, L.C. reported her own experiences during the meditation. For meditation during alpha tACS, she experienced 175, in total, spontaneous visual imagery events. Importantly, if one vision was followed by another, there was no offset button press but another vision button press. The results indicated that her visions were about 15 s long (see **Figure 5** for the average duration in each condition). For the meditation following alpha tACS, she reported that the experienced visions were strange and excessively sharp as compared to her usual experiences of vision with low resolution. On a scale from 1 (undefined, low resolution) to 5 (sharp, well defined), she rated the visions as 5 (after the session, the visions in general, not each vision individually). She reported that “*I felt like the images were invading my thoughts, sharp, very sharp images.*” During the gamma tACS, she experienced 106 visions, reporting that

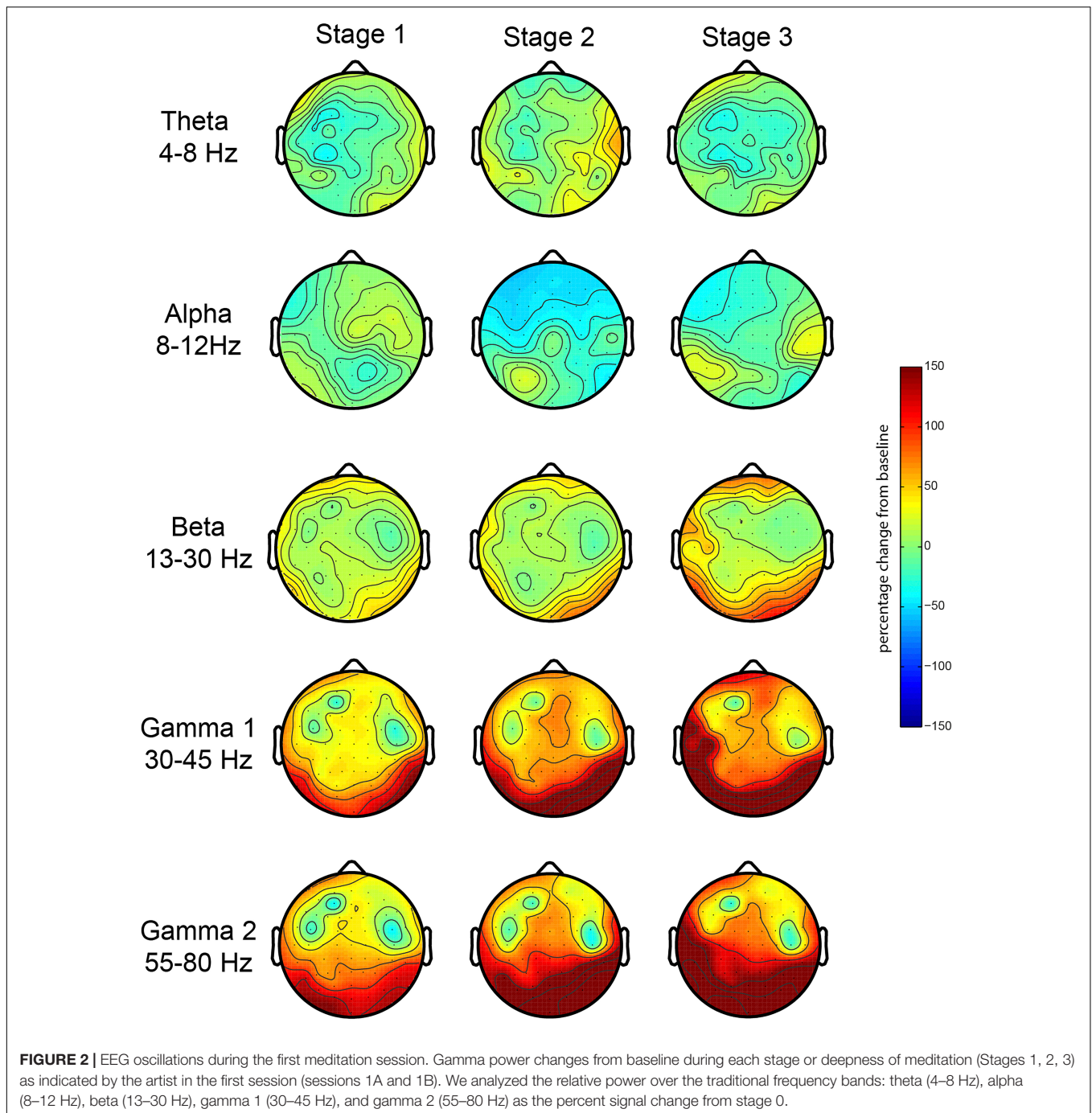


FIGURE 2 | EEG oscillations during the first meditation session. Gamma power changes from baseline during each stage or deepness of meditation (Stages 1, 2, 3) as indicated by the artist in the first session (sessions 1A and 1B). We analyzed the relative power over the traditional frequency bands: theta (4–8 Hz), alpha (8–12 Hz), beta (13–30 Hz), gamma 1 (30–45 Hz), and gamma 2 (55–80 Hz) as the percent signal change from stage 0.

“the images were more like what I usually experience during my meditation,” and rated the images (after the session was finished) as one (very blurred) on the sharpness scale mentioned previously. During the sham stimulation, she experienced 118 visions and rated them as a two on the referred scale. She reported that the visions were very similar to the ones she usually experiences. In order to statistically compare whether the brain stimulation had any effect on the duration of these visions (**Figure 5**), we conducted an one-way ANOVA comparing the *vision durations* (each individual vision as a data point,

meeting all assumptions for ANOVA) between sham, gamma and alpha tACS and found that the brain stimulation significantly modulated the duration of those visions [$F_{(2,395)} = 12.39$, $p < 0.001$]. *Post hoc* contrasts (*Bonferroni corrected*) showed that the visions were significantly shorter during alpha tACS than gamma ($p < 0.001$) and sham ($p < 0.001$) brain stimulation. There was no difference between gamma and sham stimulation ($p = 0.958$). Therefore, these findings show that alpha tACS targeted at the occipital region can modulate the duration of visual imagery during meditation.

