

SOCIALLY SITUATED? EFFECTS OF SOCIAL AND CULTURAL CONTEXT ON LANGUAGE PROCESSING AND LEARNING

EDITED BY: Pia Knoeferle, Ramesh Kumar Mishra and Marcela Pena
PUBLISHED IN: Frontiers in Psychology and Frontiers in Communication





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 978-2-88974-975-1

DOI 10.3389/978-2-88974-975-1

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

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

SOCIALLY SITUATED? EFFECTS OF SOCIAL AND CULTURAL CONTEXT ON LANGUAGE PROCESSING AND LEARNING

Topic Editors:

Pia Knoeferle, Humboldt University of Berlin, Germany

Ramesh Kumar Mishra, University of Hyderabad, India

Marcela Pena, Pontificia Universidad Católica de Chile, Chile

Citation: Knoeferle, P., Mishra, R. K., Pena, M., eds. (2022). Socially Situated? Effects of Social and Cultural Context on Language Processing and Learning. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-975-1

Table of Contents

- 05 Editorial: Socially Situated? Effects of Social and Cultural Context on Language Processing and Learning**
Pia Knoeferle, Ramesh Kumar Mishra and Marcela Peña
- 09 The Social Meaning of Contextualized Sibilant Alternations in Berlin German**
Melanie Weirich, Stefanie Jannedy and Gediminas Schüppenhauer
- 27 Common Ground Information Affects Reference Resolution: Evidence From Behavioral Data, ERPs, and Eye-Tracking**
Maria Richter, Mariella Paul, Barbara Höhle and Isabell Wartenburger
- 44 Morphological Metaphor Mapping of Moral Concepts in Chinese Culture**
Yingjie Liu, Kang Li, Lina Li, Jing Zhang, Yuerui Lin, Baxter DiFabrizio and He Wang
- 54 Explicit (Not Implicit) Attitudes Mediate the Focus of Attention During Sentence Processing**
Oleksandr V. Horchak and Margarida Vaz Garrido
- 69 The Role of Context in Language Development for Children With Autism Spectrum Disorder**
Patricia Sánchez Pérez, Anders Nordahl-Hansen and Anett Kaale
- 75 Do the Eyes Have It? A Systematic Review on the Role of Eye Gaze in Infant Language Development**
Melis Çetinçelik, Caroline F. Rowland and Tineke M. Snijders
- 91 Mood Induction Differently Affects Early Neural Correlates of Evaluative Word Processing in L1 and L2**
Johanna Kissler and Katarzyna Bromberek-Dyzman
- 107 Learning to Use Narrative Function Words for the Organization and Communication of Experience**
Gregoire Pointeau, Solène Mirliaz, Anne-Laure Mealier and Peter Ford Dominey
- 122 Effects of Dispositional Affect on the N400: Language Processing and Socially Situated Context**
Veena D. Dwivedi and Janahan Selvanayagam
- 132 Reading About Us and Them: Moral but no Minimal Group Effects on Language-Induced Emotion**
Björn 't Hart, Marijn Struiksma, Anton van Boxtel and Jos J. A. van Berkum
- 148 Face-to-Face Communication in Aphasia: The Influence of Conversation Partner Familiarity on a Collaborative Communication Task**
Willemijn Doedens, Arpita Bose, Lydia Lambert and Lotte Meteyard
- 172 Ladies First: Gender Stereotypes Drive Anticipatory Eye-Movements During Incremental Sentence Interpretation**
Ernesto Guerra, Jasmin Bernotat, Héctor Carvacho and Gerd Bohner
- 182 Attriters and Bilinguals: What's in a Name?**
Federico Gallo, Keerthi Ramanujan, Yury Shtyrov and Andriy Myachykov

188 *Integration of Social Context vs. Linguistic Reference During Situated Language Processing*

Katja Maquate and Pia Knoeferle

211 *Cognitive Processes of ESL Learners in Pragmatic Role-Play Tasks in Academic Settings*

Nick Zhiwei Bi



Editorial: Socially Situated? Effects of Social and Cultural Context on Language Processing and Learning

Pia Knoeferle^{1,2,3*}, Ramesh Kumar Mishra⁴ and Marcela Peña^{5,6}

¹ Institute of German Studies and Linguistics, Humboldt-Universität zu Berlin, Berlin, Germany, ² Berlin School of Mind and Brain, Berlin, Germany, ³ Einstein Center for Neuroscience Berlin, Berlin, Germany, ⁴ Center for Neural and Cognitive Sciences, University of Hyderabad, Hyderabad, India, ⁵ Cognitive Neuroscience Laboratory, Pontificia Universidad Católica de Chile, Santiago, Chile, ⁶ Centro Nacional de Inteligencia Artificial, Santiago, Chile

Keywords: social context, language, biases, morality, language prestige, individual differences, affect, aphasia (language)

Editorial on the Research Topic

Socially Situated? Effects of Social and Cultural Context on Language Processing and Learning

An increasing number of findings in psycholinguistics, neurolinguistics, and the cognitive sciences suggest that the (non-linguistic) socially interpreted and cultural context can influence language processing and learning. That context could include a speaker's (or bystander's) actions, facial expressions, voice or gaze, and gestures, among others. Given the wide variety of contexts (e.g., real-world, videos, still photographs, drawings, narratives, newspaper texts, poems, movies), and of writers, speakers/comprehenders (of different ages, gender, social status, linguistic, and cultural background), the extent of such social and cultural effects on language processing and learning remains unclear, partially because of the complexity to model their interactions (applying different methodologies). The submissions to this Research Topic help delineate the interplay of the socially interpreted and cultural context for language processing and learning/development.

OPEN ACCESS

Edited and reviewed by:

Manuel Carreiras,
Basque Center on Cognition, Brain
and Language, Spain

*Correspondence:

Pia Knoeferle
pia.knoeferle@hu-berlin.de

Specialty section:

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

Received: 15 January 2022

Accepted: 15 February 2022

Published: 28 March 2022

Citation:

Knoeferle P, Mishra RK and Peña M
(2022) Editorial: Socially Situated?
Effects of Social and Cultural Context
on Language Processing and
Learning. *Front. Psychol.* 13:855733.
doi: 10.3389/fpsyg.2022.855733

1. LANGUAGE DEVELOPMENT AND LEARNING

A number of contributions to this Research Topic focus on language development and learning and showcase the role of the social and cultural context in this domain. From the contributions we can take away the insight that social context is highly diverse. Modeling its influence on learning likely involves a sophisticated understanding of interdependence between language, the world that an individual perceives and acts in, and characteristics of the individual (e.g., mood and language background).

Pointeau et al. examine how robots learn about causal and temporal event relations. They use a corpus of speech from humans describing simple human-robot interactions. Algorithms are used to extract how function words link events to one another (e.g., statistics on how words like “because”, or “then” link different elements in a situation model). The recovered statistics serve as the input for robot learning (how to interact in question-answering dialogue and how to produce narratives). Other research focuses on language development in infants and with a focus on what extra-linguistic cues like a speaker's gaze can contribute to word learning: Cetincelink et al. review evidence suggesting eye-gaze is important for vocabulary development (also longitudinally), word-object reference, object, and speech processing. One key insight from this review is that word-object mapping benefits from eye-gaze. But it remains to be seen to what extent eye-gaze constitutes

a mechanism of enhancing learning even more broadly. In addition to benefits of language-world mapping for infants, second language learners also benefit from (the wider pragmatic) context. Zhiwei Bi used a role-play task involving, for instance, requesting a reference letter from a professor or scheduling a meeting with peers. The goal was to ascertain second language learners' strategies. Qualitative analyses of retrospective verbal reports uncovered a range of language-related strategies for speech acts like requests (e.g., comprehending, linking to prior experiences, or pragmatic awareness). Insights into the neurocognitive basis of language learning in social contexts come from a contribution by Kissler and Bromberek-Dyzman. They compared L1 vs. L2 comprehenders' emotion word processing as a function of mood. Mood induction influenced the very first moments of emotion word processing (stronger left-lateralization of mean amplitude in event-related brain potential negativities around 100 ms, the so-called "N1" for happy compared to sad mood). Regardless of first or second language background, valence modulation in the earliest moments emerged following happy but not sad mood induction; language background did modulate later, lexical-semantic processes. A comprehender's second language can also serve as context for, and influence, the decay of native language ("attrition"). A review by Gallo et al. focuses on how first-language attrition happens, why it occurs, and who attritors are (e.g., immigration history, linguistic behaviors, internal neurocognitive states). The authors argue that looking at attrition within the context of the bilingual mind can improve our understanding of how socio-cultural factors (that likely go hand in hand with immigration and first-language attrition) can modulate linguistic processing.

2. EXPECTATIONS AND BIASES

Social context not only influences language development and learning but also moment-to-moment language processing. The contributions to this Research Topic convey the insight that social context of different sorts seems to enable the formation of expectations. Expectation-formation has been called into question for lexical-level cues (see DeLong et al., 2017; Ito et al., 2018; Nieuwland et al., 2018; Nieuwland, 2019) and against this backdrop the convergence in anticipatory social context effects is striking. A close look at the results clarifies, however, that anticipation is not the same for all world-language relations. Emergence of biases in expectations is also striking. Guerra et al. provide evidence for expectations by exploring the role of gender attitudes and stereotypes in language comprehension. Using visual-world eye-tracking¹, they found that participants, when inspecting a display with several images, exploited the verb in German sentences to anticipate a character (out of two) that fit with verb gender-stereotype knowledge. These effects were asymmetric

in that they were larger for female than male stereotypes but they did not vary between participants depending on their gender (e.g., sexist) attitudes. Anticipation biases also emerged in research examining "common ground" effects (knowledge shared by a speaker and an addressee). Richter et al. used a referential communication game to examine whether common ground is integrated quickly or with delay, involving effort, and whether what is common knowledge vs. privileged (for just the listener) is integrated at the same time. Objects were shown visually either in privileged or common ground, and for critical trials, common ground was relevant but objects in privileged ground had to be ignored. The results from a range of methods, among them eye-tracking and event-related brain potentials, suggested that common ground had early effects, enabling the anticipation of objects; but conflicting information in privileged ground had the potential to interfere. Maquate and Knoeferle complement these insights into common ground effects with a comparison of how referentially-mediated action depictions and non-referentially mediated emotional cues (speaker face emotion) modulate visual attention and language comprehension. Effects of depicted actions were replicated and were pervasive; speaker face emotion effects were, by contrast, more subtle, highlighting the need to pay attention to the relation between language and the world in deriving predictions of (social) context effects.

3. MORALITY, LANGUAGE PRESTIGE, AND REGISTER

Context takes many facets, including that of morality and prestige. 't Hart et al. examined how facial muscle movements in response to emotionally valenced sentences vary depending on whether a sentence protagonist was described as morally good or bad (more frowning upon reading *Mark is angry* vs. *Mark is happy* when Mark was pitched as a good person, but not when he was characterized as bad). Whether the participant was part of the same group as Mark or not (in-group vs. out-group) did not modulate the frowning of the target expression, *Mark is angry/happy* (more frowning muscle activity emerged in the corrugator supercilii for angry than happy sentences). Being part of a social and age group, did, by contrast, affect the performance of participants in a language task (an implicit association test). Weirich et al. reported implicit association test results that differed for older and younger language users and for multi- vs. mono-ethnic groups. Participants, for instance, classified words (of different language register) as having bad or good valence. Experiment 1 contrasted a standard German with input labeled as a low-register German variety and Experiment 2 used the same stimuli but labeled them as standard German vs. standard French. Results revealed that older language users had a stronger association of low-register words with negative valence words when listening to low-register variety, and a smaller effect (less negative attitude) when listening to a French-native learner of German. A group of younger participants of mono-ethnic origin, by contrast, had no effect of language

¹ Visual-world eye-tracking is a paradigm in which images are presented together with spoken language; visual interrogation of the images is tracked and has been found to be guided by the interaction of visual context with language (Cooper, 1974; Tanenhaus et al., 1995; Huettig et al., 2011).

variety; but the younger multi-ethnic participants linked the low-register variant to negative-valence words more strongly for French than the low-register German language variety. Liu et al. contributed a study on Chinese (examining the identification of written Chinese characters that were either morally positive or negative valenced); they reported faster identification when positive if the characters were oriented upright or facing to the right. By contrast, immoral characters were identified faster when these were distorted or presented with a left rotation. The authors interpreted these results as suggesting that physical cues like the direction of orientation contribute to encoding social concepts in language, in line with Conceptual Metaphor Theory.

4. INDIVIDUAL DIFFERENCES: ATTITUDE, AFFECT SCORES, AND LANGUAGE BACKGROUND

The attitude participants have toward contextual information also plays an important role in language processing. This insight emerges from the results of Horchak and Vaz Garrido. The authors focused comprehenders' attention on environmental issues (*noticed ... garbage*) compared with on the emotions of protagonists (*got upset with the garbage*) or on actions (*picked up... garbage*) and assessed to what extent such a focus influenced a comprehender's sentence ratings (seriousness of the issue) and verification (fit with the picture), also as a function of participants' environmental awareness. The results suggested that focus on a topic like environmental issues boosted ratings of sentences with environmental focus, and that participants' environmental awareness can modulate attention in sentence processing. Modulation of language processing by participant characteristics was also observed in Dwivedi and Selvanayagam. They replicated increased mean amplitude negativities "N400" to semantically mismatching vs. matching words in a sentence context. Crucially, these effects were modulated by participants' affect score (PANAS). Larger N400 differences emerged for individuals with smaller negative affect scores, further highlighting the role of individual differences. Kissler and Bromberek-Dyzman reported individual differences for lexical-semantic processing as reflected in the N400, too. N400 mean amplitude differences were larger for second-language than first-language comprehenders. Together these findings highlight the role of individual differences as a modulating factor for context effects.

REFERENCES

- Cooper, R. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cogn. Psychol.* 6, 84–107.
- DeLong, K. A., Urbach, T. P., and Kutas, M. (2017). Is there a replication crisis? Perhaps. Is this an example? No: a commentary on Ito, Martin, and Nieuwland

5. CLINICAL CASES

Doedens et al. examined the role of context, and in particular familiarity with a communication partner in collaborative communication of aphasic patients compared with healthy controls. Measures of communicative efficiency like the time it took participants to complete the goal of the communicative task differed when comparing patients with aphasia to the controls. As instructors in the task, the patients were faster with an unfamiliar (vs. familiar) interlocutor (accuracy was unaffected by interlocutor familiarity). Healthy controls had higher accuracy when the partner was unfamiliar but reaction times were unaffected by the familiarity manipulation. In the listener role, patients showed a boost in accuracy for the unfamiliar interlocutor. A better understanding of how contextual factors influence communication in patients is the first step in intervention studies. A contribution that also speaks to this issue comes from Sanchez-Perez et al. who investigated vocabulary in 2 to 4-year-old children who were on the autism spectrum. They examined the children's vocabulary in at-home and pre-school contexts. Results suggest clear differences in vocabulary (size) across these two contexts, meaning that vocabulary size may be underestimated if only one context is considered.

6. SUMMARY STATEMENT

Social and cultural context influences language processing and learning during a lifespan with at least some variability across diverse language user groups.

AUTHOR CONTRIBUTIONS

PK conceptualized the editorial and provided a first draft and revised it after feedback. RM and MP provided feedback on the draft. All authors contributed to the article and approved the submitted version.

FUNDING

We acknowledge that this research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412, 416591334, Centro Nacional de Inteligencia Artificial CENIA, FB210017, and FINANCIAMIENTO BASAL PARA CENTROS CIENTIFICOS Y TECNOLOGICOS DE EXCELENCIA de ANID to MP.

(2016). *Lang. Cogn. Neurosci.* 96, 367–400. doi: 10.1016/j.neubiorev.2018.11.019

- Huetting, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* 137, 151–171. doi: 10.1016/j.actpsy.2010.11.003
- Ito, A., Pickering, M., and Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of

- English: a visual world eye-tracking study. *J. Mem. Lang.* 98, 1–11. doi: 10.1016/j.jml.2017.09.002
- Nieuwland, M. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neurosci. Biobehav. Rev.* 96, 367–400.
- Nieuwland, M., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., et al. (2018). Limits on prediction in language comprehension: a multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. *eLife* 7, e33468. doi: 10.7554/eLife.33468
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 632–634.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Knoeferle, Mishra and Peña. 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 Social Meaning of Contextualized Sibilant Alternations in Berlin German

Melanie Weirich*, Stefanie Jannedy* and Gediminas Schüppenbauer

Leibniz-Centre General Linguistics, Berlin, Germany

OPEN ACCESS

Edited by:

Ramesh Kumar Mishra,
University of Hyderabad, India

Reviewed by:

Alba Tuninetti,
Bilkent University, Turkey
Robin Dodsworth,
North Carolina State University,
United States

*Correspondence:

Melanie Weirich
weirich@leibniz-zas.de
Stefanie Jannedy
jannedy@leibniz-zas.de

Specialty section:

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

Received: 27 May 2020

Accepted: 14 September 2020

Published: 29 October 2020

Citation:

Weirich M, Jannedy S and
Schüppenbauer G (2020) The Social
Meaning of Contextualized Sibilant
Alternations in Berlin German.
Front. Psychol. 11:566174.
doi: 10.3389/fpsyg.2020.566174

In Berlin, the pronunciation of /ç/ as [ç] is associated with the multi-ethnic youth variety (*Kiezdeutsch*). This alternation is also known to be produced by French learners of German. While listeners form socio-cultural interpretations upon hearing language input, the associations differ depending on the listeners' biases and stereotypes toward speakers or groups. Here, the contrast of interest concerns two speaker groups using the [ç]–[ç] alternation: multi-ethnic adolescents from Berlin neighborhoods carrying low social prestige in mainstream German society and French learners of German supposedly having higher cultural prestige. To understand the strength of associations between phonetic alternations and social attributes, we ran an Implicit Association Task with 131 participants (three groups varying in age and ethnic background (mono- vs. multi-ethnic German) using auditory and written stimuli. In experiment 1, participants categorized written words as having a positive (*good*) or negative (*bad*) valence and auditory stimuli containing pronunciation variations of /ç/ as canonical [ç] (labeled *Hochdeutsch* [a term used in Germany for Standard German]) or non-canonical [ç] (labeled *Kiezdeutsch*). In experiment 2, identical auditory stimuli were used but the label *Kiezdeutsch* was changed to *French Accent*. Results show faster reaction times when negative categories and non-canonical pronunciations or positive categories and canonical pronunciations were mapped to the same response key, indicating a tight association between value judgments and concept categories. Older German listeners (OMO) match a supposed *Kiezdeutsch* accent more readily with negatively connotated words compared to a supposed French accent, while younger German listeners (YMO) seem to be indifferent toward this variation. Young multi-ethnic listeners (YMU), however, seem to associate negative concepts more strongly with a supposed French accent compared to *Kiezdeutsch*. These results demonstrate how social and cultural contextualization influences language interpretation and evaluation. We interpret our findings as a loss of cultural prestige of French speakers for the YMO group compared to the OMO group: younger urban listeners do not react differently to these contextual primes. YMU listeners, however, show a positive bias toward their in-group. Our results point to implicit listener attitudes, beliefs, stereotypes and shared world knowledge as significant factors in culturally- and socially situated language processing.

Keywords: sociophonetics, perception, social meaning, social context, IAT, fine phonetic detail, prestige, in-group – out-group

INTRODUCTION

In this study, we will show that listeners draw implicit associations between sub-phonemic variation or fine phonetic detail and evaluative categories in dependence to a speaker group that supposedly produced the speech form. Speech production is not merely a means of transporting propositional content, but also serves the construction of personas and reflects speakers' social identities. For simplicity reasons, here in this paper we use the binary concept of canonical vs. non-canonical, however, our conceptualization of individuals' speaking styles goes beyond the notion of complementary, binary or dichotomous categories such as formal vs. informal or casual, standard vs. non-standard, or read vs. spontaneous. Rather, we conceive speech with all its features and variants as a tool set from which speakers (sub)consciously select and chose from to position themselves in social space. Listeners pay attention to phonetic detail and either can or cannot interpret the social meaning of the variant(s).

However, many of the fine phonetic details observed in speech are produced by a speaker without much awareness. As the social dynamics change or the persona performed varies by situation or shifts over the course of a conversation, so too can the linguistic choices of the speaker. Hearers may then be in a position to draw meaningful associations between linguistic variants and social actors that use variants to create a personal style or to index a particular social persona, properties or stances. Speech researchers have long come to the realization that speakers adjust their speech in dependence to language external factors (Labov, 2001) such as the addressee (Bell, 1984), specifically in child- or animal directed speech (Burnham et al., 2002), speaker characteristics such as age (Eckert, 1989), gender identity (Weirich and Simpson, 2018), sexual orientation (Munson et al., 2006; Kachel et al., 2018), or the formality of the speech situation (Podesva, 2007). Also, a speaker's phonetic accommodation to a model talker is mediated by the other speaker's social identity and perceived attractiveness (Babel, 2012) or the participants personality (Lewandowski and Jilka, 2019). The degree of convergence has been found to be used to decrease or increase social distance (Giles, 1973; Giles et al., 1973; Bourhis and Giles, 1977). A speaker's perceived femininity or masculinity plays a role in perception (Johnson et al., 1999) as does a hearer's age (Jannedy and Weirich, 2014) or where the hearer believes the speaker is from (Niedzielski (1999), Hay and Drager (2010), Jannedy and Weirich (2014)).

The study of intra-speaker variation as a field of study has gained traction with the *Third Wave* in sociolinguistics (Eckert, 2012; Eckert and Labov, 2017) where studies focus on the speech styles of individuals as they maneuver social situations. While phonetic variation is inherently gradient, the occurrence of a phonetic form can statistically be used more often in one social situation or by one social group compared to another. Eckert and Labov (2017, p. 481) explicitly say that "Having no referential function, a phonological variable is free to take on purely contextual meaning as it ranges within the limits set by neighboring phonemes." So, once a variant has been collectively recognized by listeners as belonging to a specific speech style,

context, or social group, it can be used to index membership in this group or to index a specific context.

For example, in German, the phonological category /ç/ has two allophonic variants (throughout this manuscript we will refer to the default or canonical German fricative variant as [ç] and to the non-canonical alternant as [ç̥]). The alternation of /ç/ to a phonetic variant ranging acoustically between the palatal fricative [ç] and the post-alveolar fricative [ʃ], i.e., symbolically represented as the alveopalatal fricative [ç̥] in the youth-style multi-ethnolect Kiezdeutsch as spoken in Berlin – the sociolect investigated in the present study – serves to index membership and the identification with the multi-cultural Berlin districts Kreuzberg, Neukölln or Wedding as "their" neighborhood in the speech of adolescents (Jannedy et al., 2015). As an extension to that, for some people, it indexes a young, hip, multiethnic and urban *street* identity. For simplicity reasons, we will refer to this variant of /ç/ as [ç̥] as our work on the acoustic phonetic properties of these variants suggests that [ç̥] differs from both [ç] and [ʃ] in several spectral parameters such as center of gravity (COG, cf. section "Acoustic Characteristics of Stimulus Materials") and discrete cosine transformation coefficients (Jannedy and Weirich, 2017).

Work on language stereotypes, attitudes (Johnson et al., 1999; Niedzielski, 1999; Hay and Drager, 2010; Jannedy and Weirich, 2014), and person perception (Scherer, 1972; Schirmer, 2019) has shown that it is possible to put speakers in mind-sets in which to perceive speech. We will exploit this finding for our study, too by making listeners believe that a voice they hear either belongs to a French speaker learning German or a German speaker of Turkish decent, both groups for which stereotypes exist in dominant German language ideology (Plewnia and Rothe, 2009; Jannedy and Weirich, 2014; Jannedy et al., 2019). While German spoken with a French accent supposedly is the most favored foreign accent by Germans and generally evokes positive ratings (Plewnia and Rothe, 2009), German spoken with features believed to be of multi-ethnic origin, i.e., Arabic or Turkish seems to polarize or evoke negative stereotypes (Wiese, 2015). It is our assumption that neither the positive nor the negative associations with these two varieties of German are conscious so as to be deliberately mediated in public, and moreover, vary between individuals influenced by social factors such as age or personal background.

In this work, we investigate the relative strength of implicit associations between speech variants and evaluative categories in the context of a fictitious French vs. multi-ethnic speaker group. We have borrowed the experimental technique of the Implicit Association Task (IAT) (Greenwald et al., 1998, 2003) from psychology as we are interested in the immediate and unmediated reactions to a speech stimulus and the social information that a phonetic shape invokes. According to the *Social Connotation Hypothesis* (van Bezooijen, 2002), hearers' evaluations and reactions to language stimuli depend on social attributes and inferences drawn based on the supposed values, intentions, and attitudes that are associated with a speech variant. The IAT paradigm allows for collecting reaction time data which reveals how strongly a listener associates a specific variant with a value judgment. Due to the structure of the task, participants

should not be able to disguise which associations come closest to their own, thus revealing their implicit, rather than their explicit, associations.

There is much evidence that encountered language input is stored in memory along with social information. These remembered instances (“exemplars”) are stored in a multi-dimensional space representing a cognitive map (Goldinger, 1998; Barlow and Kemmer, 2000). Watanabe et al. (2001) stipulate that the human cognitive system is built in such a way that learning in general and by extension perceptual learning works with and without awareness through rapid adaptation to the surrounding environment. This is also corroborated by the work on sound acquisition and acquisition trajectories (Foulkes and Docherty, 2006), and sound change (Harrington, 2006). Harrington et al. (2019) for example showed that during the linguistic isolation of multi-dialectal English-speaking staff during the winter months in Antarctica, their speech begins to converge toward each other, averaging out differences in vowel production (also see Eckert, 2019 on the spread of sound change). Results like these imply that groups of speakers that have a sense of belonging to the same social group and probably identifying with it, may develop speech patterns that can eventually be interpreted in meaningful ways by hearers. Applying this train of thought to our study, we are interested in the associated information that is stored with a phonetic variant in the context of two distinct speaker groups and the way associated and implied social information shapes the attitudes associated with specific speech forms.

An example widely discussed in the literature (Campbell-Kibler, 2011, 2012 and references therein) is the English verbal suffix <-ing> which is realized as either [ɪŋ] with a velar nasal in many standard varieties of English or as [ɪŋ̹] with an alveolar nasal in non-standard varieties. It is argued that the choice of this variant by a speaker in speech production has a communicative intention (see Eckert, 2008; Campbell-Kibler, 2011; Eckert, 2012 and others). The work by Campbell-Kibler (2010; 2011; 2012) shows that addressees derive social associations such as *educated* or *intelligent* from speech variants, yet, these interpretations are highly context-sensitive and dependent on a listener’s mood and the social perception of the speaker (Campbell-Kibler, 2008).

Our own work (Jannedy and Weirich, 2014) on the [ɕ] – [ʑ] alternation in the urban context of Berlin revealed an age-graded listener bias in the categorization of stimuli taken from a 14-step acoustic continuum ranging from /ɕ/ to /ʑ/ (where [ʑ] is located along the continuum) when co-presented with the name of a Berlin neighborhood (*Kreuzberg*) known for its multi-ethnic and multi-lingual population. In this classic categorical perception task, older (mean age: 50.7) and middle (mean age: 30.2) aged listeners were biased in their responses toward the non-canonical pronunciation variant [ʑ] in the context of the prime *Kreuzberg* (the multi-ethnic and multi-lingual Berlin district), whereas younger (mean age: 22.7) listeners seemed to have been free of or have undone this bias by selecting fewer [ʑ] tokens in this condition compared to a control condition where no additional information was presented.

These results suggest that listeners have conceptualized representations or stereotypes of what people from certain

neighborhoods sound like, affectively reacting to perceptual stereotypes and creating perceptual personas. Our results corroborate findings on the social association of the -in/-ing alternation in American English (Campbell-Kibler, 2010, 2011) whereby the canonical -ing pronunciation was associated with more intelligent/educated/articulate speaker types, while the -in pronunciation was perceived to sound less formal and less likely to be gay.

Results like these show that listeners tie pronunciation variants to social attributes, a connection that is undoubtedly learned. Studies show that listeners were not able to distinguish between the standard- and non-standard varieties of languages that were unknown to them (Giles et al., 1974, 1975; van Bezooijen, 1988) in terms of the perceived pleasantness or status, showing that there is no inherent value to one form over another. In other words, one variant is not more sophisticated than another variant, it is the implicit association of speech variants with assumed, associated or stereotyped social traits of speaker groups that lets members of a speech community form value judgments (cf. *Social Connotation Hypothesis*, van Bezooijen, 2002).

To gain a better understanding of these deeply rooted implicit associations that listeners have formed on variable speech production patterns and linked to learned, assumed or stereotyped social traits, we have used a method that measures the relative strength of association between two dichotomous concept categories. The IAT (Greenwald et al., 1998; Greenwald et al., 2003; Nosek et al., 2005) investigates the immediate and affective inferences that participants draw upon being prompted with (a set of) stimuli. Classic IAT experiments were used to show the closeness of implicit associations between concepts such as *male/female* and *science/humanities*, *black/white* and *good/bad*, or *skinny/fat* and *good/bad*. This experimental paradigm has also proven to be quite valuable for testing the tight association of phonetic forms with social meanings (see Campbell-Kibler, 2012). Campbell-Kibler (2012) has used the paradigm to show that experiment participants had an awareness of the -ing/-in variable when they were presented with it in writing, associating these forms with either professions (*white-collar/blue-collar*) or with regional accents (*southern/northern states*).

Pantos (2010) pioneered a multi-modal IAT-approach, presenting auditory and visual stimuli and a combination thereof. In follow-up work, Pantos and Perkins (2012) tested in an United States-American context the implicit association between pronunciation variants and positive and negative valence words and found an implicit bias in favor of United States-accented versus Korean-accented speech. Campbell-Kibler (2012) also successfully deployed an auditory paradigm and was able to show that the -ing/-in variables were implicitly associated with *northern* vs. *southern* accented speech, respectively. Nilsson et al. (2019) tested the differences in social meaning of Swedish /i/ in two rural areas. In one of their two test-sites, their results revealed a stronger implicit association of “damped” /i/ ([i]) with ruralness and cardinal /i/ with urbanity.

In this work, we will test the implicit associations of words ending in the German adjectival suffix <-ig> (produced with the German palatal fricative [ç] varying in pronunciation between non-canonical [ʑ] and canonical [ɕ]) and positive

and negative valence words (as representatives of positive and negative attitudes toward the variants and the speaker group using this variant).

Our first hypothesis is, that listeners will associate non-canonical pronunciations with negative values and canonical pronunciations with positive values. The additional aim is to exploit the similarity of the [ç] variant in the two distinct varieties of German *Kiezdeutsch* and *French Learners' German*. In Berlin *Kiezdeutsch*, this pronunciation variant is rather salient and prevalent in the speech of multi-ethnic youth and their peers from several districts in Berlin. There are differences as to how *Kiezdeutsch* is perceived in the urban population of Berlin: to younger speakers, the multi-ethnic urban variant seems more of a default pronunciation by now and is perceived as *street* (sociolect independent of the ethnic background of the speakers) which stands for young, hip and urban. In contrast, most of the older population and more conservative views published in the press view this sociolect as polarizing, uneducated and negative, and it is shunned upon by mainstream speakers of Berlin German (Heiser, 2014). However, a similar non-canonical like variant also exists in the foreign accented speech of French learners of German which is often seen as cute and endearing and used in TV-advertisements and which generally seems to evoke more positive associations (Giles and Niedzielski, 1998; Plewnia and Rothe, 2009).

In our study we test the implicit attitudes of listener groups varying in age and ethnicity toward identical speech items (a non-canonical pronunciation variant) in the two varying socio-culturally situated contexts of associating these speech forms with multi-ethnic speakers from Kreuzberg and with learners of German from France. Thus, our second hypothesis is, that the implicit association between the non-canonical pronunciation and the negative values is stronger in the *Kiezdeutsch* context than in the *French Learners' German* context. However, this bias might vary between different listener groups. We assume that social factors such as the age or ethnic background of a listener affects his/her sensitivity to the priming conditions and moreover, his/her attitudes toward the suggested speaker groups. Thus, different biases can be explained in terms of in-group and out-group behavior: the in-group (a cohort that a speaker or listener associates with) is generally evaluated more positively and carries covert prestige compared to the out-group, that a speaker or listener feels socially distanced from Tajfel and Turner (1986) unless the out-group carries high social and cultural prestige. Therefore, the second hypothesis is modified in such a manner that we predict to find our listener groups to vary in the strength of the associations between non-canonical pronunciations and negative values across the priming conditions with regard to their age and ethnic background. Listeners with a multi-ethnic background (similar to the presumed speaker in the *Kiezdeutsch* context) should show a stronger association between non-canonical pronunciations and negative values in the *French Learners German* context (out-group) than in the *Kiezdeutsch* context (in-group). Also, through language experience within the context of urban Berlin, younger (mono-lingual, mono-ethnic German) urban listeners to some degree have overcome their bias toward multi-ethnic

and multi-lingual speakers using non-canonical phonetic forms as they themselves perform *street*, resulting in a smaller bias toward non-canonical pronunciations in general independent of the speaker group.