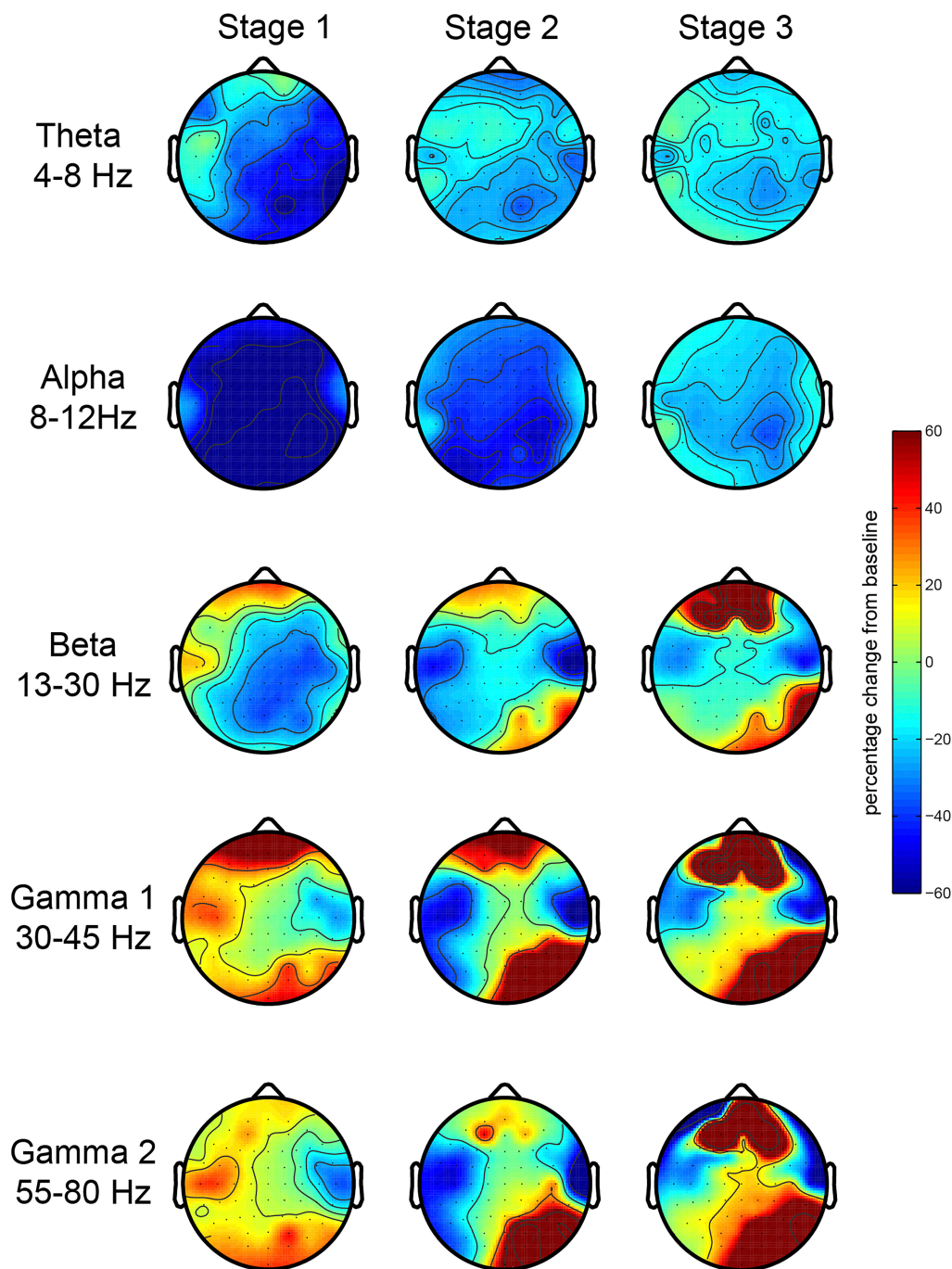
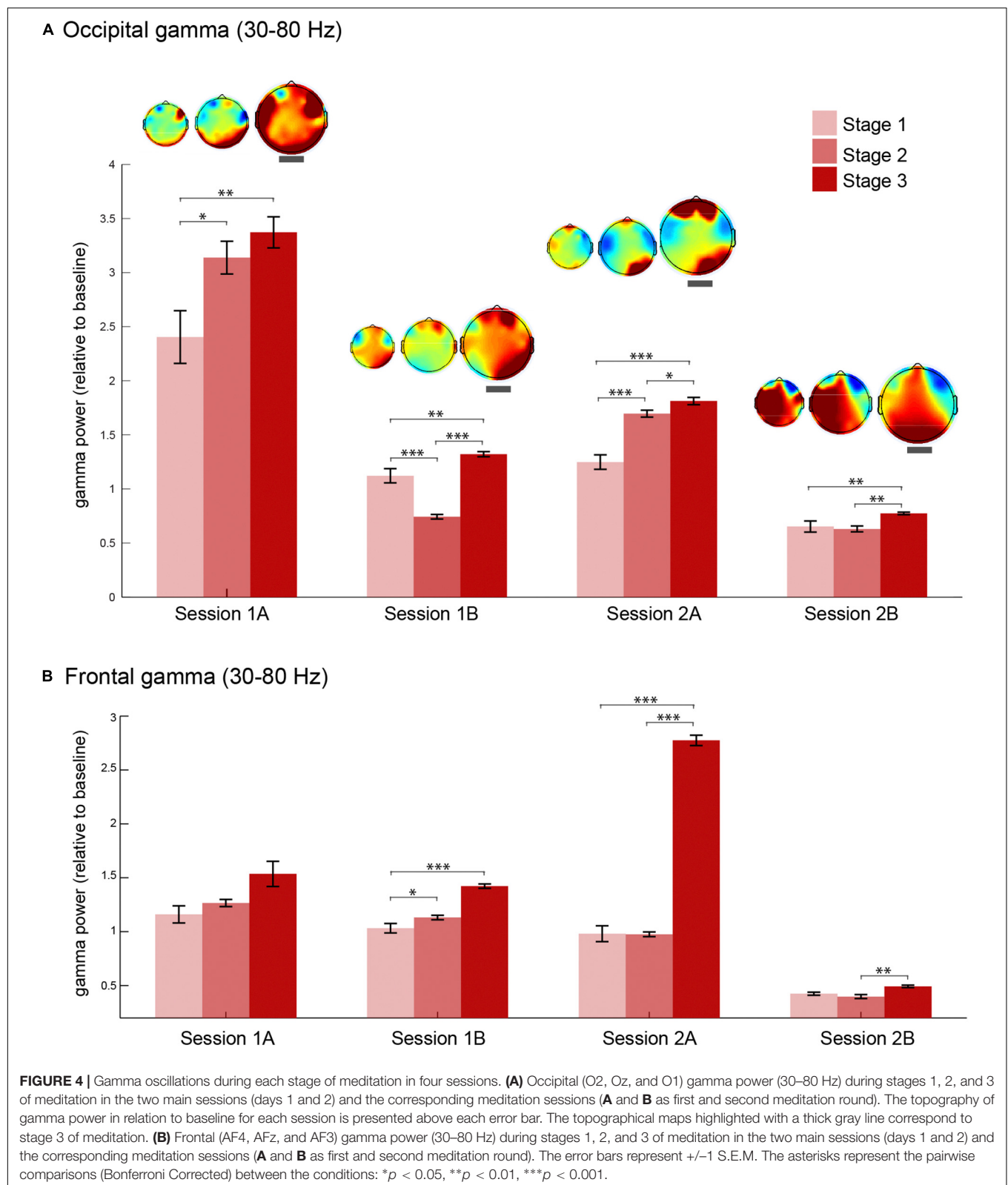


FIGURE 3 | EEG oscillations during the second meditation session. The figure displays the gamma power changes from baseline during each stage or deepness of meditation (Stages 1, 2, 3) as indicated by the artist in the second session (sessions 2A and 2B). We analyzed the relative power over the traditional frequency bands: theta (4–8 Hz), alpha (8–12 Hz), beta (13–30 Hz), gamma 1 (30–45 Hz), and gamma 2 (55–80 Hz) as the percent signal change from stage 0.

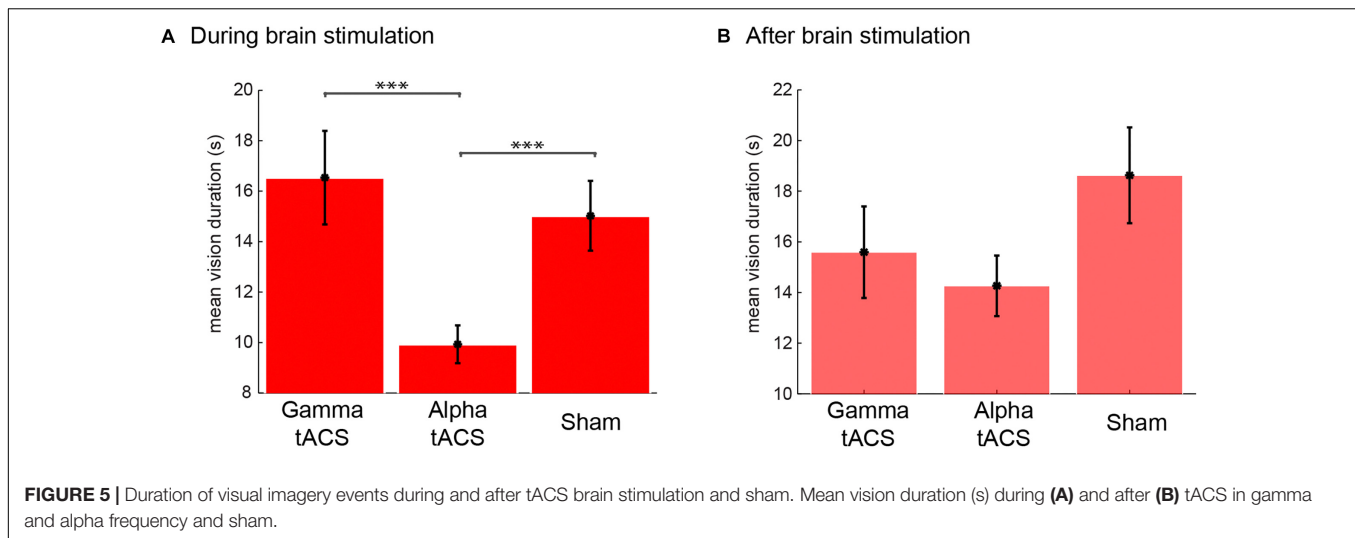
Immediately following the meditation session with brain stimulation, the EEG cap and electrodes were set up and a second meditation session was recorded in each stimulation day. There was a 35-min gap for setting up the EEG cap between the end of the stimulation and the following meditation with EEG. Although there was a trend for lower vision duration in