Based on what we have learned from the literature, we presume that hearers notice and recognize fine phonetic detail and index, interpret and evaluate it differentially. We expect our data on the attitudes associated with differentially produced speech features to show that an identical speech variant (a) indexes and receives different social meanings in dependence to the presumed speaker group and (b) receives different social meanings in dependence to the specific hearer group. The study focusses on the saliency of variation in fine phonetic detail in social interpretation and stigmatization of a speaker while exploring the role of the implicit attitudes of different hearer groups interpreting the signal. Thus, it is original and novel insofar as it explores the variance in social meaning of fine phonetic detail in the confines of an urban space in Germany, exploring the role of differences in hearer characteristics (multi-ethnic young; mono-ethnic German and young; mono-ethnic German and older) as well as in- and out-group contextual primes (*Kreuzberg* vs. *French learner of German*).

We set out to test that a phonetic variant is contextualized as it is interpreted in line with usage- and experience-based approaches to language processing (cf. Goldinger, 1998; Barlow and Kemmer, 2000) and therefore, depending on an individual's experience with and attitude toward speech forms and speaker groups, the same phonetic variant can convey differences in social meaning.

MATERIALS AND METHODS

Acoustic Characteristics of Stimulus Materials

Since the acoustic differences of the auditory stimuli are quite minute yet very crucial, we will provide a short description of the spectral features and acoustic characteristics differentiating the German sibilants /ç/ and /ʃ/ and the pronunciation variant [ç̥]. Although Standard German contrasts three voiceless sibilants phonologically: the alveolar /s/, the postalveolar /ʃ/ and the palatal /ç/, many speakers of the middle German dialects and regiolects do not differentiate between /ç/ and /ʃ/ but use the alveo-palatal pronunciation variant [ç̥] instead. The same holds for multi-ethnic speakers of *Kiezdeutsch* and French learners of German (Wottawa et al., 2016; Jannedy and Weirich, 2017).

Analyses of the acoustic characteristics of different fricatives (Evers et al., 1998; Jongman et al., 2000; Gordon et al., 2002; Nowak, 2006; Cheon and Anderson, 2008; Li et al., 2011), reveal that the acoustic differentiation between /ʃ/, /ç̥/, and /ç/ is rather difficult. For Polish, Czaplicki et al. (2016) found spectral peak (the frequency with the highest amplitude of the spectrum) and center of gravity (CoG, the mean central frequencies for the entire spectrum) to be good predictors to separate a new variant of an alveopalatal fricative from the standard Polish counterparts. Bukmaier and Harrington (2016) investigated the Polish retroflex, dental, and alveopalatal sibilants /ʃ̐ ʃ̐ ʃ̐/ and

while these three fricatives are quite distinct articulatorily, they are difficult to separate acoustically. For German, Jannedy and Weirich (2017) investigated the acoustic difference between /ç/ and /ʃ/ in three speaker groups with varying contrast realizations (one of them being *Kiezdeutsch* speakers from Berlin). While perceptually and acoustically the contrast was lost in the speech of many *Kiezdeutsch* speakers, listeners reliably differentiated these two fricatives in minimal pairs produced by mono-ethnic German speakers from Berlin.

For a visualization of the acoustic differences in this three-way contrast, **Figure 1** shows the spectral shape of [ç ʃ] produced by a female speaker of Northern Standard German. The two sounds [ç] and [ʃ] correspond to the phonemic fricative categories of Standard German, while [ç] is somewhat intermediate between /ç/ and /ʃ/ sounds found in the multi-ethnic urban Berlin variety *Kiezdeutsch* and in French learners' German. The left plot of **Figure 1** visualizes the difference between [ç] (blue) and [ç] (red), the right plot shows the difference between [ç] (red) and [ʃ] (black). The [ç] sound has more energy in the higher frequency range above 5,000 Hz than [ç] and [ʃ]. It also shows a broader band of frequencies with high energy, while both [ç] and [ʃ] reveal clearer peaks but vary in the frequency of this peak: it is higher for [ç] than for [ʃ] which is due to a lengthening of the vocal tract through labialization in [ʃ]. For fricatives, in general, the frequency range with the highest energy is influenced by the place of articulation: the more back the sound is produced, the lower is the CoG. Through the labialization of the alveo-palatal sound /ʃ/ the vocal tract between the lips and the place of constriction is lengthened and thus CoG decreases with [ʃ] having the lowest values. Since the merged sound [ç] is produced without lip rounding (in contrast to [ʃ]) and further front than [ç], its spectrum has more energy in higher frequencies and thus has a higher CoG ([ç] = 4,494 Hz, [ç] = 5,566 Hz, [ʃ] = 2,575 Hz).

In the experiments conducted in the present study we used stimuli that varied between canonical and non-canonical pronunciations of German, with the non-canonical pronunciation referring to both the *Kiezdeutsch* variety and the French learner variety of German. Therefore, the auditory stimuli used consist of two pronunciation variants of the adjectival suffix <-ig> in German, i.e., [ç] for the canonical realization, and [ç] for the non-canonical realization.

Overview of the Study

With a first group of 40 participants, we conducted an online-rating study on the auditory test items to explore if listeners (a) rate the manipulated items to be naturalistic and (b) if and how the pronunciation variations are associated with particular or specific persona types in terms of age or education.

In the main experiment, the IAT, 131 participants were asked to match the presented auditory test stimuli with either a positive or negative valence word. In line with prior IAT results where negative concepts were strongly associated with racial traits that were deemed as less desirable, we hypothesize that non-canonical pronunciations are more strongly associated with negatively connotated words and canonical pronunciations with positive valence words. In general, a non-canonical form is assumed to be evaluated as flawed and bad since it is perceived as deviating from

the norm. Thus, overall, reaction times should be faster when the common assumptions are met: when the (negatively connotated) non-canonical pronunciation (categorized as either *Kiezdeutsch* or *French Accent*) and the negative attribute category (i.e., *Bad*) share a response key and when the canonical pronunciation and the positive valence words are mapped to the same response button. A pattern of this type is indicative of an implicit bias of the respondent (Nosek et al., 2005). The robustness of such a mapping is calculated in the form of a single *D*-score per respondent (the greater the bias, the faster the reaction times and the higher the score).

In addition, the particular associations evoked by a non-canonical form are highly dependent on the interpreter, his/her background, stereotypes and beliefs. Thus, the same non-canonical form can be considered as more or less negatively or even positively valenced depending on the attributed prestige of the speaker group associated with the form. Therefore, we hypothesize that the strength of the IAT effect in the two priming conditions *Kiezdeutsch* (condition 1) and *French Accent* (condition 2) is affected by the age and the ethnic background of the listeners due to different biases toward these varieties.

Online Rating Study

Forty listeners living in Berlin (14 male, 26 female, different from the IAT experiment) were asked in an online rating study to evaluate the naturalness of the stimuli, the supposed level of education and the inferred age of a speaker on a scale from 1 to 7 based on hearing a single word. The experiment was run using *Percy* (Draxler, 2011, 2014). The stimuli tested included the ones used in the IAT experiments and the same words produced by two additional female speakers (each word in the two pronunciation variants). In addition, some filler words were added which are not part of the analysis. The three aspects (*naturalness*, *education*, and *age*) were rated separately in three subsequent blocks and stimuli were randomized over participants. Participants could listen to each stimulus maximally three times. We also collected demographic data of the listeners regarding their age, gender, language background, education, city of birth and current residency as indicated by their postal code.

IAT Study

Participants

In total, 131 German speakers participated in this IAT study, they were distributed into three groups (see **Table 1**). There were two groups of younger speakers: one was comprised of German born multi-ethnic participants of Turkish or Arab (Lebanese and Palestinian) descent and the second group of younger speakers was comprised of mono-ethnic and mono-lingual Germans. The younger multi-ethnic German listeners (YMU) all were high-school students from Wedding, a multi-ethnic district of Berlin, and stated that they were German language dominant but also often had rudimentary skills in a language other than German. Their German showed several features of the *Kiezdeutsch* variety such as the /ç/- /ʃ/ alternation (Dirim and Auer, 2004; Jannedy and Weirich, 2014), which is neither stigmatized nor recognized amongst them. Younger mono-ethnic German listeners (YMO) were beginning first semester students at Berlin universities

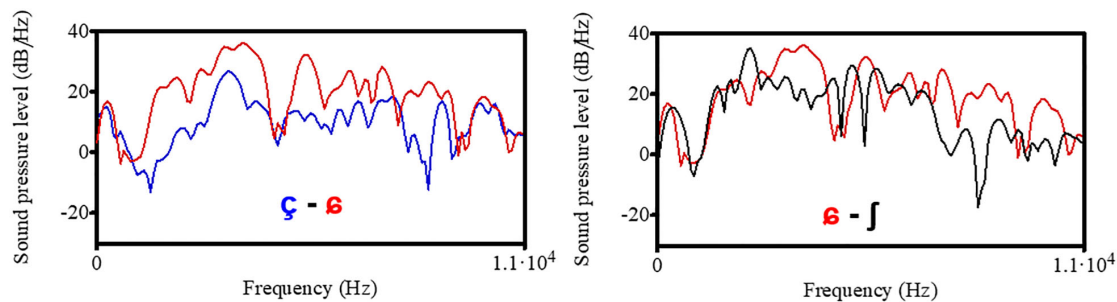


FIGURE 1 | Spectral shapes of the three fricatives [ç ç ʃ] produced by a female speaker of Northern German Standard. Different fricatives are marked by different colors.

TABLE 1 | Number of participants separated by listener groups with information on ethnic background, gender, and age.

Listener group		No (m/f)	Mean age in years (SD)
YMU	Younger M ulti-ethnic German	42 (12/30)	20.67 (4.65)
YMO	Younger M ono-ethnic German	43 (10/33)	25.02 (4.67)
OMO	Older M ono-ethnic German	46 (12/34)	51.78 (8.45)

who were either born and raised in Berlin or lived there for a significant amount of time of their life. The third group of participants were older mono-ethnic mono-lingual Germans who were born in Berlin or had lived or worked there for over 25 years. We refrained from explicitly asking participants about their familiarity with the concept of *Kiezdeutsch*, since we did not want to prime listeners in any direction. However, in Berlin, the concept of *Kiezdeutsch* is well-known and it can be assumed that everyone living in Berlin for a certain amount of time as our participants have, has at least heard about it in the news, recognizes it when hearing speakers in the tram or even knows someone using it. The same holds for the concept of *French accent* since it is widely used in mainstream media, e.g., in TV advertisement.

Given these three groups, we were able to compare differences due to ethnic background (mono- vs. multiethnic) and age (younger vs. older) (see **Table 1**). Participants were randomly distributed into the two different conditions (*Kiezdeutsch* vs. *French Accent*) of the IAT experiment.

Methodology and Explanation of Implicit Association Tasks (IAT)

A computer-based IAT requires participants to match stimuli such as orthographically or auditorily presented words with *attribute* or *concept categories* by pressing a button on the keyboard. In our case, words were rendered in the two pronunciation variations *canonical* (=Standard German which we refer to as *Hochdeutsch*) vs. *non-canonical* [(1) *Kiezdeutsch* (KD, condition 1) or (2) *French Accent* (FR, condition 2)]. It should be noted that the label *Hochdeutsch* which we used in the experiment does not contain the evaluative bias of the word *standard* as in Standard German.

These had to be matched with the *attribute categories* having a psychological valency of either good vs. bad. In other words, pronunciation variants had to be matched to the two language variants (canonical /non-canonical) and words with a positive or negative valence had to be matched to attributes (good/bad).

The trials were divided into seven blocks (see **Figure 2** and **Table 2**): In the first block, participants learn that *Hochdeutsch* pronunciations are mapped to a key (E) on the left of the keyboard while non-canonical German pronunciations are mapped to a key (O) on the right of the keyboard because the concept category *Hochdeutsch* (cf. **Figure 2**) appears in the left corner of the computer screen, effectively mapping to the left-hand response key E, and *Kiezdeutsch* (or *French Accent*) appears in the right corner, mapping the right-hand to the response key O. In the second block participants learn that words with a positive valence (good = “gut”, cf. **Figure 2**) like “*wundervoll*” (wonderful) or “*Freude* (joy) displayed in the middle of a computer screen, are mapped to the same button on the left side of the keyboard while words with a negative (bad = “schlecht”) valence (like *evil* or *failure*) are mapped to the button on the right side of the keyboard. In the third block, participants are confronted with either a written or an auditory stimulus while simultaneously seeing the label for a concept category like *Hochdeutsch* and an attribute category like *good* mapped to the left button of the keyboard. The label for the concept category *Kiezdeutsch* (or *French Accent*) were mapped to the right button just as the attribute category *bad*. For each sorting operation, the participant’s reaction time is logged. As mentioned above, the hypothesis is that combinations like *canonical variety* + *good* and *non-canonical variety* + *bad* are perceived as congruent and thus generate faster and more immediate reactions in comparison to cases where the implicit bias is violated and non-congruent (*Hochdeutsch* + *bad* and *non-canonical variety* + *good*). In other words, a faster reaction time indicates a stronger association between the paired categories (cf. Nosek et al., 2005).

In the third and also in the fourth block, all four categories appeared combined, pairing a concept category and an attribute category with one response key each. These two blocks constitute the congruent test cases (according to our hypothesis) and contain one half of the experimental trials from which the final IAT effect is calculated. The fifth block is a practice block for

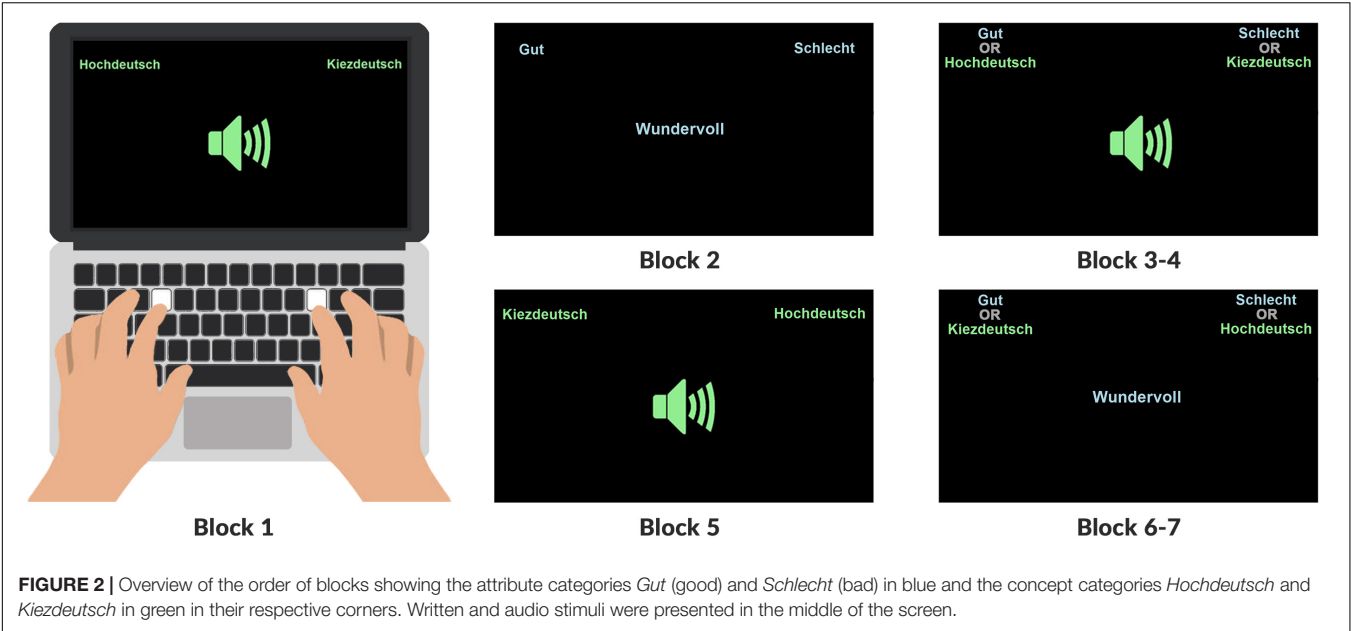


TABLE 2 | Overview and order of tasks in the IAT experiments for test order v1.

Block	Trials	Task	Left key	Right key
1	20	Practice: Audio stimuli only	Concept category <i>Hochdeutsch</i> (canonical /ç/)	Concept category <i>Kiezdeutsch</i> (or <i>Franz. Akzent</i>) (non-canonical /ç/)
2	20	Practice: Written stimuli only	Positive valence word <i>good</i>	Negative valence word <i>bad</i>
3	20	Test: Audio and written stimuli combined	<i>Hochdeutsch</i> and <i>good</i>	<i>Kiezdeutsch</i> (or <i>Franz. Akzent</i>) and <i>bad</i>
4	40	Test: Audio and written stimuli combined	<i>Hochdeutsch</i> and <i>good</i>	<i>Kiezdeutsch</i> (or <i>Franz. Akzent</i>) and <i>bad</i>
5	40	Practice reversed: Audio stimuli	<i>Kiezdeutsch</i> (or <i>Franz. Akzent</i>)	<i>Hochdeutsch</i>
6	20	Test reversed: Audio and written stimuli combined	<i>Kiezdeutsch</i> (or <i>Franz. Akzent</i>) and <i>good</i>	<i>Hochdeutsch</i> and <i>bad</i>
7	40	Test reversed: Audio and written stimuli combined	<i>Kiezdeutsch</i> (or <i>Franz. Akzent</i>) and <i>good</i>	<i>Hochdeutsch</i> and <i>bad</i>

For v2, the concept categories *Kiezdeutsch*/*Franz. Akzent* and *Hochdeutsch* switched keys resulting in incongruent test blocks first (3 and 4) and congruent test blocks second (6 and 7).

audio stimuli again, which introduces a crucial manipulation: while the attribute category mapping is kept constant throughout the experiment, the concept category mapping learned in the previous blocks is inverted by switching the position of the concept category labels (e.g., *Kiezdeutsch*/*French Accent* now maps to the left key and *Hochdeutsch* maps to the right key), effectively leading to an incongruent (according to our hypothesis) and therefore supposedly more difficult matching task. The number of trials in the fifth block is increased (40 instead of 20 trials in the practice blocks 1 and 2 before) to compensate for the learned mapping reinforced by all preceding trials (Nosek et al., 2005).

Blocks 6 and 7 combined all four category labels again while maintaining the incongruent category labels from block 5. These final two blocks provide the other half of the experimental trials needed for calculating the IAT effect size (called *D-score*). For each participant in the study, a single *D-score* value

is calculated. *D-scores* are computed as the mean difference between test blocks divided by the overall standard deviation of latencies. A detailed account of the scoring algorithm can be found in Greenwald et al. (2003). A *D-score* close to zero means no IAT effect at all. A positive *D-score* reveals a closer association between non-canonical pronunciations and “bad” and canonical pronunciations and “good” (in line with our hypothesis), while a negative *D-score* shows a closer association between canonical pronunciations and “bad” and non-canonical and “good” (opposed to our hypothesis).

In order to avoid block order effects, the starting position of the concept category labels was counterbalanced across participants. Thus, the order of test blocks – congruent to our hypothesis (*Hochdeutsch* + *good* and *non-canonical German* + *bad*) and incongruent with our hypothesis (*non-canonical German* + *good* and *Hochdeutsch* + *bad*) – was varied between participants. Half of the participants did the congruent

test blocks first (as described above), the other half did the incongruent test blocks first, leading to two versions of the experiments (v1 and v2, cf. **Table 2**). Participants were randomly assigned to the different experiment conditions (KD and FR) and order versions (v1 and v2) resulting in 10–12 participants in each of different listener groups (YMO, YMU, and OMO). Note that for participants in the FR condition all instances of *Kiezdeutsch* were replaced with *French Accent*.

The experiment was run on a Lenovo IdeaPad U330 laptop with $1,366 \times 768$ screen resolution using PsychoPy2 v1.85.3 (Peirce and MacAskill, 2018). For presentation of the auditory stimuli, Sennheiser HD590 headphones were used. The order of presentation for auditory and visual stimuli was randomized for each block across all participants. Overall, the experiment took approximately 20–30 min per participant including a questionnaire about some metadata of the participants.

Materials

The viability of using auditory stimuli in an IAT paradigm was first shown by Vande Kamp (2002) and has since been used in a variety of linguistic studies (Pantos, 2010; Campbell-Kibler, 2012, 2013; Pantos and Perkins, 2012; Loudermilk, 2015; Hilton et al., 2016; Leinonen, 2016; Llamas et al., 2016; Rosseel et al., 2018). For the current study, a female native mono-ethnic German speaker from Berlin (age 27) read 6 German adjectives ending in the syllable <-ig>: *einzig* “solely,” *fertig* “ready,” *mehlig* “floury,” *nussig* “nutty,” *körnig* “grainy,” *bündig* “concisely” in two different versions with two different pronunciation variants: *Hochdeutsch* and *Kiezdeutsch* (condition 1), doubling up as French learners’ accent in German (condition 2). As described above, in *Hochdeutsch*, the final sound is pronounced as a voiceless palatal fricative [ç], while it is pronounced as [ç̥] in *Kiezdeutsch* and in French learner varieties of German. These 12 recordings were used as the auditory stimuli for the IAT. Recordings were made in a sound attenuated room with a head-mounted Sennheiser MKH 50 P48 microphone at 44 KHz. Recordings were downsampled to 22 KHz for use in the study.

Figure 3 (left plot) shows the spectral shape of the two fricatives in /NO.SIç/ (black) and /NO.SIç̥/ (blue). The shift to the higher frequencies for the non-canonical /ç̥/ pronunciation (probably due to a more fronted articulation) can be seen and mirrors the acoustic description of the different sibilants in Section “Acoustic Characteristics of Stimulus Materials.” The greater energy in the higher frequencies can be captured by CoG values, shown in the right plot of **Figure 3**. The fricatives of all word pairs are characterized by a difference in CoG, with higher values for the alveo-palatal fricative /ç̥/ as realized in the non-canonical variety.

To use the auditory stimuli in the IAT experiment, they were temporally normalized. To do so, the stimuli were segmented into three parts: stem + /i/ + /ç̥/. Each part was manipulated to have a certain length (0.34, 0.14, and 0.19 s, respectively). This was done, so that the [Iç̥] part of each stimulus word had the same duration across all stimuli and the duration of the stem was kept constant. In addition, the stimuli were normalized in amplitude (mean intensity of 70 dB) and fundamental frequency (f0). The normalization of f0 was done in a pairwise fashion

by synthesizing the non-canonical rendition of a word pair with the extracted f0 contour of the canonical stimulus of the same word pair. This was done to control for differences between *Hochdeutsch* vs. the two non-canonical conditions but at the same time keeping the stimuli as natural as possible. Mean f0 varied between the word pairs from 204 Hz for the *einzig*-pair to 214 Hz for the *bündig*-pair. All manipulations were carried out using *Praat* (Boersma and Weenink, 2018). **Figure 4** shows spectrogram and oscillogram of the *einzig*-pair with temporally adjusted segments and normalized pitch contour (above: canonical realization, below: non-canonical realization). All test items were rated for their naturalness prior to using them in the IAT experiment (see Result section “Online Rating Study”).

Stimuli for the attribute categories consisted of 12 visually presented lexical items with either a positive or a negative valency. We selected these 12 words from a range of items suggested on the German sample IAT site “Project Implicit” hosted by Harvard University¹. The words with a positive valency were: *Freude* “joy,” *Frieden* “peace,” *Lachen* “laughter,” *Liebe* “love,” *Vergnügen* “pleasure,” *wundervoll* “wonderful.” The negatively connotated words were: *böse* “evil,” *grausam* “cruel,” *Misserfolg* “failure,” *Qual* “agony,” *Übel* “evil,” *verletzt* “hurt.” These attributes were selected because of their frequent and prior use in previous IAT studies.

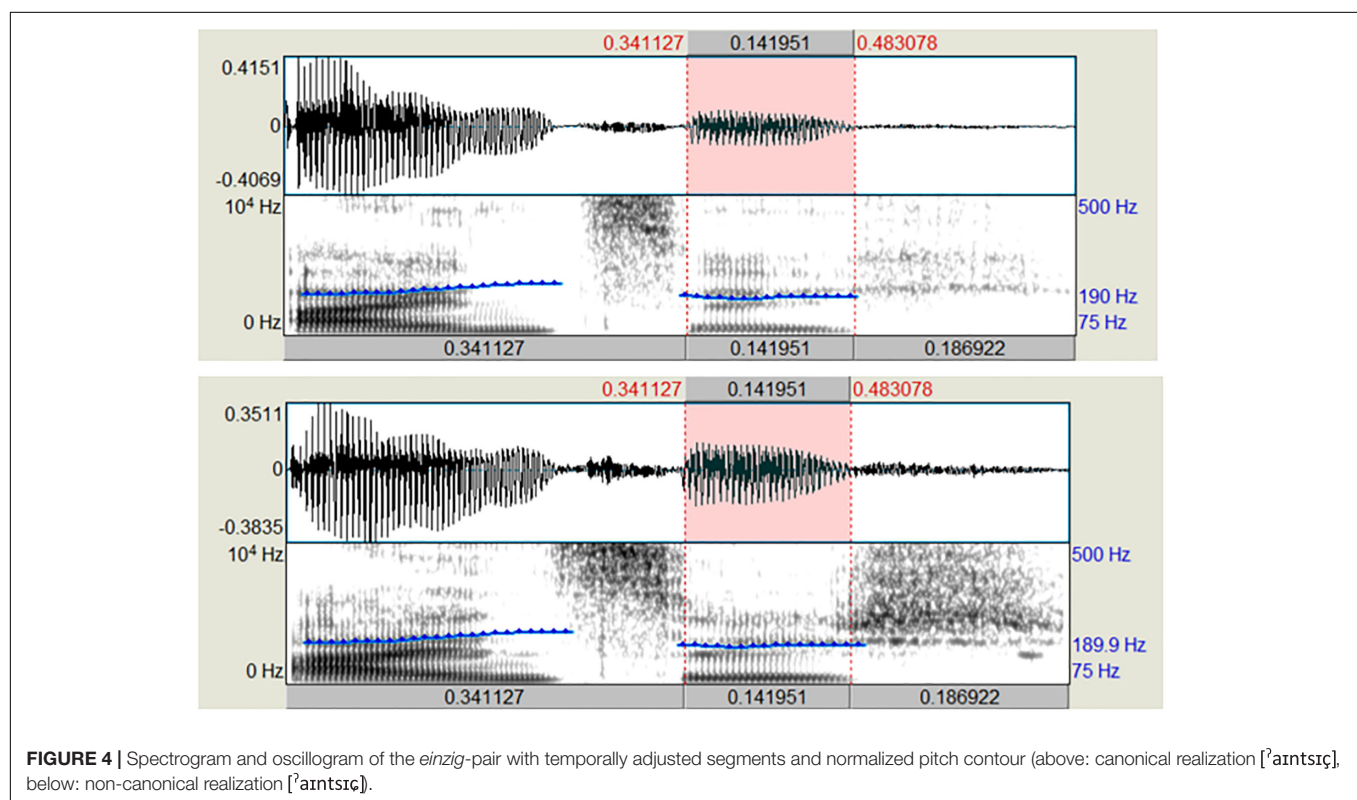
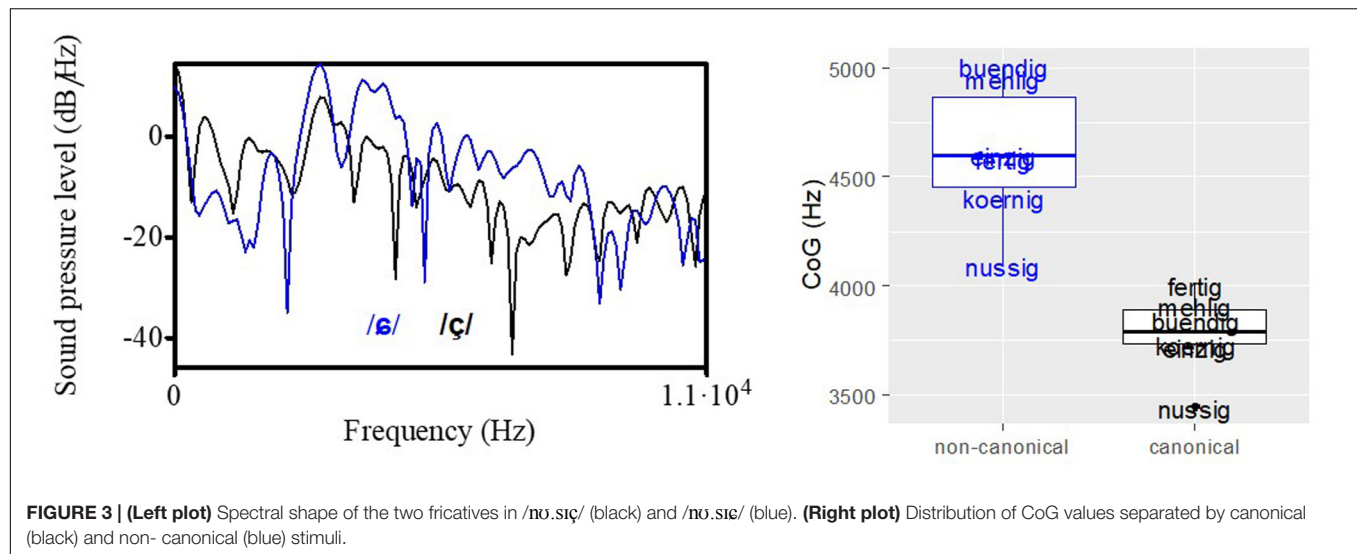
We ran two versions of this experiment: in condition 1, half of the participants saw the opposing concept categories *Hochdeutsch* and *Kiezdeutsch* while in condition 2, the second half of the participants saw the opposing categories *Hochdeutsch* and *French Accent*. Both versions of the experiment differed in the introductory text shown on the screen. In condition 1, called KD-experiment below, 68 participants (22 YMU, 20 YMO, and 26 OMO) were informed in the introduction that the auditory stimuli were recordings obtained from students at a school in the multi-ethnic district of Kreuzberg in Berlin. Accordingly, the concept categories for sorting the auditory items were labeled *Hochdeutsch* and *Kiezdeutsch*. In condition 2, called FR-experiment below, 63 participants (20 YMU, 23 YMO, and 20 OMO) read in the introduction to the experiment that they were listening to recordings of French students learning German. The concept category labels were *Hochdeutsch* and *Franz. Akzent* (*French Accent*). Both groups were told that the aim of the experiment was to sort the presented stimuli (both auditory and written) correctly into the given categories and that it was important to do this as fast as possible.

Hypotheses

Since the stimuli were the same across both versions of the experiment and for all participants, a difference in judgments reveals whether the same phonetic alternation is judged differently by each participant group depending on the information received on the origin of the auditory stimuli. As mentioned above the following hypotheses are tested:

- (1) Canonical pronunciations are associated with positive values, non-canonical pronunciations with negative values (IAT-effect).

¹<https://implicit.harvard.edu/implicit/germany/>



(2) The socially situated context (French Lerner German vs. *Kiezdeutsch*) biases listeners to interpret the identical acoustic stimuli differentially depending on the listeners' age and ethnic background.

(a) Mono-ethnic German listeners show a stronger IAT-effect in the KD experiment indicating a stronger negative bias toward this variety than in the FR experiment, while multi-ethnic listeners show a reversed pattern with a stronger negative

bias toward the supposed French variety than toward their own speech group.

(b) Younger mono-ethnic German listeners show a smaller IAT-effect than older mono-ethnic German listeners mirroring their smaller bias toward non-canonical pronunciations.

Statistical Analysis

Statistical analyses were conducted in R (R Core Team, 2016). We ran one-sample t-tests and linear (mixed) models using the

packages *lme4* (Bates et al., 2015) and *lsmeans* (Lenth, 2016). Significance testing was done by model comparison (with and without the factor or interaction in question). For the online rating test, the fixed factors included were *word*, *pronunciation variety* and *speaker*, while *listener* was added as random effect. For the IAT analysis, just one *D*-score per participant (across words) constitutes the dependent variable, thus, the calculated linear models include *listener group* and *test order* (version) as predictors, and additionally *experiment condition* in the combined data set.

RESULTS

Online Rating Study

Figure 5 (right panel) shows the results of the online rating study with 40 participants, that tested for the perceived naturalness of the auditory IAT stimuli (named *IAT_stim*) in comparison to non-manipulated items produced by two different female speakers (*sp1*, 39 years and *sp2*, 53 years). In addition to the perceived naturalness, we tested the perceived age and education of the speakers, shown in the left and middle panel of the figure. Ratings are separated by pronunciation variation (blue: canonical = /ç/, red: non-canonical = /ʤ/).

Most importantly, the manipulation did *not* show an effect on perceived naturalness: the manipulated stimuli used in the IAT did not differ significantly from the other stimuli. However, the LMM showed a main effect of word [$\chi^2(5) = 18.42, p < 0.01$], with *bündig* and *einzig* being perceived as less natural. A possible explanation is that these words are less frequent in their use than the other adjectives, especially when used out of context. More interestingly, we found an interaction of speaker and pronunciation variation [$\chi^2(2) = 12.01, p < 0.01$]: ratings were significantly less natural for the non-canonical pronunciations than for the canonical pronunciations for two out of the three

speakers (*sp2* and the *IAT_stim*, $p < 0.001$), but not for *sp1*. This might reflect the fact that *sp1* is a native speaker of a central German dialect who produces the versions of the <-ig> with a less salient perceptual contrast, whereas the other two speakers differentiated more clearly between a canonical and a non-canonical pronunciation.

For age, we also found an interaction of speaker and pronunciation variation [$\chi^2(3) = 43.32, p < 0.001$]: *Sp1* again did not show a difference in perceived age between the pronunciations. While *sp2* sounded *older* in canonical than non-canonical, the *IAT speaker* sounded *younger* in canonical than non-canonical. This opposing effect of pronunciation on perceived age in the two speakers is striking at first but might be due to the difference in biological age between the two speakers. For a young woman in her 20s a non-canonical pronunciation increases the perceived age, while it decreases the perceived age in a woman in her early 50s. For education, the interaction of word, pronunciation variation and speaker turned out significant [$\chi^2(22) = 92.09, p < 0.001$]. While there was variation in terms of differences between speakers (in some words and a certain pronunciation), for all speakers and all words the non-canonical variation was perceived as less educated than the canonical pronunciation (cf. **Figure 5**).

Error Rates and Reaction Times of the IAT Studies

Error Rates

As a first step to the analysis of the data obtained in the IAT experiment, we performed an error analysis to check whether participants were able to discriminate between the two pronunciation variants above chance level and to compare the correctness scores to the ones for the written stimuli. Overall, correctness scores were high, but as expected, they were higher in the written stimuli than in the audio stimuli. In the practice

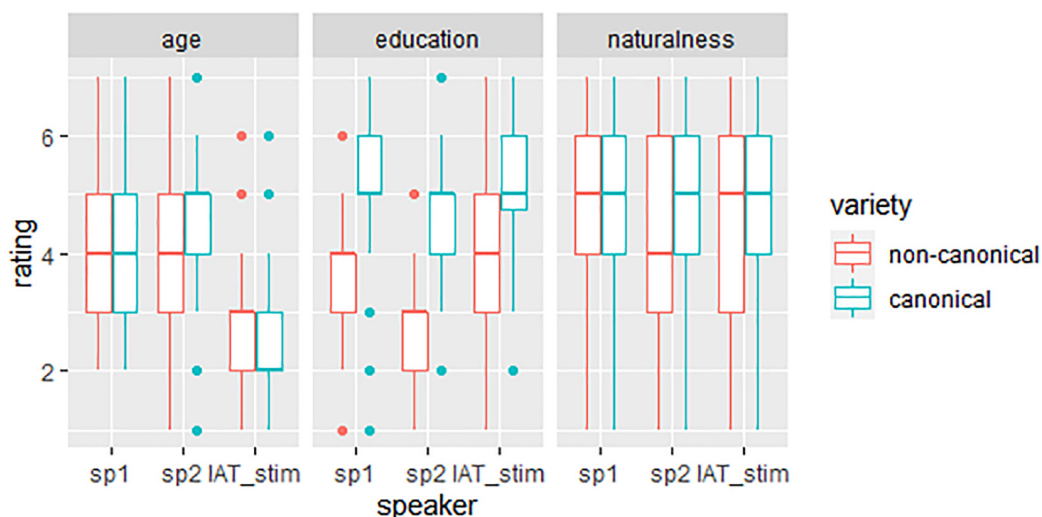


FIGURE 5 | Distribution of ratings regarding perceived age, education and naturalness separated by speaker (*sp1*, *sp2*, and *IAT_stim*) and variety (blue = /ç/, red = /ʤ/).

trials (Block 1: audio only, Block 2 written only), participants identified the audio stimuli in 84% correctly, the written stimuli in 98%. **Table 3** shows the correctness scores for the identification tasks separated by participant group and stimulus type (audio vs. written) with all blocks included. Again, the numbers show that the written stimuli were more easily correctly identified by the participants than the auditory stimuli. There were no obvious differences between the context conditions (FR vs. KD) in correctness scores, however, there is a slight tendency for the audio stimuli to be identified more reliably in the KD condition compared to the FR condition. However, this is only true for the monolingual groups YMO and OMO, but not for the YMU listeners, who do not differ between conditions and overall displayed the poorest performance in identifying the audio stimuli.

Reaction Times

Here, a short description of participants' absolute reaction times (RT) is given. Without practice trials and independent of stimulus type (audio/written) and experiment (KD/FR), RT was on average 1,144 ms (measured from the time when stimuli were displayed on screen or played) and ranged between a lower and upper quartile of 764 and 1,299 ms. **Table 4** shows the reaction times separated by experiment, stimulus type and shared key conditions. Overall, participants in the KD experiment were slightly slower than in the FR experiment across all subgroups. Note though, that these overall differences between experiment conditions do not affect the IAT effect (*D*-score), since this measure is calculated for each participant separately and is a comparative measure which takes the relation of the RTs of the different blocks into account. Also, in general, written stimuli were categorized faster than audio stimuli. However, it

should be kept in mind that the audio stimuli had a length of 670 ms and differed only in the final sound, i.e., the canonical or non-canonical fricative, and thus reducing the time participants took to decide. Most interestingly, it took participants less time to sort stimuli into categories when these categories were placed congruent to our hypotheses ("canonical form" and "good" sharing the same key) across both stimulus types and experiments.

IAT Condition 1: KD-Experiment

Figure 6 shows the distribution of *D*-scores separated by listener group (YMU, YMO, and OMO) and order of presentation (v1 and v2). As mentioned above, in v1 participants did the congruent pairing first (canonical – positive, non-canonical – negative), in v2, participants started with the incongruent pairing (non-canonical – positive, canonical – negative). Remember that a *D*-score near zero indicates that there is no effect of the experiment. A positive *D*-score, however, means analogous to our hypothesis that the non-canonical /ʒ/-pronunciation was associated with negative adjectives and the canonical /ç/-pronunciation was associated with positive adjectives. The figure shows a clear difference between the test orders (v1 vs. v2), it matters which pairing was seen and learned first. Test version v1 seems to generally result in a stronger IAT effect (higher and more positive *D*-scores) than version v2. Differences are also apparent between the listener groups. The oldest group OMO reveals the highest *D*-scores (most strongly associating the non-canonical pronunciation with negative valence words) and group YMU the lowest, while group YMO lies in between the two.

One-sample *t*-tests were carried out across v1 and v2 for each listener group separately to see whether the *D*-scores deviate from zero thereby indicating a positive IAT effect. Significant effects were found for all listener groups corroborating hypothesis 1 [YMU: $t(19) = 4.01$, p -value < 0.001, YMO: $t(19) = 3.935$, p -value < 0.001, OMO: $t = 11.261$ (21), p -value < 0.0001]. To investigate more closely the size of the IAT effect depending on the order of presentation and the listener group, a linear model was calculated with the *D*-score as dependent variable and the factors listener group (YMU, YMO, and OMO) and order of presentation of the test items (v1 and v2) as the predictors. **Table 5** shows a summary of the results. As indicated already in **Figure 6**, a significant ordering effect was found independently of listener group (even though the effect seems to be strongest for the YMO group – green in **Figure 6** – the interaction was not significant): the IAT effect was larger, meaning there was a stronger association between the non-canonical pronunciation /ʒ/ and the negative adjectives in version 1 (v1) where experiment participants first practiced the association of the /ʒ/ pronunciation with the negative adjectives.

More interestingly, a significant effect of listener group was found with the older listeners (group OMO) differing from both the younger multi-ethnic (YMU) and younger mono-ethnic (YMO) listeners, while the difference between YMO and YMU does not differ significantly. Thus, the IAT effect regarding the association between negative adjectives and the non-canonical /ʒ/-pronunciation was larger for the older listener group compared to the younger groups corroborating hypothesis 2.

TABLE 3 | Average correctness scores (in %) of audio and written stimuli calculated over the whole experiment separated by participant group and context condition (French Accent vs. Kiezdeutsch).

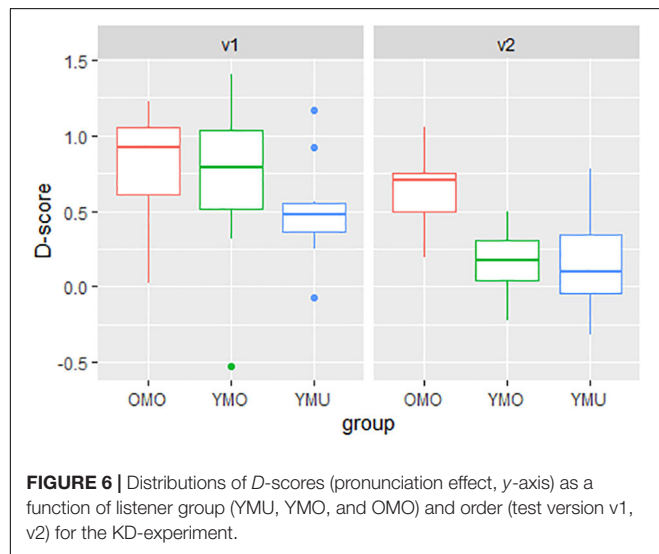
Participant group	FR (audio/written)	KD (audio/written)
YMU	80.2/92.1	79.8/95.4
YMO	88.4/95.9	93.1/95.8
OMO	86.7/96.1	91.9/96.3

TABLE 4 | Participants' mean reaction times (RT) and standard deviations (SD) separated by experiment condition (KD and FR), shared keys (congruent and incongruent to our hypothesis) and stimulus type (audio and written).

Experiment	Shared keys	Mean RT in ms (SD) audio stimuli/written stimuli
KD	Non-canonical and good	1475.9 (820.9)/1180.6 (758.0)
	Canonical and good	1196.1 (710.4)/913.0 (562.2)
FR	Non-canonical and good	1366.3 (677.4)/1063.9 (688.0)
	Canonical and good	1124.7 (452.6)/833.6 (505.7)

TABLE 5 | Summary statistics of the linear model with *D*-score as dependent variable and the influencing factors *listener group* and *test order* for the KD-experiment.

	Estimate	Std. error	t-value	Pr(> t)
(Intercept)	0.25337	0.09059	2.797	0.006991**
Group YMO vs. YMU	-0.09315	0.11140	-0.836	0.406512
Group YMO vs. OMO	0.30858	0.10884	2.835	0.006295**
Group YMU vs. OMO	0.40173	0.1088	3.691	0.000495***
Order v2 vs. v1	0.34662	0.08948	3.874	0.000275***

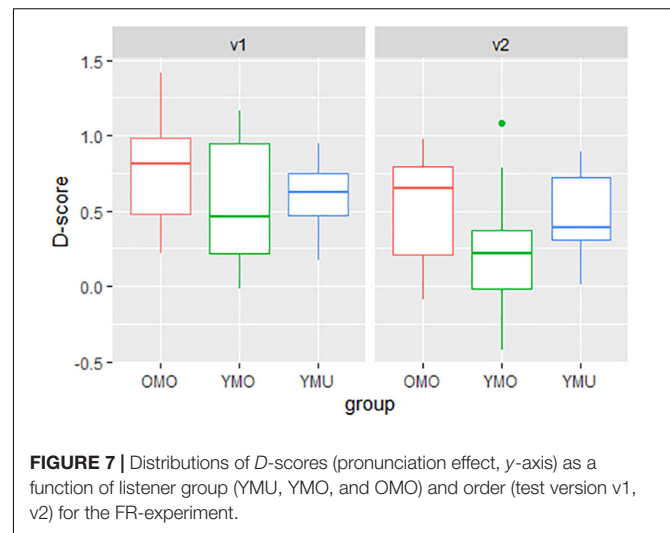
** $p < 0.01$, *** $p < 0.001$.**FIGURE 6** | Distributions of *D*-scores (pronunciation effect, y-axis) as a function of listener group (YMU, YMO, and OMO) and order (test version v1, v2) for the KD-experiment.

IAT Condition 2: FR-Experiment

Parallel to the analysis of the KD-experiment one sample t-tests were made for each listener group to see whether *D*-scores differ significantly from zero and thus indicate a positive IAT-effect. As in the KD-experiment, significant IAT-effects were found for all listener groups [YMU: $t(19) = 8.5608$, p -value < 0.0001 , YMO: $t(22) = 4.1255$, p -value < 0.0001 , OMO: $t(19) = 6.9851$, p -value < 0.0001].

Figure 7 shows the distribution of *D*-scores separated by listener group and order of presentation. Similar to the KD-experiment, variation between listener groups and versions appear to be pointing to listener- and order-specific differences in the size of the IAT-effect. However, in comparison to the KD-experiment, the intra-group variability is much greater when experiment participants believed to be listening to French learners of German. This is reflected by the larger box sizes comprising 50% of the data above and below the bold line (median) in each bar. Also, here, group YMO shows the lowest mean *D*-score whereas in the KD-experiment, YMU displayed the lowest score.

To test for the significant differences between listener groups and test orders, here too, a linear model was calculated. Again, significant main effects of the two factors *group* and *order* were found but no interaction thereof. Table 6 shows the summary statistics of the model with a stronger IAT effect in version 1 than

**FIGURE 7** | Distributions of *D*-scores (pronunciation effect, y-axis) as a function of listener group (YMU, YMO, and OMO) and order (test version v1, v2) for the FR-experiment.**TABLE 6** | Summary statistics of the linear model with *D*-score as dependent variable and the influencing factors *listener group* and *test order* for the FR-experiment.

	Estimate	Std. error	t-value	Pr(> t)
(Intercept)	0.26598	0.08788	3.027	0.00366**
Group YMO vs. YMU	0.14271	0.11157	1.279	0.20588
Group YMO vs. OMO	0.23088	0.11157	2.069	0.04290*
Group YMU vs. OMO	0.08818	0.11538	0.764	0.4478
Version 2 vs. 1	0.24347	0.09197	2.647	0.01039*

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

in version 2 and a stronger IAT effect in group OMO compared to YMO. While the *D*-scores of group YMU lie in between the other two groups, the differences fail to reach significance which we assume is also affected by the large intra-group variation that can be seen in Figure 7.

Comparing the Experiments KD and FR

We will now take a closer look at the similarities and differences between the two test conditions (KD and FR) by combining the data sets. Figure 8 gives a first impression, showing the *D*-scores for both order versions separated by listener group (YMU, YMO, and OMO) and experiment (FR and KD). Values for the younger mono-ethnic hearers YMO (green in Figure 8) are in between the two other groups, and more importantly, there is no obvious difference between the two conditions FR and KD. For the older listeners (red) and the younger multilingual listeners (blue), the effects go in different directions: while for group YMU the FR experiment reveals higher *D*-scores and thus a stronger IAT effect, for group OMO the KD experiment reveals slightly higher *D*-scores. The YMO group seems widely unaffected by the different priming conditions, the *D*-scores are above zero but rather low (with a large spread indicative of the variance in the responses) for both the KD and FR condition.

A linear model was calculated over the combined data with *D*-score as dependent variable and the factors order (v1 and v2), listener group (YMU, YMO, and OMO) and condition (KD

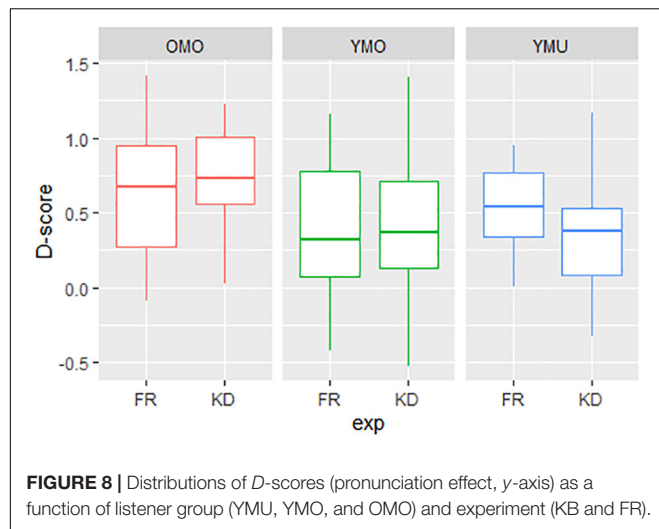


TABLE 7 | Summary statistics of the linear model with *D*-score as dependent variable and the influencing factors *listener group*, *test order* and *experiment condition*.

	Estimate	Std. error	t-value	Pr(> t)
(Intercept)	0.38310	0.08625	4.441	2.03e-05***
Condition KD vs. FR (for YMU)	-0.1968	0.11325	-1.738	0.0847
Group OMO vs. YMO	-0.1415	0.10951	-1.293	0.1985
Group OMO vs. YMU	0.08818	0.11325	0.779	0.4378
Version 2 vs. 1	0.29465	0.06408	4.59	1.08e-05***
Condition KD * Group YMO	0.23474	0.15754	1.490	0.1389
Condition KD * Group OMO	0.31356	0.15834	1.980	0.0500*

* $p < 0.05$, *** $p < 0.001$.

and FR) as potential influencing variables. **Table 7** shows the summary statistics of the model. This time, in addition to the main effect of test order, a significant interaction between listener group and condition was found as suggested by hypothesis 2. As already indicated in **Figure 8**, the difference in the IAT-effect between the experiment conditions varies between OMO and YMU: While for the YMU-listeners there was a larger IAT-effect in the FR condition, reflected by the median *D*-score in the Figure and a negative estimate in **Table 7** (line 2), for OMO-listeners, the larger IAT-effect was found in the KD-condition reflected in the significant interaction and a positive estimate in **Table 7** (last line).

This reveals that for older listeners the association between the /ç/ pronunciation and negative valency words was stronger when linked to the *Kiezdeutsch* variety, for the YMU group the association between the non-canonical variant and negative valency words was stronger when linked to a French learner variety than to *Kiezdeutsch* – the variety that many of the listeners themselves speak and are habituated to. YMO listeners lie between the other groups with no obvious difference in bias between the two cultural priming contexts.

SUMMARY AND DISCUSSION

Our results indicate that the experimental paradigm was successfully deployed to show that the implicit attitudes of the three different hearer groups not only differ but also, that the two different contexts elicited differences in implicit associations. There is a priming effect of the variant fricative forms that carry social meanings and trigger implicit attitudes. Based on the speed of reaction to the stimuli, we argue that the patterns observed here are unmediated and indicative of implicit attitudes. The method that we have chosen in our study has shown repeatedly that individuals process information implicitly (i.e., automatically or unconsciously) (Greenwald and Banaji, 1995; Bargh and Chartrand, 1999; Banaji, 2001; Greenwald et al., 2002, 2009; Fatfouta et al., 2014) as opposed to explicitly (controlled or conscious). To be more precise, we argue that our results show attitudes below the level of consciousness due to three different aspects related to the IAT method.

First, the reaction times of the responses range between a lower and upper quartile of 764 and 1,299 ms, which points to a rather rapid overall response. As a reference, in neural language processing EEG (electroencephalography) studies, a negative deflection in the ERP (event related potentials) signal at around 400–550 ms (N400) after stimulus onset indicates a detection of semantic anomalies (see for example Van Berkum et al., 2008). Given that these are very immediate and pre-motor brain responses that do not require any decision making (left or right button) or activation of motor patterns (such as lifting a finger and pressing a button), the average RT in our IAT-study seems relatively fast.

Second, participants are generally not aware of what is being measured in an IAT experiment. They might of course notice having more trouble when *good* and *Kiezdeutsch* are mapped to the same response key. However, participants might not attribute this directly to a bias they have. On the contrary, participants are often negatively surprised by their results showing a bias against a specific group of people as they are generally not aware of it (see Banaji and Greenwald, 2013). So far, it is unclear if implicit biases are based on, e.g., personal experiences, learned mainstream attitudes based on frequent confrontation with stereotypes, or internalized stereotypes against out-groups or even the own in-group.

Third, even if participants are aware of the scoring algorithm, they are unlikely able to consciously alter and adjust their behavior with the purpose of influencing their final score. For example, in case of trying to hide a bias against the *Kiezdeutsch* variant, one would have to deliberately take longer in the blocks congruent with our hypothesis (*bad* – *Kiezdeutsch* and *good* – *Hochdeutsch*) while also trying to be faster in blocks incongruent to our assumption (*good* – *Kiezdeutsch* and *bad* – *Hochdeutsch*). Moreover, this strategy would have had to have been maintained throughout the entire experiment which would have added to the cognitive load and would inevitably have led to an overall increase in RT. And as a last point, there is a systematic pattern of variation within but not across the three listener groups, showing that all of them have different implicit associations with the stimuli presented.

That is, we see that not all groups behave alike. Our first hypothesis that non-canonical pronunciations are more strongly associated with negatively valence words and canonical pronunciations with positive valence words is borne out. This finding is also corroborated by the results of the online rating experiment where stimuli from all three speakers and all words are rated less educated in the non-canonical pronunciation variant than in the canonical pronunciation. In addition, the distribution of absolute reaction times mirrors the results of the IAT experiments: participants were faster in sorting the audio and written stimuli into the corresponding categories when these categories were placed according to our hypothesis (canonical pronunciations and positive valence words, and non-canonical pronunciations and negative valence words sharing the same response keys).

Moreover, in accordance with our second hypothesis, we could show that the IAT effect differs between listener groups: it is stronger for the older age group compared to the younger groups regardless of their language and ethnic background. In addition, for the older listener group we find a greater effect (a stronger association of non-canonical pronunciation and negative valence words) when hearers believe to be listening to a speaker from Kreuzberg and a lesser effect (indicating a less negative attitude toward this speaker group and speech variant) when hearers believe to be listening to a learner of German from France. While the younger mono-ethnic German group seems indifferent toward the priming conditions (FR vs. KD), the younger multi-ethnic group links the variant pronunciation [ç] more strongly with negative valence words in the French learner condition compared to the *Kiezdeutsch* condition, showing a preference for their in-group.

It therefore seems that non-canonical speech forms must not necessarily evoke negative associations and are highly dependent on the interpreter (covert prestige). This is a complex and evolving process especially considering the ongoing diversification of the urban Berlin population but also many other urban spaces in Europe [e.g., Multicultural London English, Kerswill et al., 2008; Straattaal (Netherlands), Nortier, 2001; Rinkeby-Svenska (Sweden), Kotsinas, 1998; Kobenhavnsk Multietnolekt (Denmark), Quist, 2005; Multiethnolektales Schweizerdeutsch (Switzerland), Schmid, 2020; Kebab Norsk (Norway), Svendsen and Røynealand, 2008] where there are many antagonistic but also collective forces that build a microcosmos of their own and where world knowledge may be shared but differently evaluated, categorized or interpreted. The concepts of ethnolectal group membership (in-groups vs. out-groups, cf. Tajfel and Turner, 1986) are categories that are somewhat augmented by social affiliations with aspects of mainstream and non-mainstream culture.

Berlin prides itself with being an open-minded, diverse, friendly and multi-cultural European city with a truly international flair due to the ethnic diversity of its inhabitants and the many tourists. Especially younger people from Berlin embrace this urban feel and the flair of the hip and diverse neighborhoods. As such, there is some cultural capital (Bourdieu, 1986; Jannedy et al., 2019) associated with having international affiliations, being of multi-ethnic decent and well versed in

street-culture. It is not that the speech features described for *Kiezdeutsch* are intrinsically hip or cool – in fact, there is some evidence that the mainstream is not fond of the linguistic variation – it is the hipness of the concept of being part of the underdog, bad-boy, *street*, and youth-culture, of being shunned upon by more conservative forces and by integrating with those who in the past were not well integrated by embracing aspects of their culture, food, style, and speech. In other words, speech features that were associated with one specific social group (i.e., multi-ethnic adolescents) and that are stigmatized especially by conservative forces, were used by other parts of the younger urban population through *crossing* (Dirim and Auer, 2004, pp. 204–224; Rampton, 2014), have gained covert prestige, and were adopted as their own, indexing social orientation toward multi-ethnicity, diversity, and urbanity.

In light of this, it is feasible that younger listeners in general are more open to variation in fine phonetic detail as they are in a better position to contextualize phonetic innovations and accept these as potentially meaningful expressions of identity while older populations are more strongly attached to a fictitious standard. Our work on the /ç/ – /ʃ/ merger in Berlin (Jannedy and Weirich, 2014) corroborates these assumptions as identification patterns of older listeners showed more [ç] ratings when they believed that the speaker was from a multi-ethnic district (i.e., *Kreuzberg*) compared to a mono-ethnic German district of Berlin, while younger listeners were not receptive to the priming. Especially for the older listeners, the data strongly indicates that there is a lack of social status and prestige associated with the pronunciation of /ç/ as [ç] when attributed to a speaker group from *Kreuzberg*. The IAT results also corroborate the finding that younger mono-ethnic listeners seem to have less of a strong bias toward one variant over another with *D*-scores only slightly above zero in both conditions. We suspect this being due to hearing both versions in the ambient environment and maybe even variably producing it in contexts that situationally or functionally demand not using a canonical version of this fricative.

The interesting effect of age reflected in the results indicates that the oldest group of listeners (OMO) had the strongest associations of the non-canonical variant [ç] with negative valence words in both conditions, with a slightly stronger tendency in the *Kiezdeutsch* condition. Thus, a phonetic variant stemming from a French learner variety of German did (even though only to some extent) evoke more positive associations than the multi-ethnic variant associated with *Kiezdeutsch*. The results for the oldest listener group and the younger multi-ethnic listener group are diametrically opposed: listeners in the YMU group had stronger negative associations with the French variety compared to the *Kiezdeutsch* variety. Not only is this evidence that the associative responses are learned but also that in the case of the YMU group, the responses toward the in-group variety *Kiezdeutsch* that many of the listeners themselves speak and are habituated to were more positive (cf. Tajfel and Turner, 1986). In addition, the analysis of the error rates revealed that this group had the most problems in differentiating the two pronunciation variants, probably reflecting their own productions of the merged variant typical

for this speech community and also, a lesser awareness of the distinction in general.

For each experimental condition (KD/FR), we have tested a different group of hearers to prevent having to draw attention to the contrast between these conditions which would have biased hearers in an uncontrollable manner. To prevent this from happening and of course also due to time constraints and participants fatigue we assigned participants randomly to two groups differing in the experiment conditions. However, in order to investigate the effect of priming condition (French accent vs. *Kiezdeutsch*), we collapsed the data for these two different conditions. The important point is that we found an interaction between *listener group* and *condition* (for some groups the prime *Kiezdeutsch* evokes stronger IAT effects, for others the prime *French Accent* did so). We therefore assume that these results are due to different associations drawn by the different listener groups and not due to different participants taking part in the two experimental conditions (because then we would have expected a main effect of experiment, with an overall difference between conditions irrespective of listener group).

In addition to the hypothesized effects of priming condition and listener group, our results also show an effect of which associations listeners rated first. IAT effects were generally larger when the version with the hypothesis-congruent relations were shown first (canonical pronunciation and positive value sharing a response key) followed by the hypothesis-incongruent relations (canonical pronunciation and negative value). These order effects (i.e., associations appear stronger when they are tested in blocks 3 and 4 rather than in blocks 6 and 7) have been described earlier and a suggested improvement of IAT experiments was to increase the number of trials in block 5 to counteract these order effects (cf. Greenwald et al., 2003). In our study however, order effects still appeared despite this change in method. Greenwald et al. (2003, p. 209) suggest that order effects in IATs might be related to a phenomenon called *negative transfer* (Woodworth et al., 1954), “whereby practice at one task interferes with performance at a second task that requires giving different responses to the first task’s stimuli.” This negative transfer is assumed to result in a strengthening of associations between the hypothesis congruent categories (canonical-good) when the task that uses this association (that uses the same response keys to canonical pronunciations and positive words) is tested initially.

With respect to our study it seems that the IAT effect is somewhat leveled when listeners first “learn” the incongruent association and must redo this learning with a new (but) better fitting association (in terms of their implicit bias). When the listeners first “learn” the congruent association according to their implicit bias, the bias is strengthened and the redoing of the learned (and fitting to their stereotypical) association is even more difficult resulting in a stronger IAT effect. However, this was the case in both priming conditions (*French accent* and *Kiezdeutsch*) and no interaction of order (version) and priming was found. Also, there was no significant difference between the listener groups in terms of the order effect, even though a tendency was apparent for the YMO group in the KD experiment to show a larger difference between the versions

than the other listener groups (mainly due to a very small *D*-score and thus a low IAT effect in version 2). Thus, the YMO group shows the strongest effect of test order which might point to a greater flexibility in their associations between pronunciation variants (canonical/non-canonical) and positive or negative connotations. Their bias toward an association between canonical and positive is small and thus mostly affected by the re-learning of an assumed incongruent association such as non-canonical and positive and the process of “negative transfer.”

Nevertheless, we are aware of some limitations of our study. First, the group of hearers was not as homogenous as would have been ideal in the sense that there were differences between participants over which we had no control. Further research might highlight additional factors interacting with differences in IAT effects between individual listeners or listener groups. For example, it would have been interesting to also assess the listeners’ *explicit* attitudes toward French learners of German and of adolescent speakers from Kreuzberg with whom the tested variant is highly associated. Also, incorporating personality constructs such as openness – one of the dimensions of the Five Factor Model describing differences in personality (McCrae and John, 1992) – or the proximity of a listener to ideologies such as *conservatism* and *liberalism* (Kerlinger, 1984) – which reflect a person’s attitudes toward changes (personal or political) – might give insights into the reasons for differences in the sensitivity to priming effects and IAT effects and toward linguistic change in general.

Second, while we did look for an effect of lexical frequency of the test items on the absolute RTs, our material was not designed to investigate the effect of lexical frequency in a controlled manner. With our limited selection of stimuli, we did not find systematic variation. Nevertheless, it would be interesting to explore more the effect of lexical frequency of the different test items on the IAT-effect in the sense that more frequent words facilitate, and less frequent words inhibit the implicit association between concept categories and valence categories.

While we believe that there is a general lack of awareness that one phonetic variant belongs to a specific social group, our results strongly suggest that implicit associations are drawn between fine phonetic detail and social groups. Moreover, these associations are affected by listeners’ background, i.e., their attitudes, beliefs, stereotypes, and shared world knowledge pointing to language processing which needs to incorporate culturally and socially situated contexts.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the DGfS

(#2019-07-190625). Written informed consent was obtained from all participants. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

The idea and conceptualization of the investigation, data analysis, and writing of the manuscript was joint work by SJ and MW. The execution of the experiment was done by GS and SJ. The data curation was carried out by GS. Visualization was conducted by MW. Funding acquisition and project administration was done by SJ. All authors contributed to the article and approved the submitted version.

REFERENCES

- Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J. Phonet.* 40, 177–189. doi:10.1016/j.wocn.2011.09.001
- Banaji, M. R. (2001). "Implicit attitudes can be measured," in *The Nature of Remembering: Essays in Honor of Robert G. Crowder* H. L. Roediger, J. S. Nairne, I. Neath, and A. Surprenant eds (Washington, DC: American Psychological Association). 117–150. doi: 10.1037/10394-007
- Banaji, M. R., and Greenwald, A. G. (2013). *Blindspot: Hidden Biases of Good People*. New York, NY: Delacorte Press.
- Bargh, J. A., and Chartrand, T. L. (1999). The unbearable automaticity of being. *Am. Psychol.* 54, 462–479. doi: 10.1037/0003-066X.54.7.462
- Barlow, M., and Kemmer, S. (eds.) (2000). *Usage-Based Models of Language*. Stanford, CA: Center for the Study of Language and Information.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., et al. (2015). Package 'lme4'. *Convergence* 12:2.
- Bell, A. (1984). Language style as audience design. *Lang. Soc.* 13, 145–204. doi: 10.1017/s004740450001037x
- Boersma, P., and Weenink, D. (2018). *Praat: Doing Phonetics by Computer [Computer program]*. Version 6.0.46. Available online at: <http://www.praat.org/> (available at: 3 January 2018)
- Bourdieu, P. (1986). "The forms of capital," in *Handbook of Theory of Research for the Sociology of Education* J. Richardson ed (Westport, CT: Greenwood). 46–58.
- Bourhis, R. Y., and Giles, H. (1977). "The language of intergroup distinctiveness," in *Language, Ethnicity, and Intergroup Relations* H. Giles ed. (London: Academic Press). 119–136
- Bukmaier, V., and Harrington, J. (2016). The articulatory and acoustic characteristics of Polish sibilants and their consequences for diachronic change. *J. Int. Phonet. Assoc.* 46, 311–329. doi: 10.1017/s0025100316000062
- Burnham, D., Kitamura, C., and Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science* 296, 1435–1435. doi: 10.1126/science.1069587
- Campbell-Kibler, K. (2008). I'll be the judge of that: diversity in social perceptions of (ING). *Lang. Soc.* 37, 637–659. doi: 10.1017/S0047404508080974
- Campbell-Kibler, K. (2010). The effect of speaker information on attitudes towards (ING). *J. Lang. Soc. Psychol.* 29, 214–223. doi: 10.1177/0261927x09359527
- Campbell-Kibler, K. (2011). The sociolinguistic variant as a carrier of social meaning. *Lang. Variat. Change* 22, 423–441. doi: 10.1017/s0954394510000177
- Campbell-Kibler, K. (2012). The implicit association test and sociolinguistic meaning. *Lingua* 122, 753–763. doi: 10.1016/j.lingua.2012.01.002
- Campbell-Kibler, K. (2013). "Connecting attitudes and language behavior via implicit sociolinguistic cognition," in *Experimental Studies of Changing Language Standards in Contemporary Europe* T. Kristiansen and S. Grondelaers eds (Oslo: Novus Press) 307–329.