the session following gamma and alpha tACS (**Figure 5B**), the effect was not statistically significant [$F_{(2,330)} = 2.14$, $p = 0.120$]. However, we observed a larger number of visions following alpha tACS ($n = 141$), and the same number of visions for gamma tACS ($n = 96$) and sham ($n = 96$). *Post hoc* contrasts (*Bonferroni corrected*) between vision durations between conditions revealed



no significant differences between any of the conditions in the post stimulation session. Interestingly, after the session following the alpha tACS stimulation, L.C. reported that her visions were

more normal than during the stimulation although still sharper than usual (rated as a 3 from 1 to 5). In the sessions following gamma brain stimulation, the subject reported that the visions



were similar to what she usually experiences. However, she reported that she was very tired on the session following the sham stimulation (especially due to train delays she faced in the morning). She reported that her meditation was not successful because she was feeling fatigued. She reported that the imagery she experienced during that EEG session was “*not interesting, it was just like unconscious junk, it did not feel like proper imagery and it was sharper than usual.*” Therefore, in the EEG following sham stimulation, the nature of the imagery she reported was very distinct from the usual inspiring visual imagery she experiences during stage 3 of meditation.

Meditation EEG Following tACS: Occipital Gamma During Visions

Considering that her technique (see “Methods”) is slightly different in stages 1 and 2 as it includes the usage of stabilizing strategies such as repetition of mantra, there is a possibility that the increase in occipital gamma is a result of a change in activity during stage 3, in which she reports a pure contemplative state. Therefore, we asked whether gamma was indeed related to the intense visual imagery experienced by L.C. In order to understand the visions she experienced during the meditation, we segmented the data of stage 3 into 2 s epochs according to her indication of whether she was or not experiencing a vision. We focused on gamma power over occipital electrodes since this was the main oscillatory correlate of her meditation during stage 3. We observed that right occipital gamma power was higher during visions compared to no visions following gamma and alpha tACS, but not following sham (Figure 6). After sham, gamma power increased over the frontal but not the right occipital electrodes. Interestingly, this was the session that L.C. reported a high level of fatigue and the lowest quality of visual imagery as she described as “unconscious junk.”

Because L.C. was experiencing visions most of the time during stage 3, we limited the number of vision epochs in the analysis by randomly selecting the same number of no vision trials for each condition (following gamma, alpha, and sham tACS). We

compared right (P8, P10, PO4, PO8, and O2) occipital gamma power (30–80 Hz) during epochs with vs. without visions in stage 3 after gamma, alpha, and sham stimulation conditions (2×3 ANOVA). The three conditions differ in terms of relative gamma power [$F_{(2,322)} = 106.47, p < 0.001$], which could be related to the quality of the meditation experience during each day. Importantly, our results showed that gamma power was significantly higher while she was experiencing visual imagery [effect of visual imagery: $F_{(1,322)} = 29.39, p < 0.001$], but that interacted with stimulation condition [$F_{(2,322)} = 9.99, p < 0.001$] as the difference between vision and no vision was significant only following gamma and alpha tACS, but not following sham. Instead, there was an increase in gamma power in the temporal and frontal areas during stage three of sham. We conducted the same analysis using gamma power from fronto-temporal electrodes (F3, F4, T7, T8, C5, C6 – Figure 6B) and observed that there was significant increase in gamma power over the fronto-temporal areas following sham [effects of condition: $F_{(2,322)} = 54.94, p < 0.001$], which interacted with visual imagery [$F_{(2,322)} = 10.84, p < 0.001$] and it was larger for visions compared to no visions [$F_{(2,322)} = 16.93, p < 0.001$]. The significant contrasts (Bonferroni corrected) can be observed on Figure 6.

In order to control for other differences in gamma oscillations at stage 3 caused by the simple button press rather than meditation depth and also to test whether the general effects of meditation depth (as in Figure 4) were still present in this session, we compared the stages without separating visions and non-visions in the EEG sessions following alpha and gamma tACS (Figure 7). The results revealed gamma increased in the occipital electrodes in both sessions. The increase was also observed on the left temporal after gamma tACS and over the prefrontal electrodes following alpha tACS. In order to compare the conditions, we conducted a 2 (session: post-gamma vs. post-alpha tACS) \times 3 (stage: 1, 2, or 3) ANOVA using the occipital power values (O2, Oz, and O1) as dependent variable. The results confirmed a significant effect for stage [$F_{(2,1223)} = 5.25, p = 0.005$] but not for session [$F_{(1,1223)} = 0.46, p = 0.499$], nor interaction [$F_{(2,1223)} = 0.59, p = 0.557$]. Post hoc comparisons revealed that