FUNDING

This work was carried out with funding by the German Research Council (DFG) project number 416591334 – CRC 1412. The publication of this article was funded by the *Open Access Fund* of the Leibniz Association.

ACKNOWLEDGMENTS

We are grateful to Christoph Draxler (LMU) for the help we received with conducting the stimulus ratings on *Percy* and to Miriam Oschkinat (LMU) for patiently reading in stimulus utterances. We are very thankful to Therese Leinonen who helped with the planning and implementation of the IAT experiments in *PsychoPy*. We thank our reviewers for their thoughtful and insightful comments. All errors remain our own.

- Cheon, S. Y., and Anderson, V. B. (2008). Acoustic and perceptual similarities between English and Korean sibilants: implications for second language acquisition. *Korean Ling.* 14, 41–64. doi: 10.1075/kl.14.03syc
- Czaplicki, B., Zygis, M., Pape, D., and Jesus, L. M. (2016). Acoustic evidence of new sibilants in the pronunciation of young Polish women. *Poznan Stud. Contemp. Ling.* 52, 1–42. doi: 10.1515/psicl-2016-0004
- Dirim, Y., and Auer, P. (2004). *Türkisch Sprechen Nicht Nur Die Türken. Über Die Unschärfebeziehung Zwischen Sprache und Ethnie in Deutschland. (Turkish is not Only Spoken by Turks. On the fuzziness Between Language and Ethnicity in Germany)*. Berlin: De Gruyter.
- Draxler, C. (2011). *Percy-An HTML5 Framework for Media Rich Web Experiments on Mobile Devices*. Interspeech, Florence. 3339–3340.
- Draxler, C. (2014). "Online experiments with the Percy software framework-experiences and some early results," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)* Reykjavik. 235–240.
- Eckert, P. (1989). *Jocks and Burnouts. Social Categories and Identity in the High School*. New York, NY: Teachers College Press.
- Eckert, P. (2008). Variation and the indexical field. *J. Socioling.* 12, 453–476. doi: 10.1111/j.1467-9841.2008.00374.x
- Eckert, P. (2012). Three waves of variation. *Annu. Rev. Anthropol.* 41, 87–100. doi: 10.1146/annurev-anthro-092611-145828
- Eckert, P. (2019). The individual in the semiotic landscape. *Glossa* 4, 1–15. doi: 10.5334/gjgl.640
- Eckert, P., and Labov, W. (2017). Phonetics, phonology & social meaning. *J. Socioling.* 21, 467–496.
- Evers, V., Reetz, H., and Lahiri, A. (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *J. Phonet.* 26, 345–370. doi: 10.1006/jpho.1998.0079
- Fatfouta, R., Schröder-Abe, M., and Merkl, A. (2014). Forgiving, fast and slow: validity of the implicit association test for predicting differential response latencies in a transgression-recall paradigm. *Front. Psychol.* 5:728. doi: 10.3389/fpsyg.2014.00728
- Foulkes, P., and Docherty, G. (2006). The social life of phonetics and phonology. *J. Phonet.* 34, 409–438. doi: 10.1016/j.wocn.2005.08.002
- Giles, H., and Niedzielski, N. (1998). "Italian is beautiful, German is ugly," in *Language Myths* L. Bauer and P. Trudgill eds (Harmondsworth: Penguin Books). 85–93
- Giles, H. (1973). Accent mobility: a model and some data. *Anthropol. Ling.* 15, 87–105.
- Giles, H., Bourhis, R., and Davies, A. (1975). *Prestige Styles: The Imposed Norm and Inherent Value Hypothesis*. Walter de Gruyter: Berlin
- Giles, H., Bourhis, R., Lewis, A., and Trudgill, P. (1974). The imposed norm hypothesis: a validation. *Q. J. Speech* 60, 405–410. doi: 10.1080/00335637409383249

- Giles, H., Taylor, D. M., and Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: some Canadian data. *Lang. Soc.* 2, 177–192. doi: 10.1017/s0047404500000701
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105:251. doi: 10.1037/0033-295x.105.2.251
- Gordon, M., Barthmaier, P., and Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *J. Int. Phonet. Assoc.* 32, 141–174. doi: 10.1017/s0025100302001020
- Greenwald, A. G., and Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* 102, 4–27. doi: 10.1037/0033-295x.102.1.4
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., and Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychol. Rev.* 109, 3–25. doi: 10.1037/0033-295x.109.1.3
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480.
- Greenwald, A. G., Nosek, B. A., and Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* 85, 197–216. doi: 10.1037/0022-3514.85.2.197
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* 97, 17–41. doi: 10.1037/a0015575
- Harrington, J. (2006). An acoustic analysis of happy-tensing in the Queen's Christmas broadcasts. *J. Phonet.* 34, 439–457. doi: 10.1016/j.wocn.2005.08.001
- Harrington, J., Gubian, M., Stevens, M., and Schiel, F. (2019). Phonetic change in an antarctic winter. *J. Acoust. Soc. Am.* 146:3327. doi: 10.1121/1.5130709
- Hay, J., and Drager, K. (2010). Stuffed toys and speech perception. *Linguistics* 48, 865–892.
- Heiser, S. (2014). Berlins Bürgermeisterkandidat Saleh: Ein dubioses Hörproblem (Saleh: Berlins candidate for mayor. A dubious perception problem). taz 28.08.14. Available online at: <https://taz.de/Berlins-Buergermeisterkandidat-Saleh/!5034466/> (accessed September 23, 2020).
- Hilton, N. H., Rosseel, L., Smidt, E. M., and Coler M. (2016). “Using the IAT to understand the relationship between variant usage patterns and social meaning,” *Paper presented at Sociolinguistics Symposium 21*, Murcia.
- Jannedy, S., Mendoza-Denton, N., and Weirich, M. (2019). “Social capital in the production and perception of fine phonetic detail in Berlin,” in *The Sociolinguistic Economy of Berlin. Cosmopolitan Studies of Language, Diversity and Social Space* B. Schneider, T. Heyd, and F. von Mengden eds (Berlin: de Gruyter Mouton) 125–143. doi: 10.1515/9781501508103-006
- Jannedy, S., and Weirich, M. (2014). Sound change in an urban setting: category instability of the palatal fricative in Berlin. *Lab. Phonol.* 5, 91–122.
- Jannedy, S., and Weirich, M. (2017). Spectral moments vs discrete cosine transformation coefficients: evaluation of acoustic measures distinguishing two merging German fricatives. *J. Acous. Soc. Am.* 142, 395–405. doi: 10.1121/1.4991347
- Jannedy, S., Weirich, M., and Helmeke, L. (2015). “Acoustic analyses of differences in [ç] and [ʃ] productions in Hood German,” in *Proceedings of the 18th ICPHS*, Glasgow.
- Johnson, K., Strand, E. A., and D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *J. Phonet.* 27, 359–384. doi: 10.1006/jpho.1999.0100
- Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *J. Acous. Soc. Am.* 108, 1252–1263. doi: 10.1121/1.1288413
- Kachel, S., Simpson, A. P., Steffens, M. C. (2018). Do I sound straight? Acoustic correlates of actual and perceived sexual orientation and masculinity/femininity in men's speech. *J. Speech Lang. Hear. Res.* 61, 1560–1578. doi: 10.1044/2018_JSLHR-S-17-0125
- Kerlinger, F. N. (1984). *Liberalism and Conservatism. The Nature and Structure of Social Attitudes*. Hillsdale, NJ: Lawrence Erlbaum.
- Kerswill, P., Torgersen, E. N., and Fox, S. (2008). Reversing “drift”: innovation and diffusion in the London diphthong system. *Lang. Variat. Change* 20, 451–491. doi: 10.1017/s0954394508000148
- Kotsinas, U.-B. (1998). “Language contact in Rinkeby, an immigrant suburb,” in *Jugendsprache, Langue des Jeunes, Youth Language* J. K. Androutsopoulos and A. Scholz eds (Frankfurt am Main: Peter Lang Europäischer Verlag der Wissenschaften) 125–148.
- Labov, W. (2001). *Principles of Linguistic Change: Social Factors*, Vol 2, Hoboken, NJ: John Wiley & Sons.
- Leinonen, T. (2016). “Attitudes, salience, and accommodation – quantity in Finland-Swedish,” *Paper presented at Experimental Approaches to Phonetics & Phonology (ExAPP)*. Vienna.
- Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *J. Stat. Sofw.* 69, 1–33.
- Lewandowski, N., and Jilka, M. (2019). Phonetic convergence, language talent, personality and attention. *Front. Commun.* 4:18. doi: 10.3389/fcomm.2019.00018
- Li, F., Munson, B., Edwards, J., Yoneyama, K., and Hall, K. (2011). Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: implications for cross-language differences in speech-sound development. *J. Acoust. Soc. Am.* 129, 999–1011. doi: 10.1121/1.3518716
- Llomas, C., Watt, D., and MacFarlane, A. E. (2016). Estimating the relative sociolinguistic salience of segmental variables in a dialect boundary zone. *Front. Psychol.* 7:1163. doi: 10.3389/fpsyg.2016.01163
- Loudermilk, B. C. (2015). “Implicit attitudes and the perception of sociolinguistic variation,” in *Responses to Language Varieties: Variability, Processes and Outcomes* A. Prikhodkine and D. Preston eds (Amsterdam: John Benjamins). 137–156. doi: 10.1075/impact.39.06lou
- McCrae, R. R., and John O. P. (1992). An introduction to the Five-Factor Model and its applications. *J. Pers.* 60, 175–215.
- Munson, B., Jefferson, S. V., and McDonald, E. C. (2006). The influence of perceived sexual orientation on fricative identification. *J. Acoust. Soc. Am.* 119, 2427–2437. doi: 10.1121/1.2173521
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *J. Lang. Soc. Psychol.* 18, 62–85. doi: 10.1177/0261927x99018001005
- Nilsson, J., Leinonen, T. and Wenner, L. (2019). “Tracking change in social meaning: the indexicality of damped /i/ in rural Sweden,” in *Proceedings of the 9th International Conference on Language Variation in Europe (ICLaVE 9)*, J.-A. Villena-Ponsoda, F. Díaz-Montesinos, A.-M. Ávila-Muñoz and M. Vida-Castro eds (Malaga: ICLaVE) 46–157.
- Nortier, J. (2001). “‘Fawaka, what's up?’ Language use among adolescents in Dutch monoethnic and ethnically mixed groups,” in *Meeting at the Crossroads. Studies of multilingualism and multiculturalism in Oslo and Utrecht* A. Hvenekilde and J. Nortier eds (Oslo: Norvus Forlag) 61–73.
- Nosek, B., Greenwald, A., and Banaji, M. (2005). Understanding and using the implicit association test: II. Methods variables and construct validity. *Pers. Soc. Psychol. Bull.* 31, 166–180. doi: 10.1177/0146167204271418
- Nowak, P. M. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *J. Phonet.* 34, 139–152. doi: 10.1016/j.wocn.2005.03.001
- Pantos, A. J. (2010). *Measuring Implicit and Explicit Attitudes Toward Foreign-Accented Speech*. PhD thesis. Houston, TX: Rice University.
- Pantos, A. J., and Perkins, A. (2012). Measuring implicit and explicit attitudes toward foreign accented speech. *J. Lang. Soc. Psychol.* 32, 3–20. doi: 10.1177/0261927x12463005
- Pearce, J. W., and MacAskill, M. R. (2018). *Building Experiments in PsychoPy*. London: Sage.
- Plewnia, A., and Rothe, A. (2009). Eine Sprach-Mauer in den Köpfen? Über aktuelle Spracheinstellungen in Ost und West. *Deutsche Sprache* 37, 235–279.
- Podesva, R. J. (2007). Phonation type as a stylistic variable: the use of falsetto in constructing a persona. *J. Socioling.* 11, 478–504. doi: 10.1111/j.1467-9841.2007.00334.x
- Quist, P. (2005). “New speech varieties among immigrant youth in Copenhagen – A case study,” in *Sprachgrenzen Überspringen. Sprachliche Hybridität und*

- polykulturelles Selbstverständnis* V. Hinnenkamp and K. Meng eds (Tübingen: Narr) 145–161.
- R Core Team (2016). *R: A language and environment for statistical computing [Computer software manual]*. Vienna: R Statistics Foundation.
- Rampton, B. (2014). *Crossing: Language and Ethnicity Among Adolescents*. Routledge: Abingdon.
- Rossee, L., Speelman, D., and Geeraerts, D. (2018). Measuring language attitudes using the personalized implicit association test: a case study on regional varieties of Dutch in Belgium. *J. Ling. Geogr.* 6, 20–39. doi: 10.1017/jlg.2018.3
- Scherer, K. R. (1972). “Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences,” *Paper Presented at the Eastern Psychological Association Meeting*, Boston, MA
- Schirmer, A. (2019). *Deutsche Wortkunde: Kulturgeschichte des deutschen Wortschatzes*. (Vol. 929). Berlin: Walter de Gruyter GmbH & Co KG
- Schmid, S. (2020). Swiss German dialects spoken by second-generation immigrants: bilingual speech and dialect transformation. *J. Multiling. Multicult. Dev.* 41, 1–16. doi: 10.1080/01434632.2020.1730386
- Svendsen, B. A., and Röyneland, U. (2008). Multiethnolectal facts and functions in Oslo, Norway. *Int. J. Bilingual.* 12, 63–83. doi: 10.1177/13670069080120010501
- Tajfel, H., and Turner, J. C. (1986). The social identity theory of intergroup behaviour. in *Psychology of Intergroup Relations* S. Worchel and W. G. Austin eds (Chicago, IL: Nelson-Hall). 7–24.
- Van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.* 20, 580–591. doi: 10.1162/jocn.2008.20054
- van Bezooijen, R. (1988). “The relative importance of pronunciation, prosody, and voice quality for the attribution of social status and personality characteristics,” in *Language Attitudes in the Dutch Language Area* R. van Hout and U. Knops eds (Dordrecht: Foris) 85–103.
- van Bezooijen, R. (2002). “Aesthetic Evaluation of Dutch. Comparisons across Dialects, Accents and Languages,” in *Handbook of Perceptual Dialectology* M. Davis, and D. Preston, eds (Amsterdam: John Benjamins Publishing Company) 13–30. doi: 10.1075/z.hpd2.07bez
- Vande Kamp, M. E. (2002). *Auditory Implicit Association Tests*. Seattle: University of Washington.
- Watanabe, T., Nñez, J. E., and Sasaki, Y. (2001). Perceptual learning without perception. *Nature* 413, 844–848. doi: 10.1038/35101601
- Weirich, M., and Simpson, A. P. (2018). Gender identity is indexed and perceived in speech. *PLoS One* 13:e0209226. doi: 10.1371/journal.pone.0209226
- Wiese, H. (2015). “This migrants’ babble is not a German dialect!”: the interaction of standard language ideology and ‘us’/‘them’ dichotomies in the public discourse on a multiethnolect. *Lang. Soc.* 44, 341–368. doi: 10.1017/S0047404515000226
- Woodworth, R. S., Barber, B., and Schlosberg, H. (1954). *Experimental Psychology*. Janpath: Oxford and IBH Publishing.
- Wottawa, J., Adda-Decker, M., and Isel, F. (2016). *Putting German [ʃ] and [ç] in Two Different Boxes: Native German vs L2 German of French Learners*. San Francisco, CA: Interspeech.

Conflict of Interest: 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 © 2020 Weirich, Jannedy and Schüppenhauer. 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.



Common Ground Information Affects Reference Resolution: Evidence From Behavioral Data, ERPs, and Eye-Tracking

OPEN ACCESS

Maria Richter^{1†}, Mariella Paul^{1,2,3,4†}, Barbara Höhle^{1†} and Isabell Wartenburger^{1,3*†}

Edited by:

Pia Knoeferle,
Humboldt University of Berlin,
Germany

Reviewed by:

Ernesto Guerra,
University of Chile, Chile
Elli Tourtour,
Max Planck Institute
for Psycholinguistics, Netherlands

*Correspondence:

Isabell Wartenburger
isabell.wartenburger@uni-potsdam.de

[†]These authors share senior
authorship

*ORCID:

Maria Richter
orcid.org/0000-0001-8013-9333
Mariella Paul
orcid.org/0000-0002-5535-7141
Barbara Höhle
orcid.org/0000-0002-9240-6117
Isabell Wartenburger
orcid.org/0000-0001-5116-4441

Specialty section:

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

Received: 25 May 2020

Accepted: 10 November 2020

Published: 30 November 2020

Citation:

Richter M, Paul M, Höhle B and
Wartenburger I (2020) Common
Ground Information Affects Reference
Resolution: Evidence From Behavioral
Data, ERPs, and Eye-Tracking.
Front. Psychol. 11:565651.
doi: 10.3389/fpsyg.2020.565651

One of the most important social cognitive skills in humans is the ability to “put oneself in someone else’s shoes,” that is, to take another person’s perspective. In socially situated communication, perspective taking enables the listener to arrive at a meaningful interpretation of what is said (sentence meaning) and what is meant (speaker’s meaning) by the speaker. To successfully decode the speaker’s meaning, the listener has to take into account which information he/she and the speaker share in their common ground (CG). We here further investigated competing accounts about when and how CG information affects language comprehension by means of reaction time (RT) measures, accuracy data, event-related potentials (ERPs), and eye-tracking. Early integration accounts would predict that CG information is considered immediately and would hence not expect to find costs of CG integration. Late integration accounts would predict a rather late and effortful integration of CG information during the parsing process that might be reflected in integration or updating costs. Other accounts predict the simultaneous integration of privileged ground (PG) and CG perspectives. We used a computerized version of the referential communication game with object triplets of different sizes presented visually in CG or PG. In critical trials (i.e., conflict trials), CG information had to be integrated while privileged information had to be suppressed. Listeners mastered the integration of CG (response accuracy 99.8%). Yet, slower RTs, and enhanced late positivities in the ERPs showed that CG integration had its costs. Moreover, eye-tracking data indicated an early anticipation of referents in CG but an inability to suppress looks to the privileged competitor, resulting in later and longer looks to targets in those trials, in which CG information had to be considered. Our data therefore support accounts that foresee an early anticipation of referents to be in CG but a rather late and effortful integration if conflicting information has to be processed. We show that both perspectives, PG and CG, contribute to socially situated language processing and discuss the data with reference to theoretical accounts and recent findings on the use of CG information for reference resolution.

Keywords: perspective-taking, ERPs, eye-tracking, common ground, privileged ground

INTRODUCTION

One of the most important social cognitive skills in humans is the ability to “put oneself in someone else’s shoes,” that is, to take another person’s perspective. In communication, perspective taking enables the listener to arrive at a meaningful interpretation of what is said (sentence meaning) and what is meant (speaker’s meaning) by the speaker (Grice, 1989). Beyond linguistic information, visual and other contextual information is taken into consideration incrementally (see, for instance, the *Coordinated Interplay Account* by Knoeferle and Crocker, 2006; see also Knoeferle and Crocker, 2007; Crocker et al., 2010; Münster and Knoeferle, 2017). Especially in reference processing, the listener may have to take the speaker’s perspective in order to decode the speaker’s communicative intention. A referent can be a person, an object, or a concept, to which the speaker refers with a so-called referring expression. Speakers can choose different forms of referring expressions (e.g., a full noun phrase, a pronoun etc.) in discourse to optimize information transfer. For instance, when a referent is first introduced in discourse, the speaker commonly selects an indefinite noun phrase (e.g., a woman enters the bar). In subsequent discourse, the speaker refers back to that referent with a definite noun phrase (e.g., the woman), a pronoun (e.g., she), or another definite description (e.g., the beautiful lady), which adds information to the referent (Schumacher, 2018). Although experimental research confirmed that the speaker mostly provides sufficient but no redundant information when using referring expressions, over- and under-informative utterances occur (e.g., Deutsch and Pechmann, 1982; Engelhardt et al., 2006; Davies and Katsos, 2010; Morisseau et al., 2013). In these cases, listeners may then face multiple possible referents within the linguistic and/or non-linguistic context. In order to understand which referent the speaker was referring to, the listener has to take the speaker’s perspective. This requires the calculation of mentally and/or perceptually shared information by both interlocutors, which is often called common ground (CG) information (e.g., Clark et al., 1983). With the present study we intend to better understand, if, when, and how information in privileged and CG is integrated during utterance processing.

A body of research has been concerned with this question. For a long time, existing parsing theories took two rather different, apparently contradictory views. On the one side there were theories that assume autonomous lexical and syntactic activation with contextual and other pragmatic constraints, such as CG, entering the parsing process only at a later stage at which the different sources of information are integrated (e.g., Ferreira and Clifton, 1986; Keysar et al., 2000; Epley et al., 2004b; Barr, 2008; Kronmüller et al., 2017). We will refer to these accounts as “*late integration*” accounts. On the other side, constraint-based theories assumed that all available information sources do immediately interact during the parsing process and guide the interpretation of a sentence (e.g., Altmann and Steedman, 1988; Spivey-Knowlton and Sedivy, 1995; Trueswell et al., 1999; Nadig and Sedivy, 2002; Hanna et al., 2003; Snedeker and Trueswell, 2003; Snedeker and Yuan, 2008). We will refer to these accounts as “*early integration*” accounts.

Data supporting the assumption of a late integration of pragmatic information during the parsing process stemmed from Keysar et al. (2000). They used a version of the referential communication game, also called director’s task (Glucksberg et al., 1975; Krauss and Glucksberg, 1977). In this game, objects are placed in a vertical array. A confederate (experimenter) sits on one side of the array and instructs an addressee (participant) on the other side of the array to manipulate the objects in a certain manner. Crucially, some of the objects are hidden from the experimenter’s view, giving the participant privileged access to them. In order to follow the experimenter’s instruction correctly, that is, to pick the correct referent, participants have to consider which objects are shared for both interlocutors (i.e., are in CG). In Keysar et al. (2000) the overt responses revealed that participants based their decision on CG information in most of the cases (around 80%), that is, they picked the objects in CG. However, the eye fixation data showed that the participants initially fixated the privileged object [i.e., the competitor that was exclusively visible for the participant, that is, in privileged ground (PG)] and only later turned their eyes to the object in CG (i.e., the target). This interference effect produced by the privileged object supports the view that CG does not immediately restrict the search for referents. CG information is rather integrated late with effort, after an initial egocentric interpretation might have even led to egocentric errors, that is, picking an object that is not in CG (*Egocentrism Account*). For the (limited) effects on cultural backgrounds on egocentric errors see Wu and Keysar (2007) and Wang et al. (2019).

Further evidence for late integration accounts was obtained by Barr (2008). He used a slightly different method to instantiate the CG vs. PG objects. Here, participants directed already more fixations to the CG objects before any verbal instruction was given to them. This indicated an anticipation that the confederate would refer to CG objects. However, after the verbal instruction (e.g., “click on the bucket”) the participants needed longer time to orient their gaze to the target object when an object with a label that constituted a phonological competitor (i.e., competitor condition, e.g., bucket-buckle) was present compared to a control condition without competitor. This held true for all competitors, independent of whether they were presented in CG or PG. Crucially, when comparing the effect of interference of competitors in CG vs. PG, no differences were revealed. This suggested that CG information did not attenuate the interference of competitors, as constraint-based theories would assume. These results were interpreted in the framework of the *Autonomous Activation Account*. It proposes that listeners initially actively attempt to take a speaker’s perspective in anticipation of a linguistic expression (i.e., in the phase before any verbal instruction is given). Then they fail to fully integrate CG information, because the lexical information given by the speaker autonomously activates the information in PG (i.e., the competitor).

In contrast to these two accounts, that considered CG integration as a rather late and effortful process in which egocentric errors may occur (Keysar et al., 2000; Barr, 2008; Wang et al., 2019), earlier and cognitively less demanding effects of perspective taking on reference resolution had also been found.

This was the case in similar tasks when linguistic markers, such as color terms (e.g., red), (in)definite expressions (e.g., the/one of the), or scalar adjectives (e.g., big/small) were available to narrow down the relevant contrasts. For example, Hanna et al. (2003) used a version of the referential communication game in which a referring expression was either ambiguous with respect to two objects in CG or in which one of these objects was privileged. For instance, the confederate instructed listeners (participants) to “put the blue circle above the red triangle”. In conditions with two red triangles in CG, participants were equally likely to look at either. When one object was privileged, participants were more likely to look at the object in CG from the earliest moments and were faster to choose it, hence supporting an early integration. Also, Heller et al. (2008) presented displays which contained two pairs of size contrasting objects, for instance, a big duck (target) and a small duck (target-contrast), a big box (competitor) and a small box (competitor-contrast). There were two conditions: In the shared condition, all objects were in CG. In the privileged condition, one of the items belonging to a competitor-contrast (e.g., the small box) was in PG. Listeners received instructions with scalar adjectives, for example, “pick up the big duck”. The results showed that listeners immediately used the distinction between CG and PG. They thus integrated CG early, challenging a possible egocentric-first heuristic. This is consistent with other studies that found an early effect of CG information (e.g., Nadig and Sedivy, 2002; Hanna and Tanenhaus, 2004; Brown-Schmidt et al., 2008; Brown-Schmidt, 2012; Ferguson and Breheny, 2012). While the results of Heller et al. (2008) speak against an automatic egocentric bias in interpreting perspective-sensitive language, the authors do not claim a CG heuristic that directs attention only to mutual information. Instead they suggest that listeners are aware of the common or privileged status of information and use this distinction early in real-time reference resolution.

Other research has shown that the use of CG information, as well as a reduction of egocentric biases, is facilitated by rich discourse contexts such as when conversational context explicitly establishes what the confederate does and does not know through the use of questions (e.g., “What’s above the cow?”) (Brown-Schmidt et al., 2008). Similarly, active engagement in a task leads to earlier inferences about others’ perspectives, and boosts the immediate use of this information to anticipate others’ actions compared to passive observers (Ferguson et al., 2015). Finally, the motivation of participants plays a role: when there is a high motivation or incentive for integrating perspectives and when sufficient cognitive resources are available, participants can activate perspective taking abilities early on (Epley et al., 2004a; Cane et al., 2017). In sum, these findings indicate that CG information can be immediately processed, even involuntarily, and used early in the parsing process, contradicting late integration accounts.

Recent approaches have considered neurobiological data to disentangle early and late integration accounts of CG processing. From a neurobiological perspective the human brain enables rapid communication through a continually implemented perception-action cycle. That is, sensory input is perceived (e.g., the confederate’s speech), and generates a particular action (e.g., one’s own verbal response), which in turn results in a

self-generated sensory input, and, again, in a certain response (Bornkessel-Schlesewsky and Schumacher, 2016). Crucially for the needs of CG integration, this perception-action cycle also allows for predictive coding, and, in case of a mismatch between prediction and input, instantiates an update and the modification of the internal model (Bornkessel-Schlesewsky and Schumacher, 2016). The neuronal implementations of these mechanisms have been investigated in recent years with event-related potentials (ERPs). A late positive ERP component (starting around 400–500 ms and lasting around 1000 ms post stimulus onset) was associated with reconceptualization or repair mechanisms (Schumacher et al., 2018), and with reference processing (Schumacher, 2009). In her neurocognitive model of reference resolution, Schumacher (2009) suggested that a late positive ERP component reflects additional processing costs that arise whenever a prior discourse representation has to be updated or modified (e.g., with the emergence of a new referent). Other ERP studies investigating referential aspects of language comprehension also revealed ERP effects such as the P600 (e.g., Osterhout and Mobley, 1995; van Berkum et al., 1999; Harris et al., 2000). Hoeks and Brouwer (2014) refer to the internal model during discourse comprehension as *Mental Representation of what is Communicated*, MRC. In their view, the P600 reflects the construction or revision of an MRC. If establishing reference turns out to be impossible, or leads to an implausible interpretation, a P600 will ensue, reflecting the reorganization of the MRC. A P600 may also appear in the absence of such serious problems, when a discourse entity needs to be accommodated, or when the referring expression needs some “pre-processing” before the antecedent can be successfully identified.

In addition to the P600 ERP effects there is evidence of another ERP component involved in referential processing. Referentially ambiguous nouns (e.g., “the girl” in a two-girl context) or pronouns (e.g., “David noticed John when he stood up.”) elicited a frontally dominant and sustained negative shift, called Nref effect (van Berkum et al., 1999; Nieuwland et al., 2007, for a review see van Berkum et al., 2007). Nieuwland et al. (2007) highlight that the frontal negative shift reflects genuine referential ambiguity in the current model of the discourse. Hoeks and Brouwer (2014) instead propose that each referring expression elicits an Nref response as soon as the search for an antecedent is instantiated.

Recently, Sikos et al. (2019) reported an Nref-effect as a marker of referential ambiguity in a perspective taking task. In their study, participants were asked to pick a referent from a display of four animals (e.g., “Click on the brontosaurus with the boots”) by a speaker who could only see three of the animals. A competitor (e.g., a brontosaurus with a purse) was either mutually visible, visible only to the listener, or absent from the display. Results showed that the mutually visible competitor elicited a referential ambiguity as reflected by an Nref-effect. Crucially, when listeners had privileged access to the competitor, the ERPs did not show evidence for a referential confusion—although participants were slower when the privileged competitor was present. The authors concluded that participants did not consider the competitor in PG to be a candidate for reference. This interpretation is in line with early integration accounts that allow a rapid integration of

pragmatic information during online language comprehension and hence speak against late integration accounts. However, the finding is incompatible with “egocentric errors” in behavioral studies (e.g., Keysar et al., 2000), in which participants apply an egocentric interpretation strategy and choose the competitor in PG as the target item—and hence obviously consider it as a potential candidate for reference. Furthermore the finding is incompatible with interference effects from objects in PG that have been shown in a series of experiments by Barr (2008). Note, however, that the discussed studies also differ in the actual task design applied.

Barr (2016) called for the need to focus on the underlying processes and the use of joint data analysis routines. In the same year, Heller et al. (2016) aimed at solving the above mentioned traditional contradictions of early and late integration accounts by implementing the data of the original eye-tracking studies of Keysar et al. (2000) and Heller et al. (2008) in a Bayesian model of reference resolution. The model suggests that referring expressions are not interpreted relative to the CG or to one’s egocentric knowledge, but rather reflect the *Simultaneous Integration* of the two perspectives. In their probabilistic model, both the egocentric and the CG perspective are active in their referential domains (the referential domain is a contextually restricted set which is inferred and updated according to the current situation; here an egocentric domain and a CG domain is implemented). To gain information about the target referent, listeners simultaneously weigh evidence from both perspectives (Heller et al., 2016).

To disentangle the predictions of early and late integration accounts we here further investigated how listeners integrate egocentric and CG perspectives by adapting the well-established referential communication game of Keysar et al. (2000) to a computerized version. While we collected reaction time (RT) and accuracy data, we applied eye-tracking as well as electroencephalography (EEG) to study the timing and the underlying mechanisms of CG integration. Both methods, eye-tracking and EEG, offer a very high temporal resolution. They are therefore especially suitable to explore the temporal dynamics of the integration of CG information. Importantly, while eye movements might be affected by attentional processes that are unrelated to referent identification, EEG might be better suited to gain knowledge about the functionally distinct processes that underpin perspective taking. Our first study (ERP, Experiment 1) thus offers the opportunity to disentangle different aspects of the comprehension of referential expressions. In addition, EEG allows to draw inferences about the underlying neural mechanisms of CG integration and can be directly compared to the findings of Sikos et al. (2019). Our second study (eye-tracking, Experiment 2) with a mostly identical design to the ERP-Experiment allows for a descriptive alignment of eye-tracking results with our ERP data and provides further insights into the interaction of language comprehension and the perception of the visual world (for a recent short methodological overview see Rodriguez Ronderos et al., 2018). In addition to the ERP analysis, we appended an exploratory time-frequency analysis (TFA) of the EEG data in the **Supplementary Material**, which might provide insights about the mechanisms underlying CG processing. At the

behavioral level, Experiment 2 can be taken as an attempt to replicate Experiment 1.

The *Simultaneous Integration Account* would predict that both the egocentric and the CG perspective are active when engaged in referential communication. However, depending on the evidence triggered by the specific task or the array, either egocentric or CG behavior may be enforced (Heller et al., 2016). Since our design was very similar to that of Keysar et al. (2000), egocentric behavior may guide at least initially the perspective taking behavior. This would lead to a rather late, and effortful integration of CG information during the parsing process. If participants first consider the object in PG to be the target and then switch to the (correct) target object in CG in a competitor (here: conflict) condition, some kind of discourse updating or reconceptualization has to take place. According to previous ERP studies in the field (see above), this late and effortful integration of CG would elicit a late positivity in the ERPs. This expectation is therefore in contrast to the findings of Sikos et al. (2019) who argue that the object in PG is not considered to be a potential referent in the display and therefore would not elicit a specific ERP response (in their case an Nref component). For the behavioral and eye-tracking data, we expect to replicate the findings of Keysar et al. (2000). That is, we expect more errors and/or longer RTs in the conflict condition in which participants probably have to suppress their egocentric bias. Accordingly, eye-tracking should reveal earlier looks to the competitor in PG, and later looks to the target in CG in the conflict condition in comparison to a condition without conflict. On the other hand, if the clear instruction, the integrated practice phase, and the high repetition rate in our experimental design promotes the CG perspective taking behavior, CG information would be considered immediately. In this case we would not expect discourse updating and thus no late positivity in the ERPs. Rather, we would expect no effects in ERP signatures as a result of CG integration.

MATERIALS AND METHODS

Experiment 1: EEG

Participants

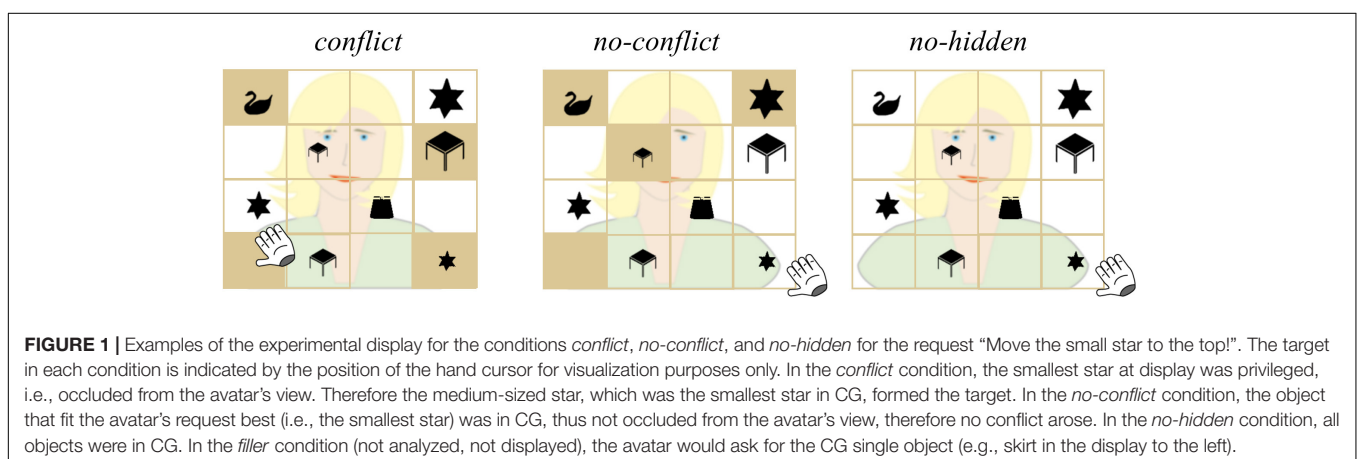
Thirty-six students of the University of Potsdam (17 female, *M* 24.6 years, age range 20–31 years) participated in the study. All participants were native German speakers, reported normal or corrected-to-normal vision, normal hearing, no neurological problems, and were right-handed as assessed by a German version of the Edinburgh Handedness Inventory (Oldfield, 1971). Nine participants were excluded for further ERP analysis due to technical problems during the recording (*n* = 2), or because less than 50% artifact-free trials survived the artifact rejection procedure in the critical conditions (*n* = 7). Thus, 27 participants entered the final ERP analysis (12 female, *M* = 24.8 years, age range 20–31 years). All participants gave written informed consent according to the local Ethics Committee of the University of Potsdam. Participants received course credits or financial compensation for their participation.

Materials and Design

Participants played a computerized version of the referential communication game (see Keysar et al., 2000). In this game, a virtual 4×4 grid was presented on a computer screen. Each display of the grid contained two object triplets with three differently sized objects (i.e., small, medium, big), two single objects (distractors), and eight empty slots (see **Figure 1**). We used two object triplets to prevent that participants would know, after a few trials, from the beginning of the display, which objects could potentially become the target. Thirty-two different objects were used to build the object triplets, and 18 other objects were used as single objects. All objects were black-and-white drawings representing man-made concrete objects like clothes, furniture or vehicles [e.g., Rock (skirt), Tisch (desk), or Zug (train)] or natural entities like celestial bodies, fruits, or animals [e.g., Stern (star), Apfel (apple), or Frosch (frog)]. The distribution of objects in the 16 slots of the 4×4 grid was fully randomized and changed across trials and participants. The virtual confederate (henceforth termed avatar) was displayed behind the grid and provided auditory instructions of the form: “[Move the] [target size] [target object] [to the top]!” (e.g., “[Move the] [big/small] [star/apple/frog/. . .] [to the top]!”; German: “[Bewege den] [großen/kleinen] [Stern/Apfel/Frosch. . .] [nach oben]!”). Notably, in German, both the determiner and the adjective are marked for gender. Therefore all nouns used were masculine, indicated by the gender-marked accusative direct determiner “den” (the), and the adjective-suffix “-en”. Due to this, a possible disambiguation before the onset of the noun was avoided. The instructions were pre-recorded by a trained native German female speaker and presented phrase by phrase with a fixed timing (0 ms[Bewege den] 1000 ms[kleinen/großen] 1650 ms[target object] 2950 ms[nach oben]) (0 ms[Move the] 1000 ms[small/big] 1650 ms[target object] 2950 ms[to the top]). Accordingly, the critical noun phrase (i.e., the [target object]) always started 1650 ms after the beginning of the auditory onset of the sentence (i.e., the [Bewege den]). The mean length of the nouns was 785 ms (± 135 ms). Nevertheless, the overall sentences sounded prosodically well-formed as the phrases were cut out of natural recordings of the full sentences that were spoken in a relatively slow speech rate by the trained speaker. Participants had to select

the target object via mouse click and had to drag and drop it on a field above the virtual grid. Dragging and dropping of the target objects were self-paced, meaning that participants were free to click on the object as soon as they had made their decision.

The crucial feature of the referential communication game is the manipulation of visual access to certain objects in the grid from the perspectives of the avatar and the participant, respectively. In three out of four conditions (conflict, no-conflict, filler), four slots in the grid contained a backboard that occluded their content from the avatar's view. In the fourth condition, no-hidden, all slots were in CG. The no-hidden condition served as a control for effects of the mere presence of occluded slots in the grid (i.e., if there would be no differences between the no-hidden and the no-conflict condition, we could conclude, that the mere presence of occluded slots in the grid did not induce some unspecific computation of ground or did not affect general attentional processes). Positions of the backboards randomly changed from trial to trial. Three out of four of these occluded slots contained objects: two contained one object of the two object triplets, and one slot contained one of the two single objects (distractor). One of the four slots was empty. Since the participant had privileged access to the objects in these slots, we term them privileged objects (PG objects). For all other objects in the grid, both the avatar and the participants had visual access. They were in CG, and are henceforth termed CG objects. We created four different conditions: in *conflict trials*, one of the privileged objects fit the avatar's request best from the perspective of the participant (e.g., the small star). In this condition the smallest star was a privileged object. Participants then had to consider which objects were visually shared, thus in CG, to select the correct object (“target”; e.g., the medium-sized star). In *no-conflict trials*, the object that fit the avatar's request best from the perspective of the participant was in CG. In *no-hidden trials*, there were no occlusions at all and therefore all objects were in CG (please see **Figure 1** for a detailed example of experimental displays). In *filler trials*, the target object was one of the two single objects (distractors) that was in CG (not at display in **Figure 1**). In all experimental conditions, the onset of the noun (i.e., [target object]) marked the point of disambiguation. In total, the EEG experiment consisted



of 256 trials, 64 per condition (*conflict*, *no-conflict*, *no-hidden*, *filler*). Filler trials were not further analyzed. The distribution of conditions throughout the experiment was fully randomized and changed across participants.

Procedure

Participants were seated approximately 70 cm in front of the computer screen. A computer mouse was placed at a comfortable distance on a desk in front of the screen. RTs and accuracy measures were obtained via mouse click. All participants used their dominant right hand to navigate the computer mouse. In advance of the experimental phase, participants were instructed to mind the avatar's perspective, which was supported by rotating the grid and showing the avatar's view on the grid. This demonstrated to the participants that the avatar was not able to see the objects that were in slots with a wooden background. In addition, participants underwent a practice phase with nine practice trials during which they received corrective feedback (two-step instruction similar to Wang et al., 2016). For instance, participants received the instruction "[Move the] [small star] [to the top]!". If they then chose the privileged object that is the smallest star at display in a *conflict* trial, the feedback they received from the avatar was: "Oh, I didn't see that star. I meant the other small star!".

Every trial started with a fixation cross, which was presented in the center of the screen for 1000 ms. Then, the empty grid with occlusions in four varying positions (with the exception of the no-hidden condition) was filled with the objects for 750 ms. Participants had time to view the grid for 500 ms. Then the avatar gave the auditory instruction, which was provided via headphones. Once the participants had made their choice, they clicked on the target with the computer mouse and dragged the object to a rectangle placed above the grid on the computer screen. Then, the next trial started.

During the test phase, the EEG was recorded and RTs and accuracy rates were measured. The RTs were measured for the first click on the target object starting from noun onset. The stimulus presentation and randomization was controlled by Presentation® software version 18.1 (Neurobehavioral Systems). After the experiment participants were asked to fill out a debrief form about their intuitions concerning the purpose of the study and the strategies they used.

EEG Recordings

The EEG was recorded with a 32-channel active electrode system (Brain Products®, Gilching, Germany). 27 electrodes were placed on the scalp within an elastic soft cap (EASY CAP®, Inning, Germany) according to the 10/20 system (American Clinical Neurophysiology Society, 2006) at the following scalp positions: F7/8, F5/6, F3/4, FC3/4, C5/6, C3/4, CP3/4, P3/4, P7/8, PO3/4, AFz, Fz, FCz, Cz, CPz, Pz, and POz. The ground electrode was placed at FP1. The electrooculogram (EOG) was recorded with four additional electrodes. To detect blinks and vertical eye movements (vertical EOG), one electrode was placed above and one electrode below the participant's right eye. To detect horizontal eye movements (horizontal EOG), electrodes were placed at the outer canthi of the left and the right eye. Impedances

were kept below 5 kOhm. The EEG data was recorded with a sampling rate of 1000 Hz. The left mastoid served as the online reference electrode, but the recording was re-referenced offline to the averaged signal of the left and right mastoids.

Behavioral Data Analysis

The behavioral data comprised accuracy and RT measures. For the accuracy data, correct and incorrect responses were counted for each participant per condition (*conflict*, *no-conflict*, *no-hidden*). The total number of correct responses per condition was then transformed into percentage values to determine the accuracy rate. The accuracy rate for each condition and participant was then averaged across participants ($n = 27$). RTs were measured in ms relative the onset of the critical noun. Prior to the analysis, RTs with negative values (i.e., reactions before the onset of the noun), wrong responses, and "double clicks" on the target were removed using MS Excel® (Version 2010). The remaining RTs were averaged for each condition per participant and then averaged across participants. To detect differences in behavior in relation to the three experimental conditions, an ANOVA with Condition (three levels: *conflict*, *no-conflict*, *no-hidden*) as within-subjects factor was run for both accuracy and RT measures separately. Whenever the main effect of condition reached significance ($p < 0.05$), *post-hoc* paired-samples *t*-tests controlled for multiple comparisons (Bonferroni corrected $p = 0.017$) were calculated. This was done to further examine the differences between conditions. Descriptive statistics as well as ANOVAs were carried out with the statistics software IBM® SPSS Statistics for Windows (Version 23.0).

ERP Data Analysis

For ERP data preprocessing, the Brain Vision Analyzer software (version 2.0.2; Brain Products®, Gilching, Germany) was used. Raw data were filtered offline by applying a Butterworth zero-phase filter (low cutoff: 0.3 Hz; high cutoff: 70 Hz; slope: 12 dB/oct) to exclude slow signal drifts and muscle artifacts. In addition, a notch filter of 50 Hz was applied to remove line noise induced by electrical devices during testing. Artifacts caused by vertical and horizontal eye movements were corrected by the algorithm of Gratton et al. (1983). An automatic artifact rejection procedure was used to reject blinks, flat signals, and drifts in the time window of -200 to 1500 ms relative to the onset of the critical noun in the target sentence. The following criteria were set to automatically mark channels as bad: Maximal allowed voltage step: $20 \mu\text{V/ms}$, maximal allowed difference of values: $75 \mu\text{V}$ per 150 ms time interval, minimal allowed amplitude: $-75 \mu\text{V}$, and a maximal allowed amplitude: $75 \mu\text{V}$, and lowest allowed activity in intervals: $0.5 \mu\text{V}$. Importantly, each trial was additionally examined visually and any remaining eye-blinks or eye movement artifacts were removed. Participants for whom less than 50% of trials in the noun onset time window survived the artifact rejection procedure were removed from further analysis ($n = 7$). Moreover, only trials in which participants selected the correct object (i.e., the target), entered the final analysis. In total, 23 trials (out of 6912 trials of the remaining 27 participants) with an incorrect response were removed (*conflict* condition: two trials, *no-conflict* condition: eight trials, *no-hidden* condition: 13

trials). Overall, the artifact rejection procedure and the deletion of incorrect trials resulted in a rejection of $21.55 \pm 12.91\%$ of trials (conflict condition: $22.05 \pm 11.94\%$, no-conflict condition: $21.12 \pm 13.42\%$, no-hidden condition: $21.47 \pm 13.38\%$). The amount of excluded trials did not differ across conditions as revealed by a repeated measures ANOVA with the factor condition as within subjects factor ($F(3,78) = 1.23$; $p = 0.303$, $\eta^2 = 0.045$).

For statistical analysis, we computed non-parametric cluster-based permutation analyses. This test calculates a cluster t -statistic that sums across temporally and spatially adjacent point-wise t -values that exceed a predefined threshold. This cluster t -statistic is then compared to a null-hypothesis distribution of cluster t -values that are generated via a Monte Carlo permutation approach. We used 1000 random permutations to generate a distribution of the null hypothesis with sufficient precision to control family wise error rate to $\alpha < 0.05$, as suggested in Maris and Oostenveld (2007). The statistics was run two-tailed and within-subjects, with a minimum number of two significant ($\alpha < 0.05$) electrodes to form a cluster. 50 ms running time windows were calculated, and considered as significant when they were significant over the entire time window. The cluster-based permutation analysis was performed with the open source software Fieldtrip for EEG/MEG analysis (Oostenveld et al., 2011) in MATLAB® (2015b, MathWorks, Natick, MA, United States).

In addition to the ERP analysis, an exploratory TFA was performed on the EEG data. This was done since amplitude increases and decreases in specific frequency bands may provide further information about the underlying brain functions. Details regarding the analysis and full results are provided as **Supplementary Materials**.

Experiment 2: Eye-Tracking Participants

Twenty-nine native speakers of German (15 female; mean age: 24.3, range: 18–34) participated in the experiment. All of them gave written consent prior to the experiment, were naïve to the purpose of the study, and did not participate in Experiment 1. The participants received either course credits or financial compensation for their participation. All participants were right-handed as assessed by a German version of the Edinburgh Handedness Inventory (Oldfield, 1971), and had normal or corrected-to-normal vision. Two participants were excluded from the experimental cohort due to technical problems during the recordings, which resulted in an experimental breakup. Thus, 27 participants (14 female, mean age: 24.2, range 18–34) entered the final behavioral and eye-tracking analyses.

Materials and Design

Materials and design were almost identical to Experiment 1 (please refer to section “Materials and Design”). However, the amount of trials was reduced from 256 to 112 trials, with 28 trials per condition (*conflict*, *no-conflict*, *no-hidden*, *filler*). In addition, the position of the target and the privileged object in the grid was constrained so that these two objects could not appear in horizontally, vertically, or diagonally neighboring slots of the

grid. This was done to minimize the misclassification of looks in the dense 16-slot-grid during data analysis. As in Experiment 1, filler trials were not further analyzed.

Procedure

The eye-tracking camera was attached at the middle of the lower edge of the PC monitor. The background screen color was set to dark gray. The participants were seated at a distance of 62–67 cm from a 22 inch (1680 × 1050 pixel) TFT PC monitor. The sitting position of the participants and the eye-tracking camera were adjusted checking that the pupils were recognized by the eye tracker in the center of a virtual box of the iViewRED-m application (SensoMotoric Instruments®, Teltow, Germany). The system was calibrated to the participants’ right eyes with a nine-point automatic calibration. For the calibration, a black dot was presented at different positions on a light gray background. In case of suboptimal calibration results, the procedure was repeated until the spatial precision of the gaze was classified as adequate by the system and by the experimenter.

Identical to Experiment 1, after calibration, a clear introduction was given by the avatar. A practice phase followed with a two-step corrective feedback (for details please see section “Procedure”). Then the experimental phase started with the presentation of four blocks á 28 trials. The task was the same as in Experiment 1. Participants used their right hand to navigate a computer mouse in order to click on the target and drag the object to a rectangle placed above the grid on the computer screen. After each block, a short, self-paced pause was inserted. Since participants were allowed to minimally move during the pause, the calibration procedure was repeated in advance of each block. During the test phase, eye-gaze data was recorded and RTs and accuracy rates were measured. The RTs were measured for the first click on the target object starting from noun onset.

Eye-Tracking Recordings

Eye movements were recorded with an SMI RED-m Eye-Tracker (SensoMotoric Instruments®, Teltow, Germany). Only the participants’ right eyes were tracked using SMI’s “smart right binocular mode”. With this mode, the system tracks gaze data every 8.33 ms (sampling frequency 120 Hz) and offers a spatial accuracy of 0.5–1°. The recovery time after track loss lies at 250 ms.

Behavioral Data Analysis

The behavioral data of the eye-tracking cohort was analyzed in the same way as the behavioral data of the EEG cohort. Please refer to section “Behavioral Data Analysis” for details of the analysis.

Eye-Tracking Data Analysis

The eye-tracking analysis (preprocessing, statistics) was performed with the free statistics software R® (R Core Team, 2015). Only trials with correct responses entered the final analysis ($M = 99.67 \pm 0.44\%$). Data points for which the eye tracker could not determine the gaze position were removed. The overall track loss was on average 4.3%. Since the objects could appear in each of the 16 slots of the vertical array, we created 16 equally sized (170 × 170 pixels) spatial areas of interest (AoI) corresponding

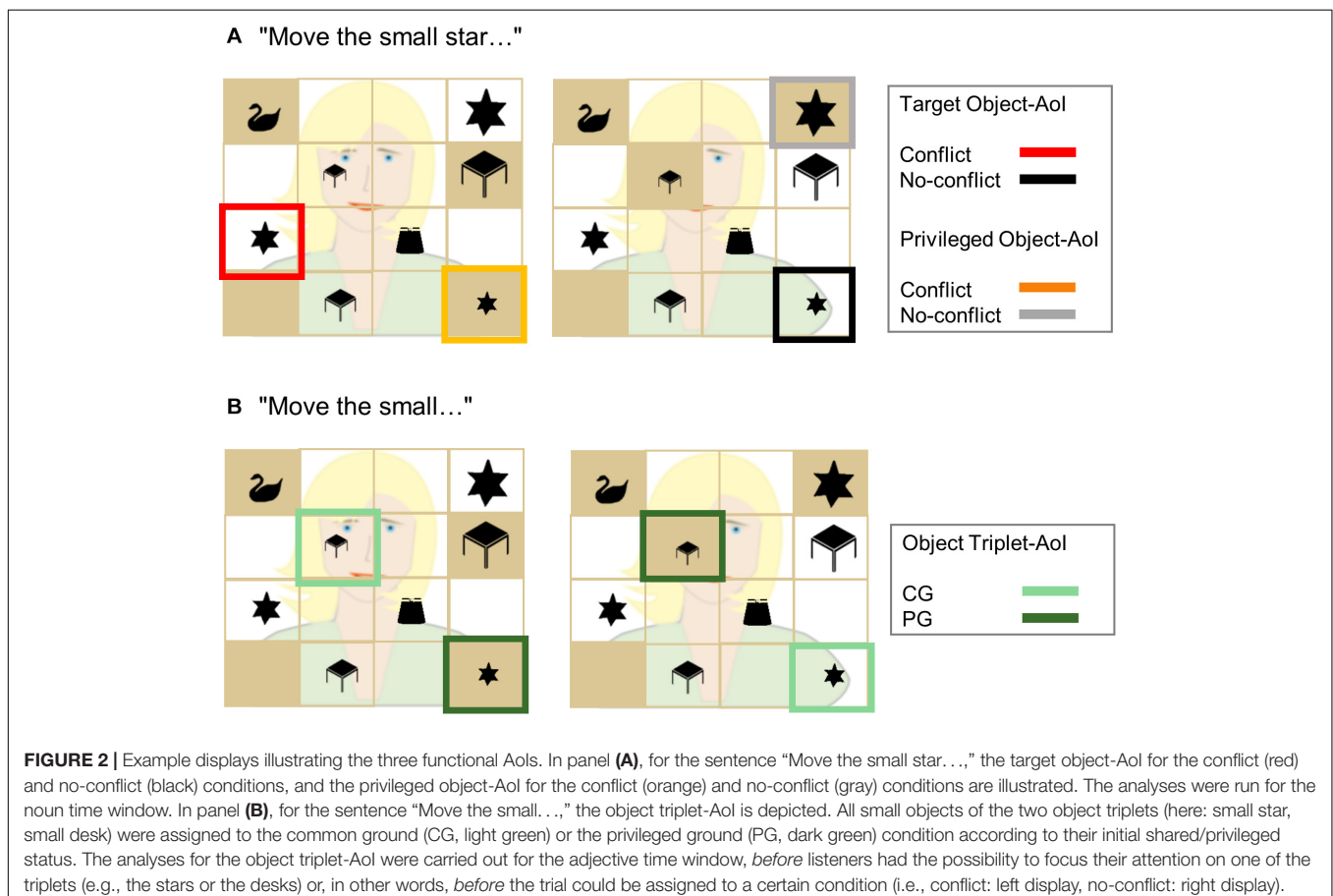
to the slots in the array. All gaze positions were automatically classified as being in one of the 16 AoIs or not.

Next, we defined functional AoIs. The first functional AoI formed the area in which the target appeared. We will call this “target object”-AoI. For instance, given the instruction “Move the small star to the top!” the target would be the medium-sized star in the conflict condition and the small star in the no-conflict condition (for an illustration please refer to **Figure 2A**). The second functional AoI formed the area in which an object in PG appeared. We will call this “privileged object”-AoI. As an example, given the instruction “Move the small star to the top!” the privileged object would be the small star in the conflict condition and the big star in the no-conflict condition, both of which were hidden from the avatar’s view. Please note that only in the conflict condition, the object in PG provided a “real” competitor to the target object, since it had the potential to interfere with the target in CG (e.g., the medium sized star). In the no-conflict condition, however, participants were not expected to look at the object in PG (“the big star”). Still, it represents the object of comparison in PG when comparing the conflict and the no-conflict conditions (**Figure 2A**) (an alternative analysis of looks to the medium-sized object is presented as **Supplementary Material**). The third functional AoI incorporated all small or big objects of the two presented object triplets of the conflict and no-conflict condition trials (e.g., the small star and the small desk

when the adjective was “small”; the big star and the big desk when the adjective was “big”; **Figure 2B**). Some of these objects were in CG and some of them were in PG.

Crucially, we created the “object triplet”-AoI to investigate early anticipatory looks to the objects in CG and PG, before the conditions (i.e., conflict, no-conflict) became evident for the participant. That is, the possibility to assign a sentence to one of the conditions only started with the presentation of the noun phrase, in dependency of the target item and the given occlusions. After the presentation of the noun, only one of the object triplets contained the target (e.g., the small star). However, in the preceding adjective time window, on which we focused in this analysis, participants could not know to which object triplet the target item belonged (e.g., the small star OR the small desk). We thus gave the third functional AoI the condition-neutral term “object triplet”-AoI (**Figure 2B**). We analyzed the looking behavior in the adjective time window comparing the looks to the objects in CG and PG of both object triplets. Crucially, if participants anticipated objects in CG to be the target, we should see more looks to the objects in CG compared to objects in PG in the adjective time window already.

The gaze data was first averaged across trials for each condition within participants, and then across participants for the grand average. Proportions of looks to the functional AoIs were calculated (values between one and zero).



Similar to the ERP analysis, we used the nonparametric cluster-based permutation analysis in order to detect reliable differences between conditions across time. Due to the strengths of nonparametric cluster-based permutation analyses (i.e., better control for multiple comparisons and the reduction of Type I errors) these analyses have also become more common in eye-tracking research (Holzen and Mani, 2012; Barr et al., 2014). For preprocessing and the statistical analysis we used the R package “eyetrackingR” (Dink and Ferguson, 2015). Non-AoI looks were treated as missing data. Then, we defined three time windows of interest: the adjective time window (1000–1650 ms post auditory onset), the noun time window (1650–2650 ms post auditory onset), and the post noun time window (2650–3650 ms post auditory onset). In each of these broader time windows, a time course based on 50 ms time bins was created, and proportion of looking times within each time-bin was summarized. Next, within the summarized time-bin data, adjacent time bins that passed the test-statistic threshold ($\alpha < 0.05$, two-tailed t -test), were assigned into groups (clusters). This output was taken for the cluster-based permutation analysis (Maris and Oostenveld, 2007). This analysis took a summed statistic for each cluster, and compared it to the “null” distribution of sum statistics obtained by shuffling the data and extracting the largest cluster from each resample. Parallel to the ERP analysis, 1000 iterations were performed in the bootstrap resampling procedure.

RESULTS

Experiment 1: EEG Behavioral Findings

Accuracy was high for all participants ($n = 27$) across all conditions ($M = 99.67 \pm 0.40\%$ accuracy rate). Errors occurred in all three conditions (accuracy conflict: $M = 99.88 \pm 0.42\%$; accuracy no-conflict: $M = 99.54 \pm 0.85\%$; accuracy no-hidden: $M = 99.25 \pm 1.40\%$). An ANOVA with Condition (3 levels: conflict, no-conflict, no-hidden) as within-subjects factor revealed no effect of condition ($F(2,52) = 2.77$; $p = 0.09$, $np^2 = 0.096$). However, perspective taking had its costs as revealed by RT measures: RTs measured relative to the onset of the critical noun (e.g., “star”) showed that participants were on average 184.22 ms slower in the conflict condition ($M = 1469.07 \pm 364.17$ ms) compared to the means of the no-conflict ($M = 1289.38 \pm 375.18$ ms) and no-hidden conditions ($M = 1280.30 \pm 367.80$ ms). An ANOVA with Condition (three levels: conflict, no-conflict, no-hidden) as within-subjects factor revealed a significant effect of condition ($F(2,52) = 31.52$; $p < 0.001$; $np^2 = 0.548$). *Post-hoc* paired-samples t -tests controlled for multiple comparisons (Bonferroni corrected $p = 0.017$) revealed that RTs differed significantly between the conflict vs. no-conflict condition ($t(26) = 6.47$, $p < 0.001$), and between the conflict vs. no-hidden condition ($t(26) = 7.49$, $p < 0.001$) with conflict trials being longer than no-conflict and no-hidden trials. There was no significant difference between the no-conflict and the no-hidden conditions ($t(26) = 0.33$, $p = 0.743$).

ERP Results

The increased processing costs for the conflict condition as evidenced in the RT data were also reflected in the ERPs through modulations of a late positivity (see **Figure 3**). The cluster-based permutation analysis (Maris and Oostenveld, 2007) of the ERP data revealed two significant positive channel-time clusters for the comparison of the conflict vs. no-conflict condition over posterior brain areas 750–850 ms relative to noun onset (cluster t -statistic = 4031, $p = 0.022$), and over anterior and posterior brain areas 900–1250 ms relative to noun onset (cluster t -statistic = 31666, $p = 0.001$) (see **Figure 3**). The comparison of the conflict vs. no-hidden conditions (see **Figure 4**) showed a positive channel-time cluster over anterior brain regions only, 1100–1200 ms relative to noun onset (cluster t -statistic = 3624, $p = 0.018$). No significant channel-time clusters were revealed for the comparison of the no-conflict vs. no-hidden conditions (see **Figure 5**).

Experiment 2: Eye-Tracking Behavioral Findings

As for Experiment 1, the accuracy results of the eye-tracking cohort were high for all participants ($n = 27$) across all conditions ($M = 99.67 \pm 0.44\%$; conflict condition: $M = 99.60 \pm 1.14\%$; no-conflict: $M = 99.21 \pm 1.51\%$; no-hidden: $M = 99.87 \pm 0.69\%$ accuracy rate). The ANOVA with condition (three levels: conflict, no-conflict, and no-hidden) as within-subjects factor revealed no significant effect of condition ($F(2,52) = 1.97$; $p = 0.158$, $np^2 = 0.070$) with respect to the accuracy rates. Thus, even in the conflict condition, participants mastered the integration of CG.

Similarly, the RTs relative to the onset of the critical noun (e.g., “star”) replicated those of the EEG cohort, with the ANOVA resulting in a significant effect of condition ($F(2,52) = 27.38$; $p < 0.001$, $np^2 = 0.513$). Participants were on average 202.30 ms slower in the conflict condition ($M = 1570.20 \pm 497.65$ ms) compared to the means of the no-conflict ($M = 1351.58 \pm 526.68$ ms) and no-hidden conditions ($M = 1384.23 \pm 489.60$ ms). Paired-samples t -tests controlled for multiple comparisons (Bonferroni corrected $p = 0.017$) indicated that RTs differed significantly for the comparison of the conflict vs. no-conflict condition ($t(26) = 7.09$, $p < 0.001$), and for the conflict vs. no-hidden condition ($t(26) = 5.41$, $p < 0.001$) with RTs in the conflict condition being longer than in the other two conditions. There was no significant difference between the no-conflict and the no-hidden condition ($t(26) = -1.08$, $p = 0.29$).

Eye-Tracking Results

As neither the RTs, nor the accuracy, nor the ERP data show a difference between the conditions no-conflict and no-hidden, we restricted the analysis of the eye-tracking data to the comparison of the conflict vs. no-conflict condition.

Looks to target object

In our experimental setup, the identification of the correct referent (e.g., the small star) could happen only after the processing of the critical noun (1650 ms after the onset of the auditory request). In the noun time window (1650–2650 ms post

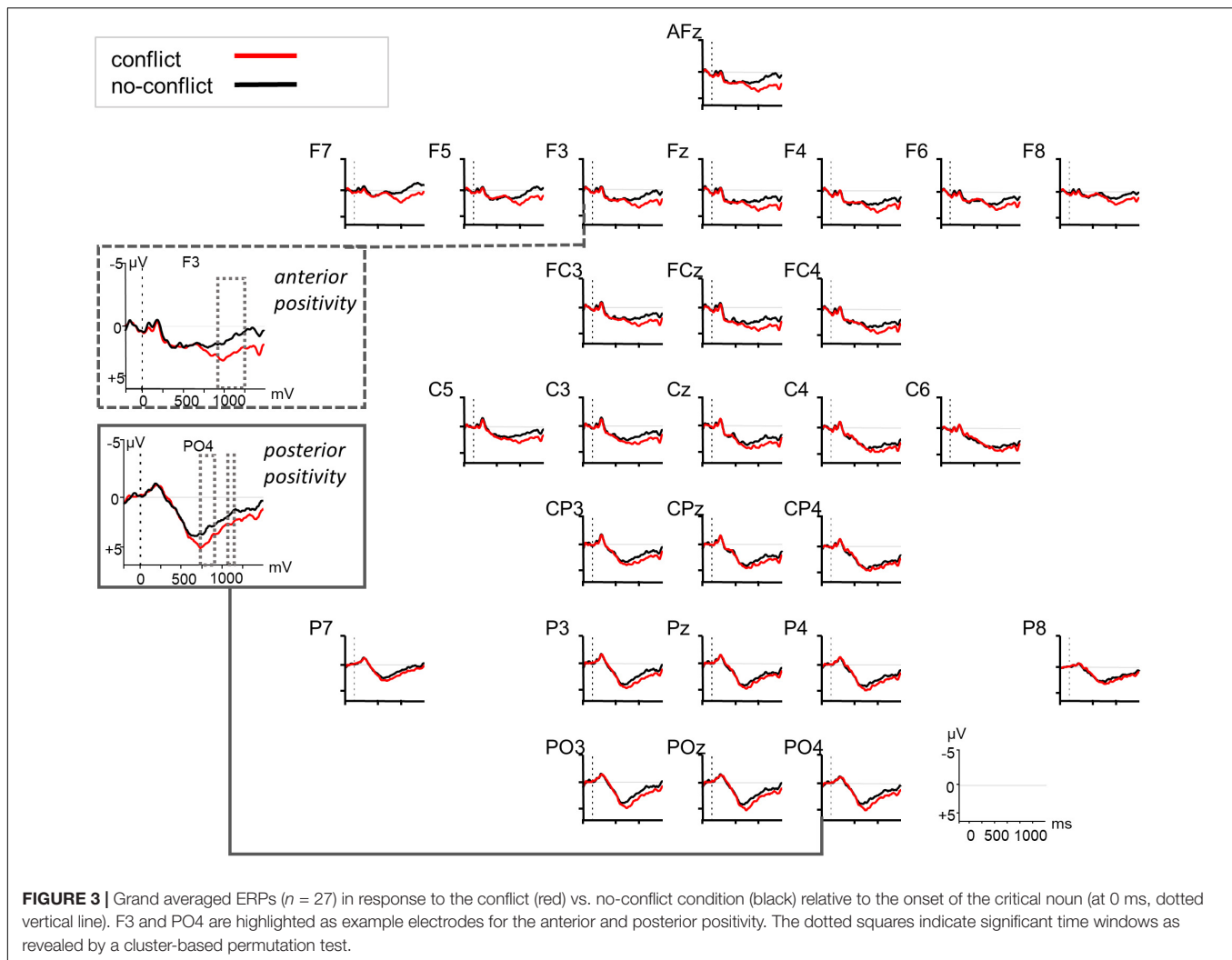


FIGURE 3 | Grand averaged ERPs ($n = 27$) in response to the conflict (red) vs. no-conflict condition (black) relative to the onset of the critical noun (at 0 ms, dotted vertical line). F3 and PO4 are highlighted as example electrodes for the anterior and posterior positivity. The dotted squares indicate significant time windows as revealed by a cluster-based permutation test.

auditory onset, i.e., 0–1000 ms post noun onset), the cluster-based permutation test revealed later and fewer looks to the target in the conflict condition compared to the no-conflict condition 1900–2400 ms after auditory onset (i.e., 250–750 ms after noun onset; cluster t -statistic: -32.37 , $p < 0.001$). In the post noun time window, 2650–3650 ms post auditory onset (1000–2000 ms post noun onset), the cluster-based permutation test revealed longer looks to the target in the conflict condition compared to the no-conflict condition 3100–3650 ms after auditory onset (i.e., 1450–2000 ms after noun onset; cluster t -statistic: 34.00 , $p = 0.002$). Thus, the integration of CG in the conflict condition delayed the looks to the target and let the participants look longer at it.