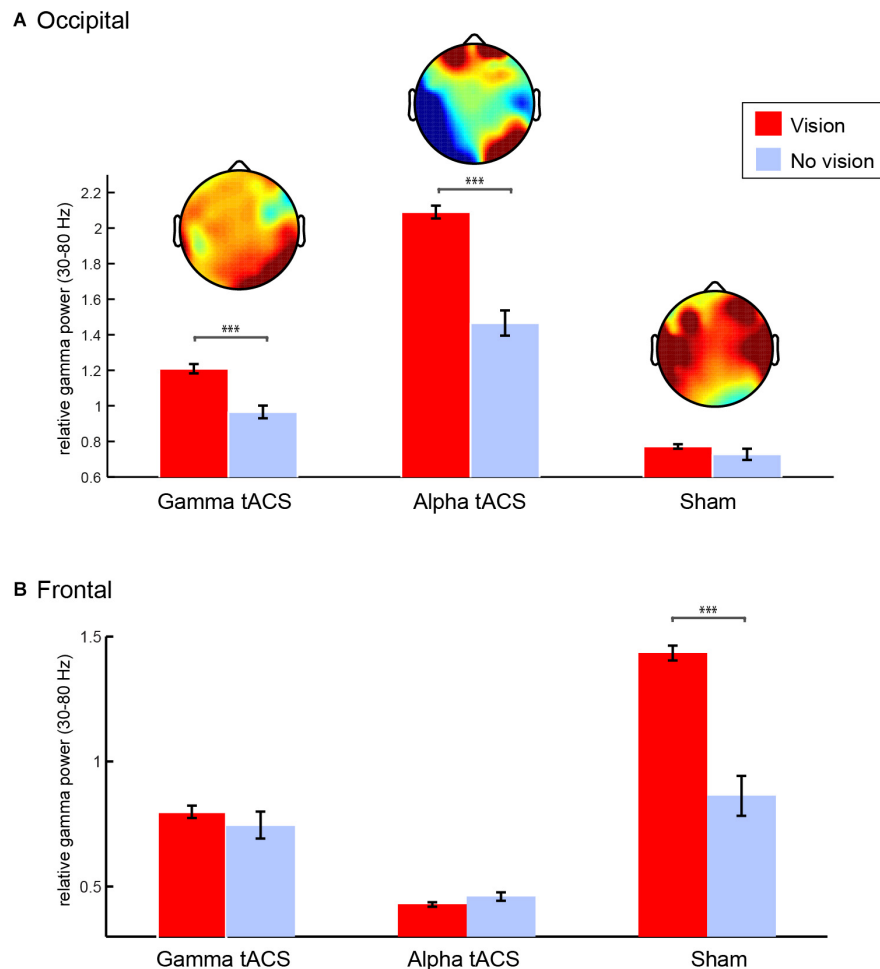


FIGURE 6 | Gamma oscillations during spontaneous visual imagery. **(A)** Right occipital gamma (35–80 Hz) power (P8, P10, PO4, PO8, O2) during visions (red) vs. no visions (blue) during meditation stage 3 following gamma, alpha, and sham tACS conditions. **(B)** Frontal gamma (35–80 Hz) power (F3, F4, T7, T8, C5, C6) during visions (red) vs. no visions (blue) during meditation stage 3 following gamma, alpha, and sham tACS conditions. Error bars represent ± 1 S.E.M across epochs (vision and no vision epochs of 2 s each). *** $p < 0.001$.

gamma increased during stage 3 more than in stage 1 and 2 ($p < 0.005$, Bonferroni corrected). In order to check whether the increase over the frontal electrodes was also significant we conducted the same factorial ANOVA using gamma power over prefrontal (AF4, AFz, and AF3) as the dependent variable. The results showed no significant effects for session, stage neither interaction between these two ($p > 0.9$), suggesting that this increase was not consistent. We also extracted the power values for the left temporal since we observed an increase during stage 3 after gamma stimulation. We conducted the same factorial ANOVA which revealed no effects for stage, session or interaction ($p > 0.5$), which suggests that the average increase in left temporal gamma was not statistically significant.

DISCUSSION

Although the spontaneous visual imagery during meditation has been previously reported in the literature as a well-known

correlate of meditation (Lo et al., 2003; Lindahl et al., 2013), very little is known about its potential neural correlates. Here we reported a case study of spontaneous visions occurring during deep stages of meditation that are considered as the source of creative inspirations for a reputed professional performing artist. In summary, our study has three main contributions that can help advance our understanding of the interface between creativity, visual imagery, and meditation: (1) we observed that occipital gamma increases in deep stage of meditation and that this increase is built up in the lower stages; (2) we showed that occipital gamma, as observed in the deep stage of meditation, is higher when L.C. experiences spontaneous visual imagery during meditation; (3) for the first time, we demonstrated that it is possible to interfere with visual imagery contents during meditation by delivering tACS to the occipital cortex. Further, by acquiring fine-grained details of the different stages of meditation over repeated sessions, our findings offer a novel insight into the meditation from first person experience, as recommended previously (Thomas and Cohen, 2014).

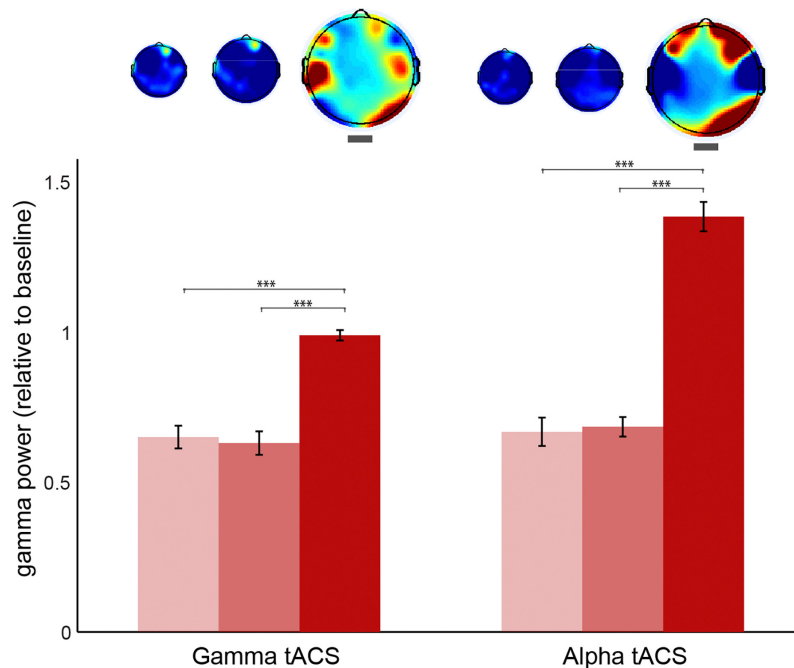


FIGURE 7 | Gamma oscillations during each stage of meditation following gamma and alpha stimulation sessions. Occipital (O2, Oz, and O1) gamma power (30–80 Hz) during stages 1, 2, and 3 of meditation (from brighter to darker red) following gamma and alpha tACS sessions. The power values are relative to baseline (stage 0). The error bars represent ± 1 S.E.M. The asterisks represent the pairwise comparisons (Bonferroni Corrected) between the conditions: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

It is important to notice that in this study we did not test the association between spontaneous visual imagery and the quality of her visual art pieces generated from it. We focused on understanding whether we could consistently identify the oscillatory correlates of her spontaneous visual imagery and whether we could modify it by stimulating specific brain oscillations using transcranial brain stimulation. Nonetheless, the participant reported that very detailed visual imagery, as she experienced during the alpha tACS session, was not very useful for her creative production. According to her, those images were “too detailed” to be used in her artwork which is abstract (see example on **Figure 1B**). Notwithstanding our limitations as a case study, this might suggest that further studies can potentially look into the association between the contents of spontaneous visual imagery and creativity in visual artists. It has been suggested that the dynamics and the contents of spontaneous thoughts or mind-wandering are important for creativity (Christoff et al., 2016). Our findings seem to suggest that the contents of the spontaneous visual imagery may be important for the creative production in visual arts. As a first case study on this, we suggest that future researchers explore ways of inducing spontaneous visual imagery in artists and investigate its association with creative outputs in visual arts.

Regarding the brain oscillations during meditation, we observed an increase in gamma power (> 30 Hz). This increase was higher during stage 3 of meditation and stronger in

the occipital areas. Importantly, the subject was able to consistently show a similar pattern (occipital gamma increase) in several meditation sessions. This increase was sometimes accompanied by an increase in prefrontal gamma oscillations, but that was a less consistent pattern across sessions. We did not observe any effects in the lower frequency bands as some studies had found (Berman and Stevens, 2015). However, there are a number of studies which found increases in gamma oscillations during meditation (Lehmann et al., 2001; Lutz et al., 2004; Cahn et al., 2010; Davis and Hayes, 2011; Berkovich-Ohana et al., 2012; Hauswald et al., 2015; Braboszcz et al., 2017) and at rest in experienced meditators (Lutz et al., 2004; Berkovich-Ohana et al., 2012; Thomas and Cohen, 2014). In particular, occipital gamma has been observed during meditation (Lutz et al., 2004; Cahn et al., 2010; Braboszcz et al., 2017). To our knowledge, no study so far has connected the documented increase in occipital gamma during meditation with spontaneous visual imagery. The process of “seeing things” during meditation has been reported as a relatively common phenomenon amongst meditators often reported as encounters with light but has never been investigated using neuroimaging methods (Lo et al., 2003; Lindahl et al., 2013).

We observed that gamma increases were most consistent in stage 3 of meditation, which challenges the idea that gamma represents the general meditation techniques rather than meditative states as suggested recently (Berman and Stevens, 2015). Instead, we suggest that gamma oscillation, in

particular occipital gamma, is one of the main mechanisms behind deep meditation states. Importantly, we found this occipital gamma to be associated with creative visual imagery experienced by our subject. When the subject was in deep meditation (stage 3), but was not experiencing these visions, this neural signature was reduced. Interestingly, in day 6 we did not observe such occipital increase in a session that the participant reported the imagery content as “junk” or not proper meditation visual imagery content. One important question is how these visions emerge: does gamma increase because of the visions or a higher gamma triggers spontaneous visual imagery? Our results seem to suggest, by looking at the previous meditation stages, that gamma starts increasing before the visions are experienced, even before the subject reaches a deep meditative state. However, this is only a hypothesis and it requires further investigation and this study has only investigated this process in a single participant, so we can only speculate that heightened occipital gamma may trigger spontaneous visual imagery in experienced meditators, which could explain the well-documented encounters with light experienced by meditators (Lindahl et al., 2013). This phenomenon has not been addressed in the neuroimaging literature up until now and it requires further investigation.

Finally, we demonstrated that by applying tACS to the occipital cortex, bilaterally, it is possible to modulate the content and duration of such visions. Unknown to the participant, alpha tACS seems to have led to unusually sharp visual imagery content (high spatial frequency) with shorter duration, whereas gamma and sham did not modulate vision duration or content. It has previously been shown (Fründ et al., 2007) that occipital alpha increases when processing sharper visual stimuli (>5 cycles per degree – cpd – of visual arc) whereas gamma increases when processing lower resolution images (<5 cpd). Previous studies have shown that it is possible to interfere with visual processing by entraining alpha (Brignani et al., 2013) and gamma (Helfrich et al., 2014; Janik et al., 2015) rhythms in the visual cortex by tACS. In our study, rather than interfering with visual perception, alpha tACS seems to have modified the visual imagery contents during meditation by making them sharper (according to L.C. subject report – blinded to the stimulation condition), which is consistent with the role of alpha oscillations in processing higher spatial frequency visual stimuli (Fründ et al., 2007). This result also evidences that spontaneous visual imagery might rely on similar neural correlates as veridical vision, in the same fashion as the observed shared processes between imagined or learned images and their actual visual processing (Albers et al., 2013; Luft et al., 2015). Considering that in this study we only had a single tACS session for each frequency, these results must be interpreted with caution since there are several factors which could have affect the visual experience of our xparticipant.