Looks to privileged object

Since looks to the target occurred later and were longer in the conflict condition, the most interesting question is, whether looks to the privileged object caused the longer latencies in the conflict condition. Cluster-based permutation analysis showed that participants indeed looked more often to the privileged object in the conflict condition as compared to the no-conflict

condition from 2100 to 2650 ms post auditory onset in the noun time window (i.e., 450–1000 ms post noun onset; cluster t -statistic: 60.01 , $p < 0.001$) and from 2650 to 3200 ms post auditory onset in the post noun time window (i.e., 1000–1550 ms post noun onset; cluster t -statistic: 41.10 , $p < 0.001$) (**Figure 6**). In addition, in the **Supplementary Materials** we compared the looks to the privileged object in the conflict-condition to the medium sized object in the no-conflict condition (see **Supplementary Figure 4**). Participants looked to the privileged object in the conflict condition more often than to the medium sized object in the no-conflict condition, but still the medium-sized object might be affected by “carry-over effects” such that it was looked at, although it never was the smallest or biggest object at display.

Anticipatory looks to CG objects within the two object triplets

Our data showed that participants anticipated objects in CG to be the referent even before the onset of the disambiguating noun (see **Figure 7**). The comparison using the cluster-based permutation revealed an anticipation of objects in CG in the adjective time window (cluster t -statistic: -50.56 , $p < 0.001$).

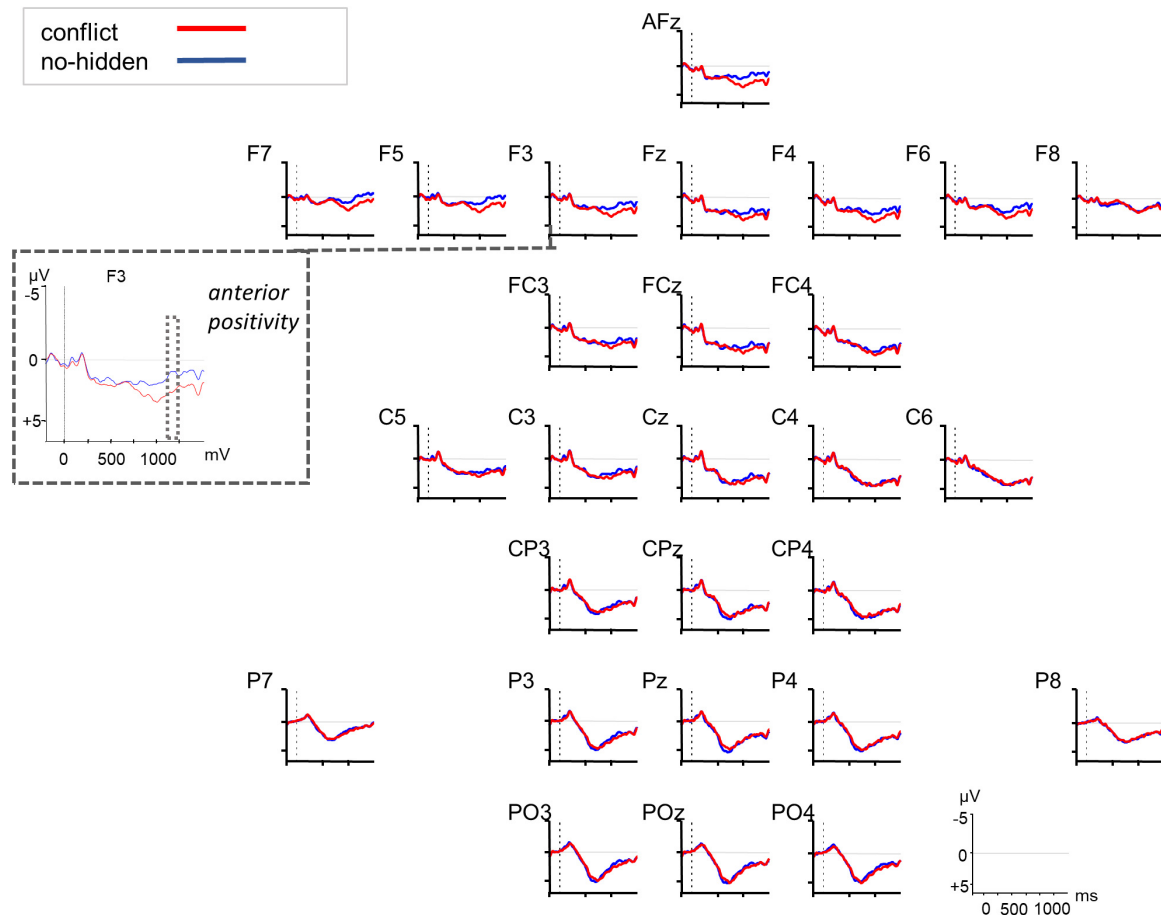


FIGURE 4 | Grand averaged ERPs ($n = 27$) in response to the conflict (red) vs. no-hidden condition (blue) relative to the onset of the critical noun (at 0 ms, dotted vertical line). F3 is highlighted as example electrode for the anterior positivity. The dotted squares indicate the significant time window as revealed by a cluster-based permutation test.

That is, 1000–1650 ms post auditory onset (i.e., right after adjective onset), participants were more likely to consider the small or big object in CG when encountering the adjective than the object in PG.

DISCUSSION

Our study aimed to verify different, partly conflicting accounts about when CG information is integrated during reference processing. Early integration accounts posit that CG information immediately constrains the domain in which utterances are typically processed (e.g., Nadig and Sedivy, 2002; Hanna et al., 2003; Brown-Schmidt et al., 2008; Heller et al., 2008). In contrast, late integration accounts suggest that CG information enter the parsing process at a later stage, either due to strategic egocentric processing strategies (*Egocentrism Account*, Keysar et al., 2000) or because lexical information is activated autonomously independent from perspective (*Autonomous Activation Account*, Barr, 2008). Recent accounts try to integrate both accounts and offer that perspective-taking is simultaneously affected by both

the egocentric and the CG perspective (*Simultaneous Integration Account*, Heller et al., 2016).

We here combined a computerized version of the well-established referential communication game (similar to Keysar et al., 2000) with behavioral, EEG (Experiment 1), and eye-tracking (Experiment 2) measures in order to better understand the temporal dynamics and the cognitive processes underlying the integration of CG information. We analyzed three conditions: conflict, in which there was a conflict between the participant's privileged information and CG information, and no-conflict and no-hidden, in which there was no such conflict. As neither the RTs, nor the accuracy, nor the ERP data show a difference between the conditions no-conflict and no-hidden, we can conclude that the mere presence of occluded slots in the grid did not lead to some unspecific computation of ground or did not affect general attentional processes. Therefore, our discussion will mainly focus on the comparison of the conflict vs. no-conflict condition. In the conflict condition, the object that fits the avatar's request best (e.g., the small star) was in PG. Thus, the consideration of CG information and the suppression of PG information were necessary in order to pick the correct

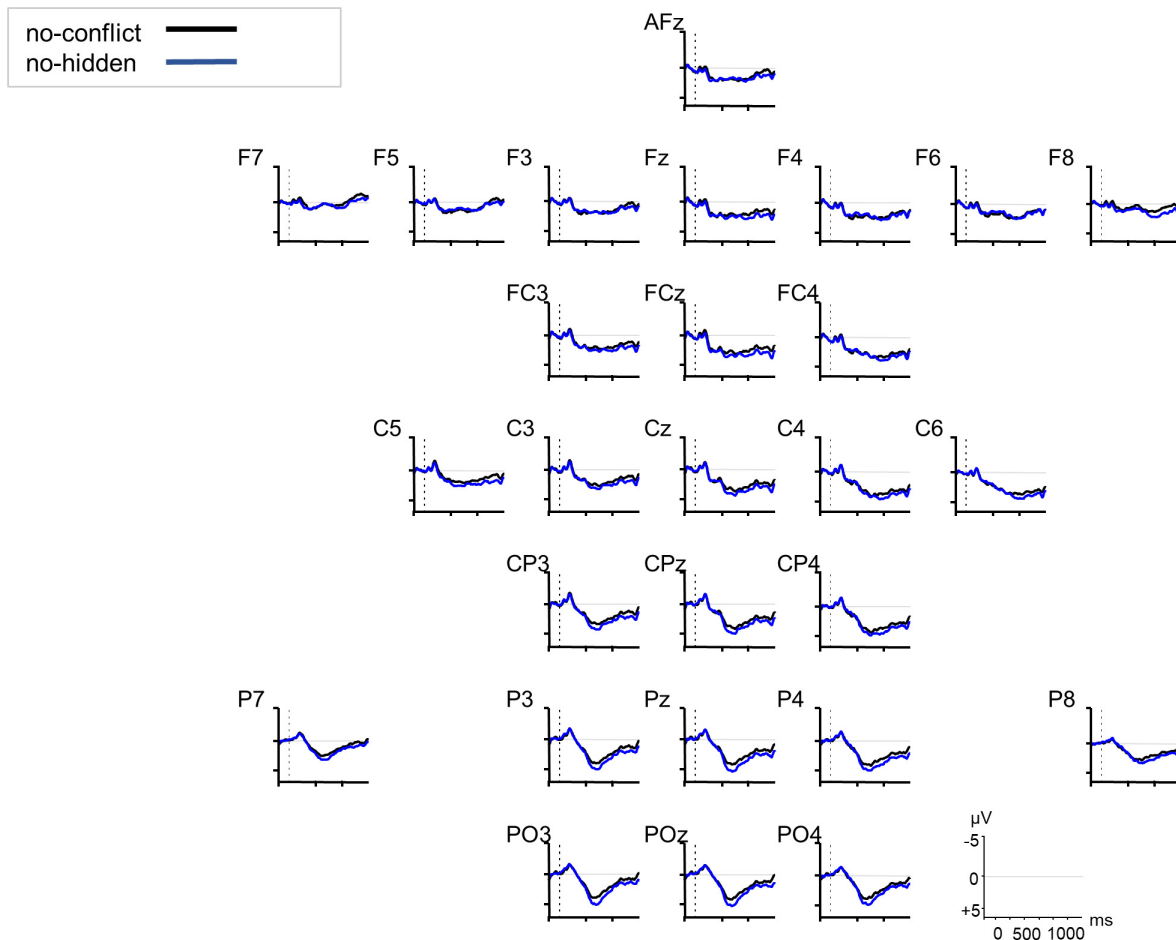


FIGURE 5 | Grand averaged ERPs ($n = 27$) in response to the no-conflict (black) vs. no-hidden condition (blue) relative to the onset of the critical noun (at 0 ms, dotted vertical line). No significant time window was revealed by the cluster-based permutation testing.

referent in CG. In contrast, in the no-conflict condition, the consideration of CG information was not necessary for reference resolution, because there was no conflicting information in PG to be resolved.

Comparing the conflict and the no-conflict conditions, RT data, ERP, and eye-tracking results all point to the notion that the integration of CG has its costs. This is the case even though participants initially anticipated objects in CG (the small/big item of the two object triplets, which was in CG) to be the referent until encountering the critical noun phrase, as revealed by the eye-tracking data. Then, 450–1550 ms after noun onset, eye-tracking also showed that participants were indeed distracted by privileged information in the conflict-condition: They considered the competitor although it was in PG. They also looked less to the target in the conflict vs. the no-conflict condition 250–750 ms post noun onset. Slightly later, that is 750–1250 ms post noun onset, ERPs revealed a late positivity when comparing the conflict and the no-conflict conditions. We propose that the late positivity resembled a P600-like response. Similarly to the increase in theta-band power, the P600 mirrors an increase in processing costs when

discourse representations have to be monitored, updated, and modified (e.g., van Herten et al., 2005; Bornkessel-Schlesewsky and Schlewsky, 2008; Schumacher, 2009; Burmester et al., 2014). This interpretation is also in line with the syntax-discourse model (SDM) (Burkhardt, 2005), and its extension, the multi-stream-model of discourse processing (e.g., Hung and Schumacher, 2012; Wang and Schumacher, 2013). Later on in the processing time line, in the post-noun-phase (1450–2000 ms post noun onset), eye-tracking revealed later or longer looks to the target in the conflict vs. the no-conflict condition. Finally, RTs of both the EEG and the eye-tracking cohort were significantly slower in the conflict (EEG: $M = 1469.07 \pm 364.17$ ms; eye-tracking: $M = 1570.20 \pm 497.65$ ms) in comparison to the no-conflict condition (EEG: $M = 1289.38 \pm 375.18$ ms; eye-tracking: $M = 1351.58 \pm 526.68$ ms)¹. Overall, the RT data confirmed

¹It is not quite clear why participants responded around 100 ms slower in the eye-tracking cohort than in the EEG cohort. One possibility might be that participants were more aware of being monitored by the eye-tracker than the EEG and hence showed a more careful responding behavior (click, drag-and-drop). In addition, as outlined above, in Experiment 2 we prevented that the target and the privileged object in the grid appeared in neighboring slots of the grid to minimize the

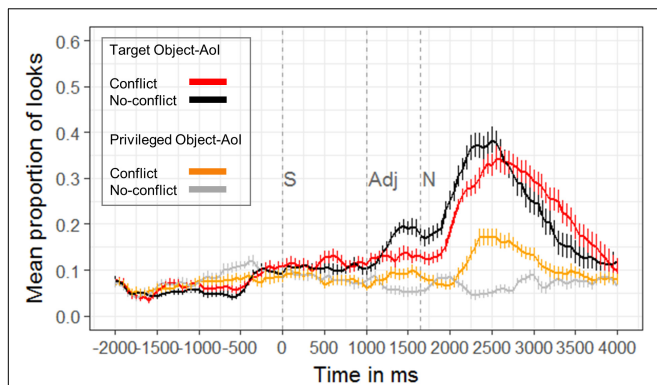


FIGURE 6 | Proportion of looks to the target (red, black) in CG and to the object in privileged ground (orange, gray) in the conflict and no-conflict conditions, respectively ($n = 27$). Trials are aligned to the onset of the sentence (S), e.g., “Move the small star to the top” at 0 ms. The onset of the sentence (S, at 0 ms), the onset of the adjective (Adj, 1000 ms post onset of the auditory request), and the onset of the noun (N, 1650 ms post onset of the auditory request) are marked by dashed vertical lines in the Figure. Cluster-based permutation analyses indicated statistically significant differences between the red and the black line 1900–2400 ms and 3100–3650 ms after auditory onset and between the orange and the gray line 2100–3200 ms after auditory onset. Error bars represent the standard error (SE). For alternative Figures see **Supplementary Materials**.

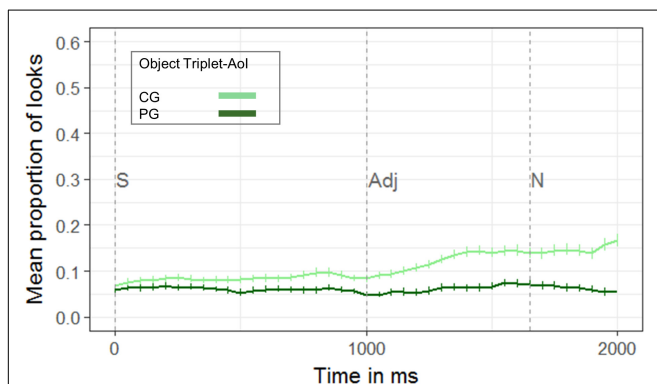


FIGURE 7 | Proportion of looks to the objects in common ground (CG, light green), and privileged ground (PG, dark green), of the two object triplets in the adjective time window 1000–1650 ms post sentence onset (S) ($n = 27$). The onset of the sentence (S, at 0 ms), the onset of the adjective (Adj, 1000 ms post onset of the auditory request) and the onset of the noun (N, 1650 ms post onset of the auditory request) are marked by vertical dashed lines in the Figure. Cluster-based permutations showed that participants were more likely to look at the objects in common ground before encountering the noun (i.e., 1000–1650 ms after auditory onset). Error bars represent the standard error (SE).

previous findings that RTs are longer when another person’s visual perspective is inconsistent with one’s own perspective than when it is consistent (e.g., Qureshi et al., 2010; Samson et al., 2010;

misclassification of looks. This also might account for slightly longer reaction times. However, the differences between the means of the conflict and the no-conflict conditions were similar in both groups (ERPs: 179.69 ms, eye-tracking: 218.62 ms), and the statistical analyses pointed to the same direction.

McCleery et al., 2011; Wang et al., 2019) and are thus also in line with late integration accounts.

Only the accuracy results did not reveal any differences for the conflict vs. no-conflict condition. Accuracy was high for both conditions in the EEG and the eye-tracking cohort. Even in the conflict condition, participants rarely made any errors (EEG: $M = 0.12 \pm 0.42\%$; eye-tracking: $M = 0.40 \pm 1.14\%$). Only 15% of those few errors were egocentric errors. These arose in the conflict condition when participants were not able to suppress privileged information when choosing a referent. Since the accuracy data does not give rise for a strong tendency to use an egocentric strategy, the *Egocentrism Account* (Keysar et al., 2000) cannot explain our data. Given that our design was most similar to that of Keysar et al. (2000), we suppose that the clear two-step instruction, the practice phase with corrective feedback (Wang et al., 2016), and the high repetition rate eliminated possible egocentric errors.

Our eye-tracking data therefore suggest that listeners initially start with the expectation that the speaker refers to an object that is shared (i.e., is in CG). This is in line with accounts that assume early effects of CG (Nadig and Sedivy, 2002; Hanna et al., 2003; Heller et al., 2008) or early attempts to take a speaker’s perspective in anticipation of a referring expression (Barr, 2008). However, with the presentation of the noun, they also consider the mentioned objects, even if they are in PG. This information seems to interfere with the earlier tendency to consider objects in CG to be the referent. The interference of privileged information happened 450–1550 ms after noun onset. There was an increase in looks to the competitor in the conflict-condition, while looks to the target were reduced (250–750 ms after noun onset). Finally, longer looks to the target were registered in the conflict vs. no-conflict condition in the later post-noun phase (1450–2000 ms after noun onset). This interference of the privileged competitor makes the late integration of CG information an effortful process, since the current discourse model has to be updated and modified (e.g., Hung and Schumacher, 2012).

Our ERP data support this view as they point to processing differences between conditions as well. Late positivities were identified that differentiated the neuronal responses to the conflict and the no-conflict condition. First, around 750–850 ms after the onset of the critical noun we found a positivity that had a posterior distribution. Second, after around 1000 ms a positivity was observed that had a more anterior distribution. These late positivities were taken as indication for increased processing costs due to the updating of discourse representations and conflict resolution (Burkhardt, 2007; Schumacher, 2009). Especially the earlier positivity may reflect a P600-like response (for a review, see Swaab et al., 2012), since it occurred over posterior brain regions, was not lateralized, and the positive slow deflection started around 500 ms post noun onset (but became significant in the statistical analysis only after 750 ms). Although modulations of the P600 were initially attributed to syntactic anomalies or ambiguities, its functional interpretation has been extended considerably in the last decades (e.g., Brouwer et al., 2012, 2017). The P600 seems to be evoked when some kind of information needs to be integrated into the unfolding interpretation of the sentence or a reanalysis has to be undertaken

due to inconsistent streams of semantic, morphosyntactic, and pragmatic information (e.g., Kuperberg et al., 2007). The same holds for executive or cognitive control mechanisms in error monitoring or information-reprocessing due to response uncertainties during language comprehension (e.g., van Herten et al., 2006). The SDM, that was first introduced for pronominal-antecedent relations by Burkhardt (2005), and extended to general discourse processing in a multi-stream-model (e.g., Hung and Schumacher, 2012; Wang and Schumacher, 2013), interprets late positivities as being induced by discourse updating and discourse modification. Finally, *P600-as-P3-accounts* (e.g., Sassenhagen et al., 2014; Sassenhagen and Fiebach, 2019) question the language-specificity of the P600 but rather see them as a (domain-general) component indexing the linkage of saliency and response selection. Our data do not allow to disentangle the underlying mechanisms, however, our results of Experiment 1 indicate that participants integrate CG information relatively late and in an effortful manner. RTs were slower for the conflict than for the no-conflict and no-hidden conditions. We interpret our ERP data as pointing to increased processing demands when CG needs to be considered (conflict condition) compared to when there is no conflict between CG and privileged information (no-conflict and no-hidden condition). As the RT data indicate that the conflict-resolution had its cost, we consider the positivity as indexing this increased processing cost. This interpretation would go in line with a domain-general view on the ERP positivities as in the *P600-as-P3-account*, in which the positivity is considered to index behaviorally relevant saliency (Sassenhagen et al., 2014).

Taken all findings together, we found very little egocentric errors, or other differences in response accuracy. Yet RTs, EEG, and eye-tracking data showed that CG integration seems to be an effortful, long-lasting, and rather late process. We saw evidence for an early anticipation of CG objects, but privileged information could not be fully neglected, resulting in a late integration of CG information during the parsing process. Therefore, our data can be explained by one of the late integration accounts, namely the *Autonomous Activation Account* of Barr (2008). This account suggests early anticipation without (or with late) integration of CG information. The *Simultaneous Integration Account* of Heller et al. (2016) posits that both egocentric and CG behavior are simultaneously active during perspective taking. Crucially, the specific design or context triggers, which behavior-egocentric or CG-dominates the task performance. As a result, the integration of CG information varies from task to task and from design to design. In our study, the privileged object best matched the referring expression, as it was the case in the Keysar et al. (2000) study. Heller et al. (2008) argue that in such cases, the “goodness of fit” to the speaker’s referring expression strongly enforces attention to the privileged object. This makes the interference with CG information more likely than in previous designs that support early integration accounts (for a thorough discussion please refer to Heller et al., 2008). Our findings thus do not speak against the *Simultaneous Integration Account* of Heller et al. (2016) but support the conclusions drawn from their work.

As outlined in the introduction, Sikos et al. (2019) used EEG to study perspective-taking using a somewhat different design

than we did and interpreted their data in the framework of early integration, constrained-based accounts. Their sentences were locally ambiguous [e.g., “Click on the brontosaurus with the boots” with two brontosauri in the display, both in CG (Common Ground Competitor condition) or one in CG and one in PG (Privileged Ground Competitor condition)]. The disambiguating noun (e.g., “boots”) was always presented at the end of the sentence in both competitor conditions. In addition, there was a No-Competitor Control condition with just one brontosaurus in CG and a perceptual control condition. The PG Competitor condition in Sikos et al. (2019) is temporally (i.e., until the onset of the disambiguating noun) similar to our conflict condition. The other conditions can’t be directly “translated” to our design: both the No-Competitor Control condition and the CG Competitor condition are similar to our no-conflict condition; in the No-Competitor condition the utterance is already disambiguated at the first noun (“brontosaurus”) while in the CG Competitor condition, the disambiguation comes later (“with the boots”). Similar to our study, Sikos et al. (2019) find increased RTs in both competitor conditions (PG Competitor, CG Competitor) as we find them in our conflict condition. However, in the ERP data they find a late, widely distributed, negativity for the CG Competitor condition. The authors interpret this as an Nref component (van Berkum et al., 1999, 2003), indicating that only the CG competitor is considered as a potential referential candidate, but not the competitor in PG. Nieuwland et al. (2007) take the frontal negative shift as reflecting genuine referential ambiguity in the current model of the discourse in a deeper sense, which is related to referential accessibility. Looking at the ERP patterns in Figure 3 of Sikos et al. (2019), (page 281) it looks as if the PG Competitor condition (i.e., the conflict condition in our terminology) leads to a greater positivity than the CG Competitor condition. In our data we interpret this as a late positivity for the conflict condition indexing updating costs. However, in Sikos et al. (2019) this difference seems not to be statistically significant and not different from the No-Competitor condition. If we apply the Nref-argument to our data and interpret our data as an Nref component, the no-conflict condition would show the Nref effect compared to the conflict condition (i.e., more negative in no-conflict than in conflict trials). However, this seems unlikely, in our opinion, as in our no-conflict condition there is clearly only ONE potential referent (e.g., the small star) which is in CG. This would contradict the assumption that the Nref indexes the accessibility of MULTIPLE potential referents as in Sikos et al. (2019). This interpretation would only apply, if the medium-sized star is not considered as “medium” but as a second small or second big star at the display leading to referential ambiguity. We cannot completely exclude this possibility. An additional analysis of looks to the medium-sized object reveals, that participants also look at this object (see **Supplementary Material**). This indicates that participants’ perspective regarding what constitutes a “small” or “big” object might be shifted, and that considering the medium-sized object as a good referent for “small” or “big” is carried over to conditions where perspective-taking is not necessary (i.e., the no-conflict condition). But, firstly, looks to the target by far exceed looks to the medium-sized object. Secondly, the supplementary analysis reveals that participants looked more

often to the privileged object (i.e., the competitor that is either the small or big object) in the conflict condition as compared to medium-sized object in CG in the no-conflict condition. Thirdly, accuracy shows that participants almost always choose the correct object, which is the smallest or biggest star at the display in the no-conflict and no-hidden condition. This indicates that our participants indeed considered the medium-sized star as medium, otherwise the potential referential ambiguity would not have been disambiguated and a referential choice could not have been made. Overall, while our study design differs from Sikos et al. (2019), we find similar behavioral effects and ERP-patterns which, superficially, look similar but which are interpreted differently. Sikos et al. (2019) conclude that the competitor in PG is not considered a potential referential candidate and that the RT effects just reflect attentional distraction effects. The latter effects are, however, not mirrored in their ERP data as Sikos et al. found no difference between the PG Competitor and the No-Competitor condition. Accuracy rates were very high in all conditions in Sikos et al. (2019) which the authors interpret as indicating that competitors in PG are not considered a potential candidate. However, if this would be generalized to all kinds of perspective taking tasks, “egocentric errors,” as have been shown in Keysar et al. (2000), or the finding of interference from information in PG (e.g., Barr, 2008) would be hard to explain. Also, if we consider our own eye-tracking data, the assumption, that PG information is not considered as a potential referent, becomes implausible: The eye-tracking data show, that the longer RTs in the conflict condition in comparison to the no-conflict condition can be attributed to two interfering effects. The first interfering effect resulted from a strong anticipation of an object in CG to be the referent: when encountering the adjective, participants showed higher proportions of looks to the object in CG compared to the object in PG in both object triplets. Then, the second interfering effect arose. When encountering the noun, participants shifted their attention to the object of comparison in PG in the conflict condition (i.e., they shifted their attention to the “egocentric competitor”): 450–1550 ms after noun onset, there was an increase in looks to the object of comparison in PG in the conflict condition (i.e., to the competitor). Only later, 1450–2000 ms after noun onset, looks to the competitor decreased while looks to the target (in CG) were reaching their peak.

To conclude, our data speak in favor of the *Autonomous Activation Account* (Barr, 2008), since our eye-tracking data reveal an early anticipation (in the adjective time window) but a late, effortful integration of CG information (in the noun time window). The re-evaluation or integration seems to be a late and effortful process reflected by increased processing costs (RTs), later and longer looks to the target, and late positive and slow brain responses. However, the data can also be aligned to the *Simultaneous Integration Account* of Heller et al. (2016), since, overall, listeners restrict their referential domain to information in CG when appropriate, but the information in PG has the potential to interfere. The *Simultaneous Integration Account* elegantly combines the contrary findings of egocentric vs. CG behavior and early vs. late integration of CG that can be found in the literature. Further, it highlights the circumstances of such performance differences. Yet, how fast CG information affects reference processing seems to depend on a variety of

factors such as the current communicative and experimental setting, the familiarity to the interlocutor (e.g., Münster and Knoeferle, 2017), the complexity of task demands, or just the readiness or motivation to take another person’s perspective. The establishment of a model of socially situated language processing, which incorporates all these factors, should be further addressed in the future.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found on the Open Science Framework, <https://osf.io/vdsxu/>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the University of Potsdam. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MR was responsible for designing and implementing the experimental paradigms and for collecting and analyzing the data and wrote the manuscript. MP was involved in data collection and data analysis and was involved in writing the manuscript. IW and BH designed the study and contributed to writing and revising the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by “XPrag.de: New Pragmatic Theories based on Experimental Evidence” (SPP 1727) funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), project number 254858842.