On the other hand, gamma tACS did not seem to affect the meditation experience, which was reported by the subject as her usual meditation experience. This finding might have occurred due to the already heightened gamma oscillations

in the occipital cortex during meditation since it was found that the tACS effects on the oscillations are highly dependent on endogenous brain states on the stimulated frequency (Neuling et al., 2013). In particular, it was observed that individualized alpha peak (IAF) tACS stimulation only enhanced IAF power under conditions in which the endogenous IAF power was naturally low (Neuling et al., 2013). Therefore, it could be that by stimulating gamma frequency, which is naturally higher during her meditation stage 3, we were not able to enhance them. Further studies could explore the possibility of enhancing gamma oscillations for triggering spontaneous visual imagery in beginners or intermediate meditators since they still have not developed such a self-induced high gamma power increase during meditation. Questions such as whether this would increase the depth of meditation or elicit creative visual imagery are of interest. Another interesting possibility is to induce occipital gamma in order to trigger spontaneous visual imagery for creative purposes in artists.

Some limitations of this study must be kept in mind. First, although the gamma band correlates of stage 3 were replicated in different days/sessions, we did not test the tACS effects in a second experiment. Therefore, the effects of alpha tACS on vision duration and precision should be interpreted with caution. Future studies investigating visual imagery on meditators could test this protocol further in a different order. Second, we cannot rule out differences between visions and non-visions in other frequency bands. In this study, we focused on the gamma band because it was the neural correlate of stage 3 meditation. Other frequency bands could be affected by the visions, but they were not investigated in the present study. Third, we understand that as a single case report, there is a need of more studies investigating the neural correlates of visual imagery and how that can affect the creative process in visual arts. Our study provides preliminary evidence that spontaneous episodes of visual imagery experienced in deep meditation are associated with higher occipital gamma, but new studies with other participants having similar experiences are needed. Importantly, our study raises the possibility of using brain stimulation for interfering with visual imagery contents, a relevant new venue to explore to modulate meditation experience.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Goldsmith's ethics committee with written informed consent from our research participant Lia Chavez. She gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Goldsmiths ethics committee.

AUTHOR CONTRIBUTIONS

CL, JB, and MB designed the research. CL, IZ collected the data. CL analyzed the data. CL, JB, MB, and IZ wrote the manuscript.

FUNDING

This was supported by the CREAM project funded by European Commission grant 612022. This publication reflects the views only of the authors, and the European Commission cannot be held responsible for any use which may be made of the information contained therein. MB was supported by the ESRC (ES/K00882X/1).

REFERENCES

- Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., and de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Curr. Biol.* 23, 1427–1431. doi: 10.1016/j.cub.2013.05.065
- Ambrus, G. G., Al-Moyed, H., Chaieb, L., Sarp, L., Antal, A., and Paulus, W. (2012). The fade-in – Short stimulation – Fade out approach to sham tDCS – Reliable at 1 mA for naïve and experienced subjects, but not investigators. *Brain Stimul.* 5, 499–504. doi: 10.1016/j.brs.2011.12.001
- Antal, A., and Paulus, W. (2013). Transcranial alternating current stimulation (tACS). *Front. Hum. Neurosci.* 7:317. doi: 10.3389/fnhum.2013.00317
- Austin, J. H. (2003). *Chase, Chance, and Creativity: The Lucky Art of Novelty*. Cambridge, MA: MIT Press.
- Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W., Franklin, M. S., and Schooler, J. W. J. P. S. (2012). Inspired by distraction: mind wandering facilitates creative incubation. *Psychol. Sci.* 23, 1117–1122. doi: 10.1177/0956797612446024
- Berkovich-Ohana, A., Glicksohn, J., and Goldstein, A. (2012). Mindfulness-induced changes in gamma band activity - implications for the default mode network, self-reference and attention. *Clin. Neurophysiol.* 123, 700–710. doi: 10.1016/j.clinph.2011.07.048
- Berman, A. E., and Stevens, L. (2015). EEG manifestations of nondual experiences in meditators. *Conscious Cogn.* 31, 1–11. doi: 10.1016/j.concog.2014.10.002
- Braboszcz, C., Cahn, B. R., Levy, J., Fernandez, M., and Delorme, A. (2017). Increased gamma brainwave amplitude compared to control in three different meditation traditions. *PLoS One* 12:e0170647. doi: 10.1371/journal.pone.0170647
- Brignani, D., Ruzzoli, M., Mauri, P., and Miniussi, C. (2013). Is transcranial alternating current stimulation effective in modulating brain oscillations? *PLoS One* 8:e56589. doi: 10.1371/journal.pone.0056589
- Cahn, B. R., Delorme, A., and Polich, J. (2010). Occipital gamma activation during *Vipassana meditation*. *Cogn. Process.* 11, 39–56. doi: 10.1007/s10339-009-0352-1
- Chodron, T. (2007). *Guided Meditations on the Stages of the Path*. Ithaca, N.Y: Snow Lion Publications.
- Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N., and Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nat. Rev. Neurosci.* 17, 718–731. doi: 10.1038/nrn.2016.113
- Davis, D. M., and Hayes, J. A. (2011). What are the benefits of mindfulness? A practice review of psychotherapy-related research. *Psychotherapy* 48, 198–208. doi: 10.1037/a0022062
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Dentico, D., Cheung, B. L., Chang, J. Y., Guokas, J., Boly, M., Tononi, G., et al. (2014). Reversal of cortical information flow during visual imagery as compared to visual perception. *Neuroimage* 100, 237–243. doi: 10.1016/j.neuroimage.2014.05.081
- Fründ, I., Busch, N. A., Körner, U., Schadow, J., and Herrmann, C. S. (2007). EEG oscillations in the gamma and alpha range respond differently to spatial frequency. *Vis. Res.* 47, 2086–2098. doi: 10.1016/j.visres.2007.03.022
- Hauswald, A., Ubelacker, T., Leske, S., and Weisz, N. (2015). What it means to be Zen: marked modulations of local and interareal synchronization during open monitoring meditation. *Neuroimage* 108, 265–273. doi: 10.1016/j.neuroimage.2014.12.065
- Helfrich, R. F., Knepper, H., Nolte, G., Strüber, D., Rach, S., Herrmann, C., et al. (2014). Selective modulation of interhemispheric functional connectivity by HD-tACS shapes perception. *PLoS Biol.* 12:e1002031. doi: 10.1371/journal.pbio.1002031
- Janik, A. B., Rezlescu, C., and Banissy, M. J. (2015). Enhancing anger perception with transcranial alternating current stimulation induced gamma oscillations. *Brain Stimul.* 8, 1138–1143. doi: 10.1016/j.brs.2015.07.032
- Laakso, I., and Hirata, A. (2013). Computational analysis shows why transcranial alternating current stimulation induces retinal phosphenes. *J. Neural Eng.* 10:046009. doi: 10.1088/1741-2560/10/4/046009
- Laczo, B., Antal, A., Niebergall, R., Treue, S., and Paulus, W. (2012). Transcranial alternating stimulation in a high gamma frequency range applied over V1 improves contrast perception but does not modulate spatial attention. *Brain Stimul.* 5, 484–491. doi: 10.1016/j.brs.2011.08.008
- Lehmann, D., Faber, P. L., Achermann, P., Jeanmonod, D., Gianotti, L. R., and Pizzagalli, D. (2001). Brain sources of EEG gamma frequency during volitionally meditation-induced, altered states of consciousness, and experience of the self. *Psychiatry Res.* 108, 111–121. doi: 10.1016/S0925-4927(01)00116-0
- Lindahl, J. R., Kaplan, C. T., Winget, E. M., and Britton, W. B. (2013). A phenomenology of meditation-induced light experiences: traditional buddhist and neurobiological perspectives. *Front. Psychol.* 4:973. doi: 10.3389/fpsyg.2013.00973
- Lo, P.-C., Huang, M.-L., and Chang, K.-M. (2003). EEG alpha blocking correlated with perception of inner light during zen meditation. *Am. J. Chin. Med.* 31, 629–642. doi: 10.1142/S0192415X03001272
- Luft, C. D., Meeson, A., Welchman, A. E., and Kourtzi, Z. (2015). Decoding the future from past experience: learning shapes predictions in early visual cortex. *J. Neurophysiol.* 113, 3159–3171. doi: 10.1152/jn.00753.2014
- Luft, C. D. B., Zioga, I., Thompson, N. M., Banissy, M. J., and Bhattacharya, J. (2018). Right temporal alpha oscillations as a neural mechanism for inhibiting obvious associations. *Proc. Natl. Acad. Sci. U.S.A.* 115, E12144–E12152. doi: 10.1073/pnas.1811465115
- Lutz, A., Greischar, L. L., Rawlings, N. B., Ricard, M., and Davidson, R. J. (2004). Long-term meditators self-induce high-amplitude gamma synchrony during mental practice. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16369–16373. doi: 10.1073/pnas.0407401101
- Lutz, A., Lachaux, J. P., Martinerie, J., and Varela, F. J. (2002). Guiding the study of brain dynamics by using first-person data: synchrony patterns correlate with ongoing conscious states during a simple visual task. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1586–1591. doi: 10.1073/pnas.032658199
- Neuling, T., Rach, S., and Herrmann, C. S. (2013). Orchestrating neuronal networks: sustained after-effects of transcranial alternating current stimulation depend upon brain states. *Front. Hum. Neurosci.* 7:161. doi: 10.3389/fnhum.2013.00161
- Singer, I. (2011). *Modes of Creativity: Philosophical Perspectives*. Cambridge, MA: MIT Press.
- Taylor, M. (2011). “Imagination,” in *Encyclopedia of Creativity*, eds M. A. Runco, S. R. Pritzker, and R. Reiter-Palmon (Cambridge, MA: Academic Press), 637–643.
- Thomas, J. W., and Cohen, M. (2014). A methodological review of meditation research. *Front. Psychiatry* 5:74. doi: 10.3389/fpsyg.2014.00074
- Ward, T. B. (1994). Structured imagination: the role of category structure in exemplar generation. *Cogn. Psychol.* 27, 1–40. doi: 10.1006/cogp.1994.1010

ACKNOWLEDGMENTS

We would like to thank Lia Chavez for her important contributions to this research. Lia Chavez's participation in the research was active and went beyond being the traditional guinea pig model. She was vital in initiating and giving direction to the research and its continued development. We would also like to thank Tegan Penton for helping in the brain stimulation sessions.

Ward, T. B. (1995). "What's old about new ideas?" in *The Creative Cognition Approach*, eds S. M. Smith, T. B. Ward, and R. A. Finke (Cambridge: MIT Press), 157–178.

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 Luft, Zioga, Banissy and Bhattacharya. 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.

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