ACKNOWLEDGMENTS

We are very grateful to Tom Fritzsche who supported the analysis of the eye-tracking data. We further thank Lu Zhang and Choonkyu Lee for helpful discussions of our study design and data. We also thank Jan Ries for programming and Robin Schäfer, Franziska Machens, Helene Killmer-Rumpf, and Susanne Pelke for help with recruiting participants, running the experiments, and data analysis. We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (German Research Foundation) and the Open Access Publishing Fund of the University of Potsdam.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.565651/full#supplementary-material>

REFERENCES

- Altmann, G., and Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition* 30, 191–238. doi: 10.1016/0010-0277(88)90020-0
- American Clinical Neurophysiology Society (2006). Guideline 5: guidelines for standard electrode position nomenclature. *J. Clin. Neurophysiol.* 23, 107–110. doi: 10.1097/00004691-200604000-00006
- Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: listeners anticipate but do not integrate common ground. *Cognition* 109, 18–40. doi: 10.1016/j.cognition.2008.07.005
- Barr, D. J. (2016). “Visual world studies of conversational perspective taking,” in *Visually Situated Language Comprehension*, eds P. Knoeferle, P. Pykkönen-Klauck, and M. W. Crocker (Amsterdam: John Benjamins Publishing Company), 261–290. doi: 10.1075/aicr.93.10bar
- Barr, D. J., Jackson, L., and Phillips, I. (2014). Using a voice to put a name to a face: the psycholinguistics of proper name comprehension. *J. Exp. Psychol.* 143, 404–413. doi: 10.1037/a0031813
- Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Res. Rev.* 59, 55–73. doi: 10.1016/j.brainresrev.2008.05.003
- Bornkessel-Schlesewsky, I., and Schumacher, P. B. (2016). “Towards a neurobiology of information structure,” in *The Oxford Handbook of Information Structure*, eds C. Féry and S. Ishihara (Oxford: Oxford University Press).
- Brouwer, H., Crocker, M. W., Venhuizen, N. J., and Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cogn. Sci.* 41(Suppl. 6), 1318–1352. doi: 10.1111/cogs.12461
- Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Res.* 1446, 127–143. doi: 10.1016/j.brainres.2012.01.055
- Brown-Schmidt, S. (2012). Beyond common and privileged. Gradient representations of common ground in real-time language use. *Lang. Cogn. Process.* 27, 62–89. doi: 10.1080/01690965.2010.543363
- Brown-Schmidt, S., Gunlogson, C., and Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition* 107, 1122–1134. doi: 10.1016/j.cognition.2007.11.005
- Burkhardt, P. (2005). *The Syntax-Discourse Interface: Representing and Interpreting Dependency*. Amsterdam: John Benjamins Publishing Company.
- Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport* 18, 1851–1854. doi: 10.1097/WNR.0b013e3282f1a999
- Burmester, J., Spalek, K., and Wartenburger, I. (2014). Context updating during sentence comprehension: the effect of aboutness topic. *Brain Lang.* 137, 62–76. doi: 10.1016/j.bandl.2014.08.001
- Cane, J. E., Ferguson, H. J., and Apperly, I. A. (2017). Using perspective to resolve reference: the impact of cognitive load and motivation. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 591–610. doi: 10.1037/xlm0000345
- Clark, H. H., Schreuder, R., and Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *J. Verb. Learn. Verb. Behav.* 22, 245–258. doi: 10.1016/S0022-5371(83)90189-5
- Crocker, M. W., Knoeferle, P., and Mayberry, M. R. (2010). Situated sentence processing: the coordinated interplay account and a neurobehavioral model. *Brain Lang.* 112, 189–201. doi: 10.1016/j.bandl.2009.03.004
- Davies, C., and Katsos, N. (2010). Over-informative children. Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua* 120, 1956–1972. doi: 10.1016/j.lingua.2010.02.005
- Deutsch, W., and Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition* 11, 159–184. doi: 10.1016/0010-0277(82)90024-5
- Dink, J. W., and Ferguson, B. (2015). *EyetrackingR: An R Library for Eye-tracking Data Analysis*. Available online at: <http://www.eyetrackingr.com> (accessed January 11, 2017).
- Engelhardt, P. E., Bailey, K. G., and Ferreira, F. (2006). Do speakers and listeners observe the Gricean maxim of quantity? *J. Mem. Lang.* 54, 554–573. doi: 10.1016/j.jml.2005.12.009
- Epley, N., Keysar, B., van Boven, L., and Gilovich, T. (2004a). Perspective taking as egocentric anchoring and adjustment. *J. Pers. Soc. Psychol.* 87, 327–339. doi: 10.1037/0022-3514.87.3.327
- Epley, N., Morewedge, C. K., and Keysar, B. (2004b). Perspective taking in children and adults. Equivalent egocentrism but differential correction. *J. Exp. Soc. Psychol.* 40, 760–768. doi: 10.1016/j.jesp.2004.02.002
- Ferguson, H. J., Apperly, I., Ahmad, J., Bindemann, M., and Cane, J. E. (2015). Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task. *Cognition* 139, 50–70. doi: 10.1016/j.cognition.2015.02.010
- Ferguson, H. J., and Breheny, R. (2012). Listeners’ eyes reveal spontaneous sensitivity to others’ perspectives. *J. Exp. Soc. Psychol.* 48, 257–263. doi: 10.1016/j.jesp.2011.08.007
- Ferreira, F., and Clifton, C. (1986). The independence of syntactic processing. *J. Mem. Lang.* 25, 348–368. doi: 10.1016/0749-596X(86)90006-9
- Glucksberg, S., Krauss, R. M., and Higgins, E. T. (1975). “The development of referential communication skills,” in *Review of Child Development Research*, ed. F. E. Horowitz (Chicago: University of Chicago Press), 305–345.
- Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484. doi: 10.1016/0013-4694(83)90135-9
- Grice, H. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Hanna, J. E., and Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task. Evidence from eye movements. *Cogn. Sci.* 28, 105–115. doi: 10.1207/s15516709cog2801_5
- Hanna, J. E., Tanenhaus, M. K., and Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *J. Mem. Lang.* 49, 43–61. doi: 10.1016/S0749-596X(03)00022-6
- Harris, T., Wexler, K., and Holcomb, P. (2000). An ERP investigation of binding and coreference. *Brain Lang.* 75, 313–346. doi: 10.1006/brln.2000.2318
- Heller, D., Grodner, D., and Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition* 108, 831–836. doi: 10.1016/j.cognition.2008.04.008
- Heller, D., Parisien, C., and Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition* 149, 104–120. doi: 10.1016/j.cognition.2015.12.008
- Hoeks, J. C., and Brouwer, H. (2014). “Electrophysiological research on conversation and discourse processing,” in *The Oxford Handbook of Language and Social Psychology*, ed. T. M. Holtgraves (Oxford: Oxford University Press).
- Holzen, K. V., and Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *J. Exp. Child Psychol.* 113, 569–586. doi: 10.1016/j.jecp.2012.08.001
- Hung, Y.-C., and Schumacher, P. B. (2012). Topicality matters: position-specific demands on Chinese discourse processing. *Neurosci. Lett.* 511, 59–64. doi: 10.1016/j.neulet.2012.01.013
- Keysar, B., Barr, D. J., Balin, J. A., and Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychol. Sci.* 11, 32–38. doi: 10.1111/1467-9280.00211
- Knoeferle, P., and Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cogn. Sci.* 30, 481–529. doi: 10.1207/s15516709cog0000_65
- Knoeferle, P., and Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: evidence from eye movements. *J. Mem. Lang.* 57, 519–543. doi: 10.1016/j.jml.2007.01.003
- Krauss, R. M., and Glucksberg, S. (1977). Social and nonsocial speech. *Sci. Am.* 236, 100–105. doi: 10.1038/scientificamerican0277-100
- Kronmüller, E., Noveck, I., Rivera, N., Jaume-Guazzini, F., and Barr, D. (2017). The positive side of a negative reference: the delay between linguistic processing and common ground. *R. Soc. Open Sci.* 4:160827. doi: 10.1098/rsos.160827
- Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., and Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: evidence from event-related potentials. *Brain Lang.* 100, 223–237. doi: 10.1016/j.bandl.2005.12.006
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

- McCleery, J. P., Surtees, A. D. R., Graham, K. A., Richards, J. E., and Apperly, I. A. (2011). The neural and cognitive time course of theory of mind. *J. Neurosci.* 31, 12849–12854. doi: 10.1523/JNEUROSCI.1392-11.2011
- Morisseau, T., Davies, C., and Matthews, D. (2013). How do 3- and 5-year-olds respond to under- and over-informative utterances? *J. Pragmat.* 59, 26–39. doi: 10.1016/j.pragma.2013.03.007
- Münster, K., and Knoeferle, P. (2017). Extending situated language comprehension (accounts) with speaker and comprehender characteristics: toward socially situated interpretation. *Front. Psychol.* 8:2267. doi: 10.3389/fpsyg.2017.02267
- Nadig, A. S., and Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychol. Sci.* 13, 329–336. doi: 10.1111/j.0956-7976.2002.00460.x
- Nieuwland, M. S., Otten, M., and van Berkum, J. J. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *J. Cogn. Neurosci.* 19, 228–236. doi: 10.1162/jocn.2007.19.2.228
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869.
- Osterhout, L., and Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773. doi: 10.1006/jmla.1995.1033
- Qureshi, A. W., Apperly, I. A., and Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: evidence from a dual-task study of adults. *Cognition* 117, 230–236. doi: 10.1016/j.cognition.2010.08.003
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rodríguez Ronderos, C., Münster, K., Guerra, E., Kreysa, H., Rodríguez, A., Kröger, J., et al. (2018). Eye tracking during visually situated language comprehension: flexibility and limitations in uncovering visual context effects. *J. Vis. Exp.* e57694. doi: 10.3791/57694
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., and Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1255–1266. doi: 10.1037/a0018729
- Sassenhagen, J., and Fiebach, C. J. (2019). Finding the P3 in the P600: decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage* 200, 425–436. doi: 10.1016/j.neuroimage.2019.06.048
- Sassenhagen, J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain Lang.* 137, 29–39. doi: 10.1016/j.bandl.2014.07.010
- Schumacher, P. B. (2009). “Definiteness marking shows late effects during discourse processing: evidence from ERPs,” in *DAARC 2009: LNAI 5847*, eds S. L. Devi, A. Branco, and R. Mitkov (Berlin: Springer-Verlag), 91–106. doi: 10.1007/978-3-642-04975-0_8
- Schumacher, P. B. (2018). “Semantic-pragmatic processing,” in *The Handbook of Psycholinguistics*, 1st Edn, eds E. M. Fernández and H. S. Cairns (New York, NY: John Wiley & Sons), 392–410. doi: 10.1002/9781118829516.ch17
- Schumacher, P. B., Brandt, P., and Weiland-Breckle, H. (2018). *The Semantics of Gradability, Vagueness, and Scale Structure: Experimental Perspectives*. Cham: Springer International Publishing.
- Sikos, L., Tomlinson, S. B., Heins, C., and Grodner, D. J. (2019). What do you know? ERP evidence for immediate use of common ground during online reference resolution. *Cognition* 182, 275–285. doi: 10.1016/j.cognition.2018.10.013
- Snedeker, J., and Trueswell, J. (2003). Using prosody to avoid ambiguity. Effects of speaker awareness and referential context. *J. Mem. Lang.* 48, 103–130. doi: 10.1016/S0749-596X(02)00519-3
- Snedeker, J., and Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *J. Mem. Lang.* 58, 574–608. doi: 10.1016/j.jml.2007.08.001
- Spivey-Knowlton, M., and Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition* 55, 227–267. doi: 10.1016/0010-0277(94)00647-4
- Swaab, T. Y., Ledoux, K., Camblin, C. C., and Boudewyn, M. A. (2012). “Language-related ERP components,” in *Oxford Handbook of Event-Related Potential Components*, eds E. S. Kappenman and S. J. Luck (Oxford: Oxford University Press), 397–440.
- Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition* 73, 89–134. doi: 10.1016/S0010-0277(99)00032-3
- van Berkum, J., Koornneef, A. W., Otten, M., and Nieuwland, M. S. (2007). Establishing reference in language comprehension: an electrophysiological perspective. *Brain Res.* 1146, 158–171. doi: 10.1016/j.brainres.2006.06.091
- van Berkum, J. J., Brown, C. M., and Hagoort, P. (1999). Early referential context effects in sentence processing. evidence from event-related brain potentials. *J. Mem. Lang.* 41, 147–182. doi: 10.1006/jmla.1999.2641
- van Berkum, J. J., Zwitserlood, P., Hagoort, P., and Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cogn. Brain Res.* 17, 701–718. doi: 10.1016/S0926-6410(03)00196-4
- van Herten, M., Chwilla, D. J., and Kolk, H. H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *J. Cogn. Neurosci.* 18, 1181–1197. doi: 10.1162/jocn.2006.18.7.1181
- van Herten, M., Kolk, H. H. J., and Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Brain Res. Cogn. Brain Res.* 22, 241–255. doi: 10.1016/j.cogbrainres.2004.09.002
- Wang, J. J., Ali, M., Frisson, S., and Apperly, I. A. (2016). Language complexity modulates 8- and 10-year-olds' success at using their theory of mind abilities in a communication task. *J. Exp. Child Psychol.* 149, 62–71. doi: 10.1016/j.jecp.2015.09.006
- Wang, J. J., Tseng, P., Juan, C.-H., Frisson, S., and Apperly, I. A. (2019). Perspective-taking across cultures: shared biases in Taiwanese and British adults. *R. Soc. Open Sci.* 6:190540. doi: 10.1098/rsos.190540
- Wang, L., and Schumacher, P. B. (2013). New is not always costly: evidence from online processing of topic and contrast in Japanese. *Front. Psychol.* 4:363. doi: 10.3389/fpsyg.2013.00363
- Wu, S., and Keysar, B. (2007). The effect of culture on perspective taking. *Psychol. Sci.* 18, 600–606. doi: 10.1111/j.1467-9280.2007.01946.x

Conflict of Interest: 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 © 2020 Richter, Paul, Höhle and Wartenburger. 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.



Morphological Metaphor Mapping of Moral Concepts in Chinese Culture

Yingjie Liu¹, Kang Li¹, Lina Li^{1*}, Jing Zhang¹, Yuerui Lin¹, Baxter DiFabrizio² and He Wang^{1*}

¹ School of Psychology, North China University of Science and Technology, Tangshan, China, ² College of William and Mary, Williamsburg, VA, United States

OPEN ACCESS

Edited by:

Marcela Pena,
Pontificia Universidad Católica
de Chile, Chile

Reviewed by:

Luis Faisca,
University of Algarve, Portugal
Akira Utsumi,
The University
of Electro-Communications, Japan

*Correspondence:

Lina Li
llina3725582@163.com
He Wang
wanghe429@163.com

Specialty section:

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

Received: 21 April 2020

Accepted: 02 November 2020

Published: 17 December 2020

Citation:

Liu Y, Li K, Li L, Zhang J, Lin Y,
DiFabrizio B and Wang H (2020)
Morphological Metaphor Mapping
of Moral Concepts in Chinese Culture.
Front. Psychol. 11:554061.
doi: 10.3389/fpsyg.2020.554061

According to conceptual metaphor theory, individuals are thought to understand or express abstract concepts by using referents in the physical world—right and left for moral and immoral, for example. In this research, we used a modified Stroop paradigm to explore how abstract moral concepts are metaphorically translated onto physical referents in Chinese culture using the Chinese language. We presented Chinese characters related to moral and immoral abstract concepts in either non-distorted or distorted positions (Study 1) or rotated to the right or to the left (Study 2). When we asked participants to identify the Chinese characters, they more quickly and accurately identified morally positive characters if they were oriented upright or turned to the right and more quickly and accurately identified immoral characters when the characters were distorted or rotated left. These results support the idea that physical cues are used in metaphorically encoding social abstractions and moral norms and provided cross-cultural validation for conceptual metaphor theory, which would predict our results.

Keywords: moral concept, shape, rotation, Chinese culture, conceptual metaphor theory, morphological metaphor

INTRODUCTION

Morality is a reflection of fundamental judgments about good and bad, right and wrong. It plays an indispensable role in enhancing human well-being by accurately assessing the social desirability of a behavior or belief (Haidt and Algor, 2004). The moral concept is a very important abstract concept in social life, and its understanding is also realized through a relatively concrete concept. Conceptual metaphor theory (CMT) proposes that humans tend to rely on metaphors as an efficient way to ground abstract concepts in a physical, relatable reality (Lakoff and Johnson, 1999, 2008).

Metaphors connect abstract concepts (which CMT calls the “target domain”) to more concrete concepts (the “origin domain”) to make abstract concepts more understandable (Lakoff, 2008). In CMT, an origin domain refers to the cognitive domain consisting of specific events we are familiar with or that can be directly experienced. A target domain refers to the cognitive domain consisting of abstract concepts hard to understand or perceive, which often rely on the vocabulary and imagery of the origin domain to be expressed (Lakoff and Johnson, 2008).

Metaphor is a powerful cognitive tool for understanding abstract moral concepts (Landau et al., 2010). Lakoff and Johnson (1999) proposed that that humans organize and construct conceptual systems through a few specific basic concepts, including a set of spatial relationships (such as up and down, left and right), a set of physical ontological concepts (such as

entities, containers), and a set of behaviors (such as eating and walking). All of these basic concepts are derived from human sensorimotor experience. Through cross-domain mapping, we construct and understand abstract concepts via projections and consecutive applications of basic concrete concepts. Thus, the conceptual metaphors of morality and immorality are clarified in terms of some common contrastive categories from our bodily experience in the physical environment. For example, in English, “high-minded” or “on the up and up” is used to describe a moral person, and “down and dirty,” “low-minded,” or “underhanded” is used to describe an immoral person (Lakoff and Johnson, 1999). In addition to this spatial position metaphor, morality is also represented in metaphors by concrete concepts, such as brightness, colorfulness (Hill and Lapsley, 2009; Sherman and Clore, 2009), and cleanness (Haidt and Joseph, 2004; Zhong and Liljenquist, 2006; Schnall et al., 2008). In one study, when the word “God” (morality) was presented in the center of the screen, the attention of the individual would be diverted to the above and right spatial field of vision, and the display of “demon” related words (immorality) shifted the individual’s visual attention to the below and left spatial field of vision, presumably in accordance with the learned association of right and left (“sinister” is the Latin word for “left”) (Chasteen et al., 2010). To sum up, these studies involve color, light spectrum contrasts, and directional and spatial metaphors to help describe morality and its social concomitants. However, the origin domain does not only include the dimensions as mentioned earlier but also includes shape, location, and other orienting properties. For instance, the English expression “an upright position” has been used to describe a person with strong morals and seems to come from the perspective of “morphology.” The morphology refers to the appearance or image of objects in existence or the form of expression under certain conditions. Morphological metaphors are different from common spatial metaphors; it has to do with its own properties. According to the morphological characteristics of objects, we can divide it into skewness, rotation, stand upside down, orientation, or its own size. Further, morphological metaphor can be extended from the perspectives of space metaphor, location metaphor, color metaphor, etc., and expand the research scope of moral metaphor.

Research on the metaphorical characteristics of moral cognition has focused mainly on the English language (Camgöz et al., 2002; Meier et al., 2004). However, the cultural background also plays an important role in the study of moral concepts and their metaphors. Culture influences not only the personal embodied metaphor (Gibbs and Berg, 1999) but also the metaphorical representation of thinking and language (Xinya and Zhongyi, 2015). It follows that language may actually dictate how moral metaphors are encoded and how concepts from the origin domains are configured to express the target domains—inherently creating differences in moral concepts across languages and cultures. The Chinese perspective on the metaphor of moral imagination deserves special attention, as it may indeed have markedly different representations of moral concepts because of its linguistic distinctiveness. Both from the structure of Chinese characters and its extended meaning, it has special

morphological metaphor characteristics. In Chinese, “正义” (zhèng yì, “justice”), “正直” (zhèng zhí, upright), “正大光明” (zhèng dà guāng míng, “aboveboard”), and “改邪归正” (gǎi xié guī zhèng, “on the straight”) are often used to describe a person’s moral integrity; all these words encompass the character “正” (zhèng), which means “standard, not deviating, not bending, properly proportioned.” We also use “歪主意” (wāi zhu yì, bad ideas), “歪门邪道” (wāi mén xié dào, crooked ways), and “歪风邪气” (wāi fēng xié qì, evil winds) to describe a person’s bad moral character; all these words contain the character “歪” (wāi), which means “skew, distortion.” In English, people often use “straight” and “upright” to describe honest and reliable, whereas using “devious” and “oblique,” which mean crooked, to describe something that is unjust or immoral—for example, “achieve the goal by means of a devious path” or “an oblique political maneuver.” Chinese also regard immoral behaviors as the distortion of moral concepts; hence, uprightness and distortion correspond, respectively, to morality and immorality.

Does our cognitive architecture effectively connect metaphorical representations of moral concepts with morphology? If the CMT model is an accurate representation of how we articulate and encode abstractions, we would expect that the particular property of objects traditionally associated with a certain concept (for example, “straight,” “up,” “right,” “clear”) would cue faster recognition of those concepts (“morality,” “honesty,” “valor,” etc.). We would also expect the CMT to predict that the particular property of objects traditionally associated with concepts with lower social desirability or negative valence (“deceit,” “cruelty,” “betrayal,” “rage”) would also be more readily recognizable in their stereotypical characteristics (“crooked,” “left,” “down,” “distorted”). We would explore whether the morphological characteristics of objects or words can also be associated with positive moral or negative, immoral concepts.

Orientation is a vital task in the human cognitive system, and many abstract concepts rely on spatial metaphors to be construed clearly (Lakoff and Johnson, 2008). There are three diametrically opposed, generally asymmetric mental axes in the body: up and down, left and right, and front and back (Tversky, 2008). The left and right directional indicators have been mapped concurrently onto metaphorical representations of moral concepts. Chinese traditional culture often refers to “以右为尊” (yǐ yòu wéi zūn, “take right side as honor”). The host should take the initiative to stay on the left side and make the right side free for guests, honoring their presence. Men take the initiative to make their right side available for women when standing together, out of respect. Juniors should give the right-of-way and their right-side standing position up to their elders. Many cultural phenomena further indicate that horizontal spatial orientation has been related to moral concepts (“right” referring to moral rectitude and “left” referring to immoral or aberrant behavior). These perceptual cues or orientation of the objects can also be redundantly mapped onto each other: right can also refer to up, and left refers to down.

In Chinese, we often use the word “无出其右” (wú chū qí yòu, “second to none”) to refer to one’s moral integrity, whereas “旁门左道” (páng mén zuǒ dào, “heterodoxy”) is commonly used

to refer to impure or improper methods or ways. Even today, the words such as “座右铭” (zuò yòu míng, motto) and “意见相左” (yì jiàn xiāng zuǒ, “difference of opinion”) still indicate that the right side is honored. The Chinese characters “左” (zuǒ) and “右” (yòu) mean left and right, respectively. Not only in the background of Chinese culture but also in the expression of English, “right” often has a positive connotation, such as intelligence (the Latin word for right-handed was “dexter,” from which the word dexterous comes from). For example, “a right-hand man,” “to right a wrong.” On the other hand, “left” often express a negative meaning, such as clumsiness, socially suboptimal, or socially undesirable conduct. For example, “two left feet,” “out in left field,” and one from English, a “left-handed compliment.” Chasteen et al. (2010) showed that when people process the words related to their god or holy items, they responded faster to the stimulus on the right than left. Contrarily, when people processed the words related to the devil, they responded faster to the stimulus on the left than right. It is plausible to infer that left and right could be perceptual, orientational signals that cue the moral concepts of right and wrong, especially as they, like a number line, instruct a sort of directionality and imply a visual contrast. Left and right might be metaphorical representations of moral concepts in Chinese culture, but then again, they might not.

In the present research, we used an experimental paradigm borrowing the basic principle of the Stroop paradigm (Stroop, 1938). We manipulated the congruence words’ meanings and their visual display and used response time to measure facilitation or interference effects on word semantic classification (moral vs. immoral). We applied this experimental paradigm to test the possible facilitation of the proposed morphological metaphor mapping for moral concepts present in Chinese culture. Thus, we tested whether moral concepts are more readily pairable with “upright” characters and immoral with distorted characters in Study 1. In Study 2, we checked to see if the left–right association also mapped onto immoral and moral concepts, respectively. Detecting meaningful associations along these lines would lend credibility to the morphological metaphor mapping concept as a psychologically generalized reality.

STUDY 1 THE MORPHOLOGICAL METAPHOR OF MORAL CONCEPT: FROM THE UPRIGHT AND SKEWED PERSPECTIVE

The aim of study 1 was to investigate whether moral words are easier to read when clear and immoral words are comparatively easier to read when distorted, compared with the inverse (moral words distorted vs. immoral words non-distorted). We hypothesized participants would more quickly recognize non-distorted moral words and distorted immoral words (congruent condition) and that they would take more time to parse non-distorted immoral words and distorted moral words (incongruent condition).

MATERIALS AND METHODS

Participants

Thirty-four college students voluntarily participated in study 1 [14 males, 20 females; their mean age was 18.26 years, and standard deviation (SD) was 0.89]. All were right-handed. Participants were recruited through advertisements posted in school social spaces. Participants contacted us and volunteered to participate in the experiment. All participants had normal vision or corrected vision. At the end of the experiment, participants were paid according to their winnings and debriefed.

Stimulus Construction and Evaluation Methodology

Moral and Immoral Words

To create stimuli, we took 20 two-character moral words from the Chinese edition of modern Chinese frequency dictionary, such as honesty, nobility, purity, etc., and 20 two-character words describing immoral states of immoral concepts, such as betrayal, dishonor, dirtiness, ridicule, punishment, etc., (see **Table 1**). For example, the word frequency of 诚实 (honest) was 0.00076 (the 0.00076 represents the word occurred 0.00076 times in a million). There was no significant difference in the words’ frequency of usage between moral words ($M = 0.00086$) and immoral words ($M = 0.00102$), $t(38) = -0.87$, $p = 0.390$. Word frequency information was obtained from “The Corpus System of Modern Chinese Research” of Beijing Language and Culture University¹. There was no significant difference in the number of strokes

¹<http://corpus.zhonghuayuwen.org/index.aspx>

TABLE 1 | List of moral words and immoral words.

Moral words	Word frequency	Immoral words	Word frequency
廉洁 (White-handed)	0.00023	混乱 (Chaos)	0.00205
恩惠 (Grace)	0.00023	恶劣 (Scurviness)	0.00198
厚道 (Honest and kind)	0.00023	欺骗 (Deception)	0.00183
刚毅 (Courageous)	0.00038	腐朽 (Rotten)	0.00167
洁净 (Spotless)	0.00038	可耻 (Shame)	0.00122
坚毅 (Fortitudinous)	0.00046	惩罚 (Punishment)	0.00114
纯真 (Trueness)	0.00046	可恶 (Detestable)	0.00114
坚贞 (Faithful)	0.00046	功利 (Utilitarian)	0.00107
恭敬 (Respectful)	0.00053	腐败 (Corruption)	0.00107
朴实 (Plain)	0.00053	狼狈 (Discomfiture)	0.00099
谦逊 (Modest and moderate)	0.00068	肮脏 (Sordid)	0.00053
诚实 (Honest)	0.00076	毒害 (Poison)	0.00099
公平 (Fair)	0.00091	狂妄 (Conceit)	0.00084
慷慨 (Generous)	0.00099	践踏 (Trample)	0.00068
淳朴 (Artlessness)	0.00053	嘲讽 (Sarcasm)	0.00061
谦虚 (Modest)	0.00129	粗暴 (Rude)	0.00061
高尚 (Noble)	0.00167	堕落 (Degenerate)	0.00053
援助 (Aid)	0.00190	讥讽 (Ridicule)	0.00053
纯洁 (Pure)	0.00221	敲诈 (Blackmail)	0.00053
感激 (Grateful)	0.00237	丑陋 (Ugliness)	0.00046

making up the characters of our moral words ($M = 17.74$) and immoral words ($M = 18.50$), $t(38) = -0.71$, $p = 0.51$.

Before the formal experiment, we assessed the adequacy (the degree of validity and recognition) of the moral and immoral words we chose during a pretest (Yang et al., 2017; Zhai et al., 2018). We asked a different set of 30 participants, who were not part of the formal experiment, to rate the moral valence of these words on a nine-point scale, from 1 for “highly immoral” to 9 for “highly moral.” We found that the mean score for moral words ($M \pm SD = 7.33 \pm 0.45$) was significantly higher than 5, the middle point of the nine-point morality rating scale, $t(19) = 23.18$, $p < 0.001$. The mean score of immoral words ($M \pm SD = 3.225 \pm 0.71$) was significantly lower than 5, $t(19) = -11.164$, $p < 0.001$. The chosen moral and immoral words aligned with participants’ experience of them and met the requirements of the experiment—specifically, our moral and immoral language samples discretely and reliably represented their intended corpora.

Distortions of Verbal Stimuli

We randomly selected five moral and five immoral words from the experimental materials we had chosen and used Adobe Photoshop’s built-in distortion function to *morph* each character, twisting each character into 45° and 90° distortions to both their left and right sides—as if the center of the character was a whirlpool’s epicenter, bending the characters around its nexus. We then assessed how our stimuli would feel in terms of their degree of distortion and ease of recognition with a separate pretest. Another 30 participants, neither part of the formal experiment nor the prior pretest, rated the degree of distortion on a seven-point scale from 1 for “no distortion at all” to 7 for “very distorted.” They also rated the relative ease of recognition on a seven-point scale from 1 for “it was hard to identify” to 7 for “it was easy to identify.” Statistical analysis using a one-way analysis of variance (ANOVA) showed that the main effect of the degree of distortion was significant, and the mean score of 90° distortion to the left ($M \pm SD = 5.47 \pm 0.73$) was significantly higher than other distortions, $F(3) = 102.02$, $p < 0.001$, $\eta^2 = 0.73$. The difference in the degree of recognition was not significant, $F(3) = 0.10$, $p = 0.964$; words were generally as recognizable across the 45° and 90° distortions. Therefore, we chose to distort the words 90° to the left for our formal experimental materials.

Experimental Design and Procedure

We used 2 (word type: morality vs. immorality) \times 2 (word shape: distortion vs. no distortion) within-subjects design. The dependent variables were participants’ (1) accuracy of identifying the word’s moral category and (2) reaction time (RT). We based our procedure and implementation of the experiment on that of prior research (Yu et al., 2016; Yang et al., 2017).

We used the Stroop inhibitory strategy to probe the effects of deforming the words on the time it took participants to recognize them (Stroop, 1938; Federica and Tagini, 2017). Target words in the experimental task were presented in the center of the computer screen with the background set as 50% grayscale (set the red green blue values of 50% grayscale as red = 128, green = 128, and blue = 128). All words were in Song typeface,

with a font size of 48 pounds, and then every single word was processed into an image of 550 \times 300 pixels. The distorted words in the experiment were generated by the “Distortion” function in Adobe Photoshop. Every distorted word was distorted 90° to the left. No distorted words were presented in a familiar or standard form. Unaltered words were presented simply in the Song typeface without further processing. Each of the 20 moral words and the 20 immoral words was presented in both “distortion” and “no distortion” states only once, for a total of 80 trials. Stimuli were presented in completely random order.

At the beginning of the experiment, participants were asked to sit in front of the computer approximately 30 cm away from the screen and place their left index finger on the key “F” of the keyboard and their right index finger on the key “J.” They were told that during the experiment, they would have to make accurate and rapid judgments about the inherent characteristics of the words—“Does the word you see describe a Moral or Immoral concept?”

The experimental procedure consisted of two parts: the practice and the experimental trials. Before the formal experiment, non-experimental stimuli (10 moral words and 10 immoral words) were used to help participants practice the task—and the computer would automatically give feedback (right or wrong category) about their responses. Participants completed the 20 practice trials; after the practice, they saw a screen reading, “Press any key to start the formal experiment.” During the formal experiment, participants pressed buttons according to the instructions, and the target words were presented randomly—this time, without feedback as to whether they sorted the word into the correct category or not. Participants were instructed to press “F” for immoral words and press “J” for moral words. The key assignments were counterbalanced between participants, and the computer measured participants’ RT and accuracy for each trial.

Before each stimulus was presented, a red “+” fixation point was displayed in the center of the computer screen for 800 ms. Then the target word was presented, and participants were asked to make corresponding keystrokes according to the category (morality or immorality) the word belonged to. For example, 诚实 (chéng shí, meaning “honesty”) corresponds to morality. A blank screen was presented for 500 ms during the intertrial interval (see Figure 1).

RESULTS

The accuracy rate of all participants in Study 1 was above 97%. We excluded trials with RT shorter than 300 ms and deleted 3 SDs below and above the mean (approximately 2% of trials) and the inaccurate trials. A repeated-measures ANOVA using a 2 (word type: morality vs. immorality) \times 2 (word shape: distortion vs. no distortion) design was conducted on RT. A significant main effect of word type was revealed [$F(1, 33) = 4.81$, $p = 0.035$, $\eta^2 = 0.13$]. Specifically, the RT for morality was faster than immorality (see Table 2). The main effect of word shape was significant [$F(1, 33) = 7.20$, $p = 0.011$, $\eta^2 = 0.18$]. The RT for distortion was slower than no distortion (see Table 2). Importantly, a significant

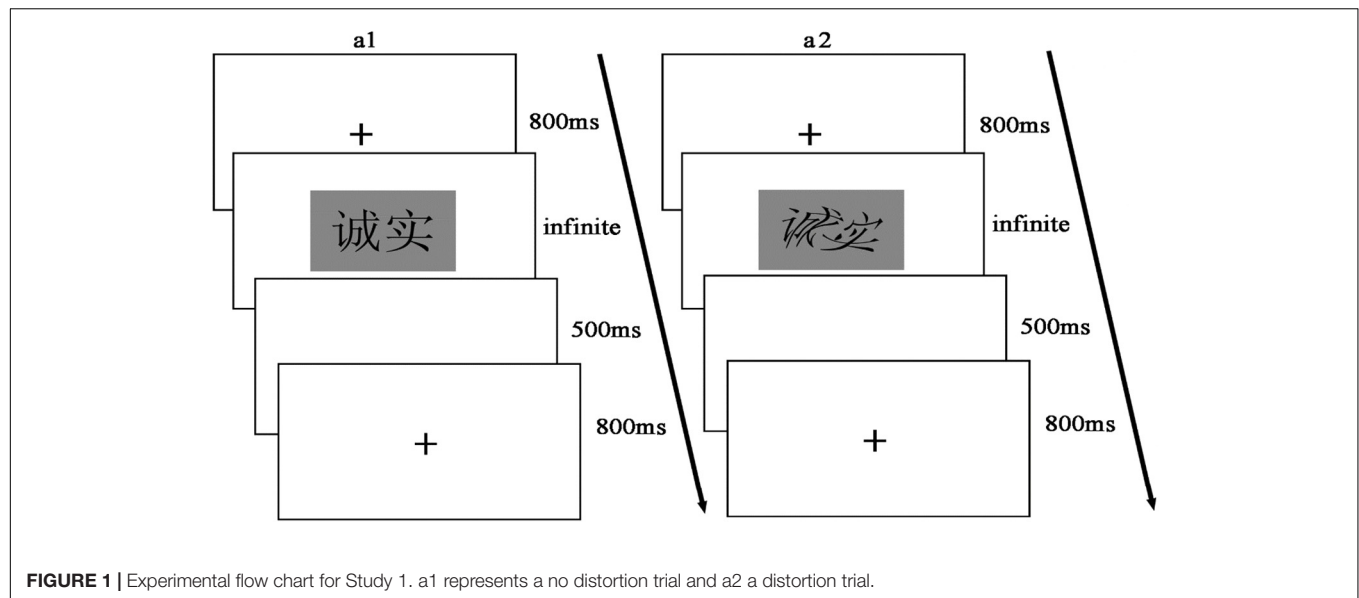


TABLE 2 | Means and standard deviations of RT and accuracy.

		RT(ms)($M \pm SD$)	Accuracy($M \pm SD$)
Word type	Morality	718.69 \pm 113.21	0.99 \pm 0.02
	Immorality	752.65 \pm 124.95	0.96 \pm 0.04
Word shape	Distortion	746.69 \pm 118.59	0.98 \pm 0.03
	No distortion	718.76 \pm 115.91	0.97 \pm 0.04

interaction between word type and word shape was observed [$F(1, 33) = 9.89, p = 0.004, \eta_p^2 = 0.23$]. *Post hoc* comparisons revealed that the RT for morality concepts under distortion ($M \pm SD = 747.24 \pm 117.83$) was slower than it under no distortion ($M \pm SD = 690.13 \pm 118.66$), $p < 0.001$; the RT showed no difference between immorality concepts under distortion ($M \pm SD = 746.13 \pm 133.82$) and immorality under no distortion ($M \pm SD = 747.39 \pm 132.63$), $p = 0.937$. Furthermore, we also found that the RT showed no difference between morality under distortion and immorality under distortion, $p = 0.940$; but the RT for immorality under No distortion was slower than morality under no distortion, $p = 0.002$ (see **Figure 2**).

A repeated-measures ANOVA using a 2 (word type: morality vs. immorality) \times 2 (word shape: distortion vs. no distortion) design was conducted on accuracy. The main effect of word type was significant [$F(1, 33) = 17.92, p < 0.001, \eta_p^2 = 0.35$], the accuracy for morality was higher than immorality (see **Table 2**). The main effect of word shape was not significant [$F(1, 33) = 0.42, p = 0.523$]. A significant interaction between word type and word shape was observed [$F(1, 33) = 7.92, p = 0.008, \eta_p^2 = 0.19$]. *Post hoc* comparisons revealed that the accuracy for morality under distortion ($M \pm SD = 0.98 \pm 0.23$) was lower than it under no distortion ($M \pm SD = 0.99 \pm 0.12$), $p = 0.044$; the accuracy for immorality under distortion ($M \pm SD = 0.97 \pm 0.17$) was higher than it under no distortion ($M \pm SD = 0.95 \pm 0.41$), $p = 0.050$ (see **Figure 2**).

DISCUSSION

Using a modified Stroop paradigm, we asked participants to make judgments about the underlying social quality (morality or immorality) of words that we presented either in a non-distorted or a distorted state. Results indicate that participants' average RT for morality words was significantly faster than immorality words. Participants' average RT for distortion was slower than that of no distortion ("Upright"). Furthermore, their RT for morality under distortion was slower than it was under no distortion. Average RT did not differ between immorality words under distortion and it under no distortion.

In terms of accuracy of identification, morality was higher than immorality, and morality under distortion was lower than it under no distortion. Accurately identifying immorality words was more likely in the distortion condition than in the no distortion condition. These findings support the conclusion that word shape and presentation (non-distortion vs. distortion) may have metaphorical representations consistent with abstract moral concepts. To further scrutinize this phenomenon, in Study 2, we explored whether Chinese moral concepts also evinced the same sort of differential facilitation and hindrance effects along the orientations of right vs. left.

STUDY 2 THE MORPHOLOGICAL METAPHOR OF MORAL CONCEPT: FROM THE PERSPECTIVE OF LEFT ROTATION OR RIGHT ROTATION

Study 2 meant to investigate whether the morphological metaphor mapping of moral concept might help us predict the cognitive behavior of participants seeing moral and immoral words presented rotated toward the right or the left, seemingly triggering an association with right and wrong. We hypothesized

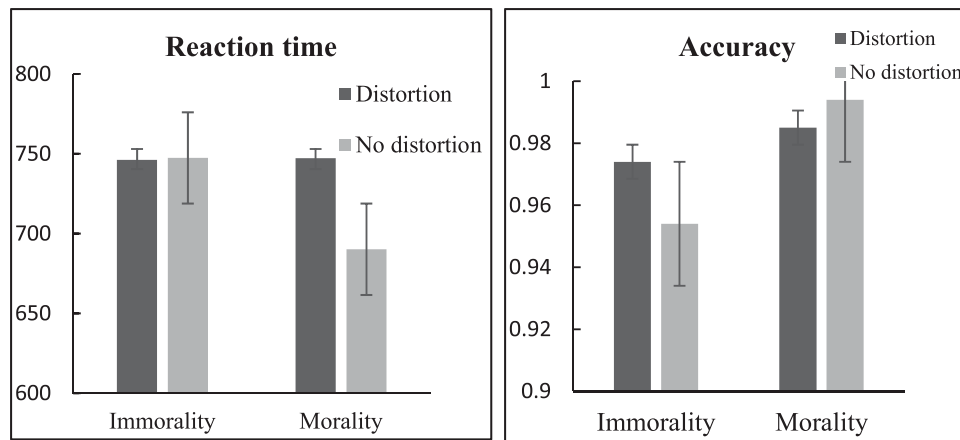


FIGURE 2 | RT and accuracy of morality vs. immorality judgments under word shape conditions (distorted vs. non-distorted). Error bars = SEM.

participants would have a faster reaction time to moral words that were rotated to the right and immoral words that were rotated to the left (“congruent”) and a slower reaction time to immoral words that were rotated to the right and moral words that were rotated to the left (“incongruent”).

MATERIALS AND METHODS

Participants

Thirty-eight college students voluntarily participated in study 2 (16 males, 22 females; their mean age was 18.30 years, and SD was 0.86); all were right-handed. Participants were recruited in the same way as Study 1. At the end of the experiment, participants were paid according to their winnings and debriefed.

Experimental Material and Evaluation Method

The corpus of moral and immoral words for Study 2 came from Study 1. However, in Study 2, each experimental word was rotated either 45° to the left or to the right in Adobe Photoshop. Prior research shows that mental rotation with an angle of less than 60° can affect word recognition (Koriat and Norman, 1985). Some researchers had also studied the recognition of Chinese characters and found that the mental rotation of 45° had an impact on the recognition of Chinese characters (Bolin et al., 1995), so we used 45 degrees of rotation for our experimental material.

Each of the 20 moral and immoral words was presented in the states of both 45° rotation to the left and 45° rotation to the right. Every single word was processed into an image of 550 by 350 pixels to ensure that the font could be fully displayed on the screen after the rotation. Each word is presented twice (once tilted to the right, once tilted to the left). There were a total of 80 trials, and the presentation order was completely random.

Experimental Design and Procedure

We used a within-subjects experimental design of 2 (word type: morality vs. immorality) × 2 (word rotation: left rotation vs. right

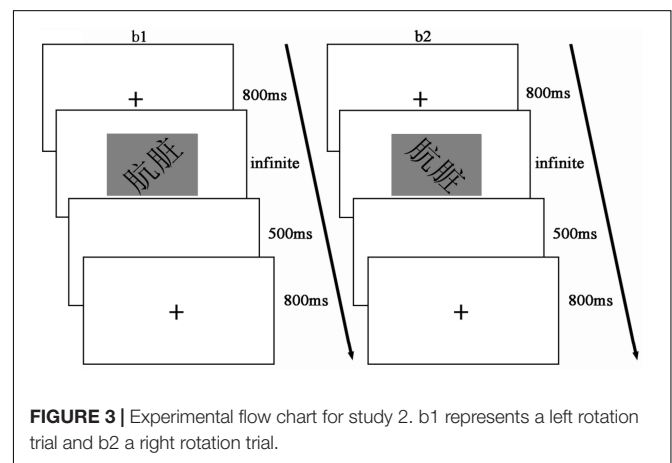


FIGURE 3 | Experimental flow chart for study 2. b1 represents a left rotation trial and b2 a right rotation trial.

rotation) within-subjects design. The dependent variables were the accuracy and RT.

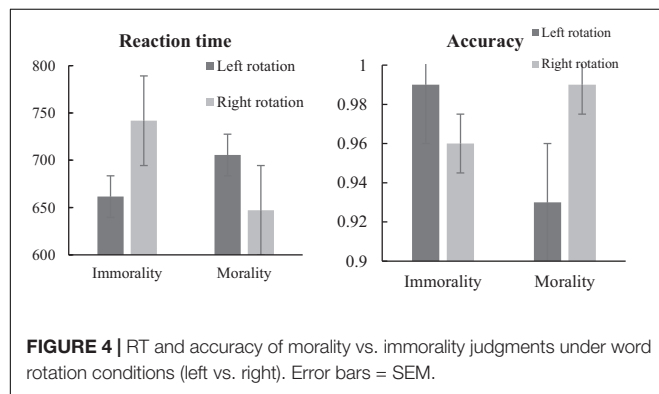
The experimental procedure in Study 2 was essentially that of Study 1, except that the independent variable of “word shape” in Study 1 now became “word rotation” in Study 2 (see **Figure 3**). Although we physically altered the shape of words in Study 1, in Study 2, the words themselves were unchanged but merely rotated 45°. The participants were presented with two completely different forms of vision (see **Figure 3**). We referred to previous empirical studies that recommended the rotation factor of 45° (Ścigała and Indurkha, 2016; Wang et al., 2016; Li and Cao, 2017; Zhai et al., 2018; Schneider et al., 2020). For example, 肮脏 (ang zang, means dirty) corresponding to immorality.

RESULTS

The accuracy rate of all participants of Study 1 was above 95%. We excluded trials with RT shorter than 300 ms and deleted 3 SDs below and above the mean (approximately 3% of trials) and the inaccurate trials. A repeated-measures ANOVA using a 2 (word type: morality vs. immorality) × 2 (word rotation:

TABLE 3 | Means and standard deviations of RT and accuracy.

		RT(ms)($M \pm SD$)	Accuracy($M \pm SD$)
Word type	Morality	676.31 \pm 55.01	0.960 \pm 0.09
	Immorality	701.66 \pm 79.85	0.980 \pm 0.02
Word rotation	Left rotation	683.55 \pm 69.94	0.963 \pm 0.09
	Right rotation	694.42 \pm 67.87	0.976 \pm 0.02



left rotation vs. right rotation) design was conducted on RT. A significant main effect of word type was revealed [$F(1, 37) = 4.50, p = 0.041, \eta_p^2 = 0.11$]; the RT for morality was faster than immorality (see **Table 3**). The main effect of word rotation was not significant [$F(1, 37) = 0.79, p = 0.377$]. However, a significant interaction between word type and word rotation was observed [$F(1, 37) = 31.26, p < 0.001, \eta_p^2 = 0.46$]. *Post hoc* comparisons revealed that the RT for morality under left rotation ($M \pm SD = 705.56 \pm 84.20$) was slower than it under right rotation ($M \pm SD = 647.06 \pm 66.58$), $p = 0.001$; the RT for immorality under left rotation ($M \pm SD = 661.55 \pm 87.72$) was faster than it under right rotation ($M \pm SD = 741.78 \pm 105.13$), $p < 0.001$ (see **Figure 4**).

Similar to the analysis of RT, to scrutinize accuracy of participants' judgments across conditions, we conducted a repeated-measures ANOVA. The main effect of word type was not revealed [$F(1, 37) = 2.11, p = 0.155$] (see **Table 2**). The main effect of word rotation was not significant [$F(1, 37) = 0.83, p = 0.368$]. However, a significant interaction between word type and word rotation was observed [$F(1, 37) = 6.64, p = 0.014, \eta_p^2 = 0.15$]. *Post hoc* comparisons revealed that the accuracy for morality under left rotation ($M \pm SD = 0.93 \pm 0.19$) was lower than it under right rotation ($M \pm SD = 0.99 \pm 0.02$), $p = 0.041$; the accuracy for immorality under left rotation ($M \pm SD = 0.99 \pm 0.01$) was higher than it under right rotation ($M \pm SD = 0.96 \pm 0.04$), $p < 0.001$ (see **Figure 4**).

DISCUSSION

Study 2 used a modification of the Stroop paradigm to explore the accessibility of moral and immoral concepts when presented in the right or left rotations (considered the metaphorically "congruent" condition) vs. left and right rotations (considered "incongruent"). RT for morality was faster than immorality.

The interaction effect showed that the RT for morality words was faster under the right rotation than when under left rotation; immorality words under left rotation were identified faster than when under the right rotation. For accuracy, only the interaction effect was significant. Participants classified morality concepts less accurately under the left rotation than it under the right rotation. They identified immoral concepts under left rotation was more accurately than under the right rotation.

Word rotation, such as left–right rotation, may have metaphorical representations consistent with abstract moral concepts, supporting the domain-generalization of CMT. We also found the scientific evidence for the concept of “以右为尊” (take right side as honor) in traditional Chinese culture.

GENERAL DISCUSSION

In the current research, we used a variation on the Stroop paradigm to conduct two studies investigating the morphological metaphor mapping of moral concepts against the backdrop of Chinese cultural expression. Specifically, we looked at moral and immoral concept vocabulary that were either presented with a non-distortion or a distortion or rotated either right or left to see if the CMT's theoretical framework would accurately predict human cognitive behavior when identifying these words. Participants identified words that were distorted or on the left faster if they were about immoral concepts, and they identified words describing moral concepts more readily if the words were non-distorted and rotated to the right. Moral concepts were easier to classify than immoral concepts and generally were identified faster. In “incongruent” conditions, identifying the moral and immoral words took longer (moral \times distorted/left or immoral \times non-distorted/right), even though the words were basically controlled for by shape and number of brushstrokes.

Accuracy of identifying the concepts also paralleled the congruent or incongruent condition—it was easier for participants to identify words if they were in a metaphorically stereotyped posture or shape. These results suggest that word shape and word rotation had metaphorical representations that were consistent with abstract moral concepts. These results further confirmed the morphological metaphor of moral concepts had psychological reality.

In the study, there is an interaction between word type and word shape, i.e., the RT for morality words was faster under the right rotation than when under the left rotation; immorality words under the left rotation were identified faster than when under the right rotation. The traditional Chinese etiquette usually takes the right as the top, for respect, and the left as the bottom for inferiority. The right usually represents high status, positive and positive, whereas the left represents low status, negative and derogatory. Studies have also shown that when there is a phenomenon, the right side is good, and the left side is bad in Western culture. Positive valence concepts such as intelligence and nobility are usually associated with the right of the space, whereas

negative valence concepts such as clumsiness and inferiority are associated with the left of the space (Jewell and McCourt, 2000). Also, in popular belief, morality is always just, and immorality is always oblique. In our other study, it was still found the RT for morality under distortion was slower than it was under no distortion. The result also further supported the difference in the association between moral and immoral words with different morphology. Language is encoded in the body of knowledge and information (Louwerse, 2008). The formation of conceptual meaning is not determined by a single factor but involves a variety of coded types of information, some with perceptual characteristics (e.g., embodied, modal) and some with non-perceptual characteristics (e.g., verbal symbols, modeless) (Dove, 2010). In the perception of moral words, people often accept that morality is positive and that distorted morality does not correspond to reality.

According to the theory of experiential cognition and perceptual simulation, perceptual representation is automatically activated when the vocabulary is processed (Barsalou, 2011). In our studies, people identified moral words faster than immoral ones. Some theorize that identifying moral, socially desirable behavior is primal and “direct” and hence inherently faster (Chen et al., 2018). Moral words could be the metaphorical prototypes of concepts that are deformed to create words for immoral concepts. On the other hand, Hill and Lapsley (2009) posit that individuals pay more attention to *immoral* events to protect themselves from betrayal and injury, a phenomenon called immorality bias. Under this model, participants focus more on immoral words and do more perceptual processing, resulting in a slower RT in our tasks. Positively valenced words are usually processed faster than negative words (Kauschke et al., 2019), which would suggest that moral words are processed faster than words about immoral abstractions.

The availability of concrete referents could also drive the observed RT difference. Real-life referents are more easily simulated or visualized by perception because they have realistic counterparts grounded in perceptual experience, and they could have just been the default association for moral concepts (Holcomb et al., 1999; Binder et al., 2005; Wang et al., 2010; Laszlo and Federmeier, 2011). Our research cannot decide which of these views is correct.

Lakoff and Johnson (1999) argue that abstract concepts are based on concrete sensory experience. People might consider that the expression or representation of morality should hence be positive or upright. This mapping of moral conceptual metaphors could automatically activate the corresponding spatial representations and subtly influence an individual's cognition through their coding of that spatial representation. Studies have shown that concrete concepts were more easily simulated or visualized by perception because they have realistic counterparts to provide a perceptual experience (Holcomb et al., 1999; Binder et al., 2005; Wang et al., 2010; Laszlo and Federmeier, 2011). In life, people might pay more attention to morality than immorality and could be more sensitive to concepts related to morality, whereas the immoral concepts had not

formed an obvious metaphorical representation in people's minds. Moral intuition is essentially a stable, innate moral belief, knowledge, or ability (Sinnott-Armstrong et al., 2010). Moreover, Chinese cultural iconography and language are profoundly metaphorical. People generally use words such as “一身正气” (yī shēn zhèng qì, upright) to describe moral people. CMT's predictions—that we use concrete features of the physical world (the origin domain) to imbue a sense of physicality to abstract moral and social concepts (the target domain)—bore out using the Chinese language, providing new evidence of CMT's generalizability as a theory of human moral cognition and communication (Williams et al., 2009; Landau et al., 2010; Pecher et al., 2011).

From the perspective of vision, it has been found that words related to “moral” can effectively activate the response of people to “up,” whereas words related to “immoral” can activate the response of individuals to “down,” namely “moral up, immoral down” (Chasteen et al., 2010). Moreover, the concept of “以右为尊” (take the right side as honor) has existed in traditional Chinese culture since ancient times. People associate the word “右” (right) with positive things. Metaphor and metonymy play important roles in the representations of objects and events and the constructions of mental images in terms of basic everyday experiences (Landau et al., 2010).

Limitations of this study highlight the need for future research. Firstly, our research only focused on word shape and word rotation. Other sorts of deformations may help clarify the facilitating effect of the physical reference point as a metaphorical guide for identifying symbolic language. Secondly, there may be some potential confounding effects in the manipulation of the valence of moral and immoral words. The positive valence is inherent to moral stimuli as well as negative valence to immoral stimuli. The “right side” metaphorical advantage for moral words could be because they are positive valued words and the “right” advantage works as well for other positive valued abstract concepts. We should pay more attention to this effect and make a clearer distinction between the potency and manipulation of words in future studies. Thirdly, our experiments could be tinged with error coming from individual's highly variant skills with written language. For future research, we could adopt other experimental paradigms, such as the implicit association test, using images and audiation paradigms to see how effectively CMT predicts individual behavior across the senses or in different cognitive frameworks. From the perspective of experimental technology, advanced cognitive neural technologies, for example, ERP and fMRI, could also be used to shed new light on how the mapping between moral concepts and morphological metaphors leads to firm inferences.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by North China University of Science and Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YiL conceived the manuscript, ran statistical analyses, and contributed to the manuscript. KL contributed to the manuscript. LL conceived, modified the manuscript, and contributed to the

manuscript. JZ, YuL, BD, and HW conceived the manuscript and contributed to the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Hebei Social Science Foundation (HB20SH013), Hebei provincial universities basic scientific research operating expenses project (JSQ2019006), Hebei Social Science development research Project (20200303120), and the National Natural Science Foundation of China (31970990).

REFERENCES

- Barsalou, L. W. (2011). Integrating Bayesian analysis and mechanistic theories in grounded cognition. *Behav. Brain Sci.* 34, 191–192. doi: 10.1017/S0140525X11000197
- Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., and Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *J. Cogn. Neurosci.* 17, 905–917. doi: 10.1162/0898929054021102
- Bolin, Y., Zhao, L., and Shulin, Z. (1995). Effect of complexity on the recognition of skewed Chinese characters. *Acta Psychol. Sin.* 27, 337–343.
- Camgöz, N., Yener, C., and Güvenç, D. (2002). Effects of hue, saturation, and brightness on preference. *Color Res. Appl.* 27, 199–207. doi: 10.1002/col.10051
- Chasteen, A. L., Burdzy, D. C., and Pratt, J. (2010). Thinking of God moves attention. *Neuropsychologia* 48, 627–630. doi: 10.1016/j.neuropsychologia.2009.09.029
- Chen, R., Sai, J., Zhu, Q., Zhou, R., Li, P., and He, S. (2018). Horizontal spatial metaphors for morality: a cross-cultural study of han Chinese students and ethnic minority hui students in china. *Front. Psychol.* 9:1145. doi: 10.3389/fpsyg.2018.01145
- Dove, G. (2010). On the need for embodied and dis-embodied cognition. *Front. Psychol.* 1:242. doi: 10.3389/fpsyg.2010.00242
- Federica, S., and Tagini, S. (2017). The Stroop color and word test. *Front. Psychol.* 8:557. doi: 10.3389/fpsyg.2017.00557
- Gibbs, R. W., and Berg, E. A. (1999). Embodied metaphor in perceptual symbols. *Behav. Brain Sci.* 22, 617–618. doi: 10.1017/S0140525X99312140
- Haidt, J., and Algor, S. (2004). “Moral amplification and the emotions that attach us to saints and demons,” in *Handbook of Experimental Existential Psychology*, eds J. Greenberg, S. L. Koole, and T. Pyszczynski (New York, NY: Guilford), 322–335.
- Haidt, J., and Joseph, C. (2004). Intuitive ethics: how innately prepared intuitions generate culturally variable virtues. *Daedalus* 133, 55–66. doi: 10.2307/20027945
- Hill, P. L., and Lapsley, D. K. (2009). The ups and downs of the moral personality: why it's not so black and white. *J. Res. Pers.* 43, 520–523. doi: 10.1016/j.jrp.2009.01.014
- Holcomb, P. J., Kounios, J., Anderson, J. E., and West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 721–742. doi: 10.1037/0278-7393.25.3.721
- Jewell, G., and McCourt, M. E. (2000). Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia* 38, 93–110. doi: 10.1016/S0028-3932(99)00045-7
- Kauschke, C., Bahn, D., Vesker, M., and Schwarzer, G. (2019). The role of emotional valence for the processing of facial and verbal stimuli—positivity or negativity bias? *Front. Psychol.* 10:1654. doi: 10.3389/fpsyg.2019.01654
- Koriat, A., and Norman, J. (1985). Reading rotated words. *J. Exp. Psychol. Hum. Percept. Perform.* 11, 490–508. doi: 10.1037/0096-1523.11.4.490
- Lakoff, G. (2008). “The neural theory of metaphor,” in *The Cambridge Handbook of Metaphor and Thought*, ed. R. W. Gibbs Jr. (New York, NY: Cambridge University Press), 17–38.
- Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.
- Lakoff, G., and Johnson, M. (2008). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Landau, M. J., Meier, B. P., and Keefer, L. A. (2010). A metaphor enriched social cognition. *Psychol. Bull.* 136, 1045–1067. doi: 10.1037/a0020970
- Laszlo, S., and Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology* 48, 176–186. doi: 10.1111/j.1469-8986.2010.01058.x
- Li, H., and Cao, Y. (2017). Who's holding the moral higher ground: religiosity and the vertical conception of morality. *Pers. Individ. Diff.* 106, 178–182. doi: 10.1016/j.paid.2016.11.016
- Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychon. Bull. Rev.* 15, 838–844. doi: 10.3758/PBR.15.4.838
- Meier, B. P., Robinson, M. D., and Clore, G. L. (2004). Why good guys wear white: automatic inferences about stimulus valence based on brightness. *Psychol. Sci.* 15, 82–87. doi: 10.1111/j.0963-7214.2004.0150202.x
- Pecher, D., Boot, I., and Van Dantzig, S. (2011). Abstract concepts: sensory-motor grounding, metaphors, and beyond. *Psychol. Learn. Motiv.* 54, 217–248. doi: 10.1016/B978-0-12-385527-5.00007-3
- Schnall, S., Benton, J., and Harvey, S. (2008). With a clean conscience: cleanliness reduces the severity of moral judgments. *Psychol. Sci.* 19, 1219–1222. doi: 10.1111/j.1467-9280.2008.02227.x
- Schneider, I. K., Stapels, J., Koole, S. L., and Schwarz, N. (2020). Too close to call: spatial distance between options influences choice difficulty. *J. Exp. Soc. Psychol.* 87:103939. doi: 10.1016/j.jesp.2019.103939
- Ściagała, K., and Indurkha, B. (2016). “The influence of verticality metaphor on moral judgment and intuition,” in *Proceedings of the IEEE International Conference on Cognitive Infocommunications*, (Piscataway, NJ: IEEE).
- Sherman, G. D., and Clore, G. L. (2009). The color of sin: white and black are perceptual symbols of moral purity and pollution. *Psychol. Sci.* 20, 1019–1025. doi: 10.1111/j.1467-9280.2009.02403.x
- Sinnot-Armstrong, W., Young, L., and Cushman, F. (2010). Moral intuitions as heuristics. *Moral Psychol. Handb.* 26, 246–272. doi: 10.1111/j.1466-769X.2010.00478.x
- Stroop, J. R. (1938). Factors affecting speed in serial verbal reactions. *Psychol. Monogr.* 50, 38–48. doi: 10.1037/h0093516
- Tversky, B. (2008). “Embodied spatial cognition,” in *Proceedings of the “XXIX International Congress of Psychology”*, Berlin.
- Wang, H. L., Lu, Y. Q., and Lu, Z. Y. (2016). Moral-up first, immoral-down last: the time course of moral metaphors on a vertical dimension. *Neuroreport* 27:247. doi: 10.1097/WNR.0000000000000528
- Wang, J., Conder, J. A., Blitzler, D. N., and Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: a meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 31, 1459–1468. doi: 10.1002/hbm.20950

- Williams, L. E., Huang, J. Y., and Bargh, J. A. (2009). The scaffolded mind: higher mental processes are grounded in early experience of the physical world. *Eur. J. Soc. Psychol.* 39, 1257–1267. doi: 10.1002/ejsp.665
- Xinya, Y., and Zhongyi, L. (2015). The embodiment and culture of metaphor. *Psychol. Sci.* 38, 1081–1086.
- Yang, J. P., Guo, X. M., and Wang, X. C. (2017). Metaphorical representation of moral concepts: evidence from red/white color, left/ right position and upright/ skew font. *Acta Psychol. Sin.* 49, 875–885. doi: 10.3724/SP.J.1041.2017.00875
- Yu, N., Wang, T., and He, Y. (2016). Spatial subsystem of moral metaphors: a cognitive semantic study. *Metaphor Symbol* 31, 195–211. doi: 10.1080/10926488.2016.1223470
- Zhai, D. X., Guo, Y. L., and Lu, Z. Y. (2018). A dual mechanism of cognition and emotion in processing moral-vertical metaphors. *Front. Psychol.* 9:1554. doi: 10.3389/fpsyg.2018.01554
- Zhong, C. B., and Liljenquist, K. (2006). Washing away your sins: threatened morality and physical cleansing. *Science* 313, 1451–1452. doi: 10.1126/science.113072

Conflict of Interest: 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 © 2020 Liu, Li, Li, Zhang, Lin, DiFabrizio and Wang. 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.



Explicit (Not Implicit) Attitudes Mediate the Focus of Attention During Sentence Processing

Oleksandr V. Horchak* and Margarida Vaz Garrido

Centro de Investigação e Intervenção Social, Iscte-Instituto Universitário de Lisboa, Lisbon, Portugal

OPEN ACCESS

Edited by:

Pia Knoeferle,
Humboldt University of Berlin,
Germany

Reviewed by:

Lotte Meteyard,
University of Reading,
United Kingdom
Likun Zhan,
Beijing Language and Culture
University, China

*Correspondence:

Oleksandr V. Horchak
Oleksandr.Horchak@iscte-iul.pt

Specialty section:

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

Received: 15 July 2020

Accepted: 07 December 2020

Published: 23 December 2020

Citation:

Horchak OV and Garrido MV
(2020) Explicit (Not Implicit) Attitudes
Mediate the Focus of Attention During
Sentence Processing.
Front. Psychol. 11:583814.
doi: 10.3389/fpsyg.2020.583814

Many studies showed that comprehenders monitor changes in protagonists' emotions and actions. This article reports two experiments that explored how focusing comprehenders' attention on a particular property of the protagonist dimension (e.g., emotional or action state) affects the accessibility of information about target objects mentioned in the sentence. Furthermore, the present research examined whether participants' attitudes toward the issues described in the sentence can modulate comprehension processes. To this end, we asked participants to read sentences about environmental issues that focused comprehenders' attention on different mental and physical attributes of the same entities (protagonists and objects) and then self-report their own thoughts on the topic of environment by responding to the items assessing their environmental awareness. Importantly, we manipulated the task requirements across two experiments by administering a self-report task (Experiment 1), which required the participants to rate the seriousness and the frequency of the problem mentioned in a sentence; and administering a sentence-picture verification paradigm (Experiment 2), which required the participants to merely indicate if the object depicted in the picture (related to a certain environmental problem) was mentioned in the preceding sentence. The results of these experiments suggest that the focus of a sentence on the environmental problem (rather than the protagonist's emotion and action) enhances the accessibility of information about environmental issues (e.g., plastic garbage); that the comprehender's level of environmental awareness influences one's attention during sentence processing; and that comprehender characteristics significantly modulate comprehension processes only when the measures tap into explicit (and not implicit) processes.

Keywords: situation model, protagonist dimension, comprehender characteristics, language comprehension, environmental attitude

INTRODUCTION

Since the introduction of the construct of mental or situation model (Kintsch and van Dijk, 1978; Johnson-Laird, 1983), the notion that language comprehension requires the construction of mental representations of the agents, objects, locations, events, and actions described in a text has become a mainstream position adopted by many researchers in the areas of linguistics, psychology, and more generally, cognitive science. This perspective has generated many interesting lines of research that allowed us to come closer to answering the question how comprehenders understand the meaning of language. According to Zwaan et al.'s (1995a) event-indexing model, comprehenders

monitor information described in a story at the event level, whereby each event can be indexed on the following five dimensions: time (Magliano and Schleich, 2000; Rinck et al., 2003), space (Glenberg et al., 1987; Morrow et al., 1989; Rinck and Bower, 2000), causation (Trabasso and Sperry, 1985; Trabasso and Suh, 1993), motivation (Lutz and Radvansky, 1997; Rinck and Bower, 2004), and protagonists and objects (Myers et al., 1994; Cook et al., 1998). Thus, from a situation model perspective comprehenders create a coherent representation of story events by means of the immersive experience of the story world, which is very similar to how people track and process events in real-life (Zwaan, 2004; Mulcahy and Gouldthorp, 2016).

In their review of situation model research, Zwaan and Radvansky (1998) argued that such entities as protagonists and objects form the “meat” of situation model construction. These entities, in turn, are defined by their corresponding properties (e.g., physical and mental attributes) that are the most relevant for meaning-making processes during language comprehension. Many experiments confirmed compellingly this claim. With regard to such properties as character traits, Albrecht and O’Brien (1993) showed that participants’ reading times were slower when they read about a vegetarian ordering a hamburger, thus suggesting that comprehenders store the protagonists’ traits in memory during reading. With regards to emotional states, Gernsbacher et al. (1992) demonstrated that critical sentences were read more slowly when they contained an emotion word that is inconsistent with the emotional state implied by the story (e.g., processing the word “pride” after reading how someone was fired). Similarly, Komeda and Kusumi (2006) discovered that discontinuities in the emotional dimension (e.g., *worry-relief* vs. *relief-relief*) lead to significant increases in reading times, further implicating a constant situation model updating during reading. With regards to goals, Long et al. (1992) tested and confirmed a hypothesis that understanding a sentence like “The dragon kidnapped the daughters” leads to the generation of superordinate goal inferences (e.g., the dragon will eat the daughters). Rinck and Bower (2004) found that objects relevant to the current goal (e.g., a xerox machine when reading about the action of printing) were more accessible for readers than those that are irrelevant. Thus, altogether these results support the claim that the protagonists, objects, and their properties are the core around which situation models are created.

The above studies on the importance of entities and their properties (e.g., emotional states, action) made significant headway in assessing the influence of dimensional focus during situation model construction. As reviewed above, one common finding in this literature is that reading times increase when there are some discontinuities along a particular single property, such as, for instance, mismatching emotion or goal-irrelevant action (see Theriault and Rinck, 2007, for discussion). This finding is consistent with a processing load hypothesis of the Event-indexing model (Zwaan et al., 1995a) and an Event Segmentation theory (Zacks et al., 2007), which suggest that comprehenders should find it more difficult to integrate the current event into their situation model when there are few indices that are shared between the past and present events. Therefore, increased reading times reflect the fact that readers need to update a situation model

to be able to incorporate new information that mismatches, even if partially, the situation described by the previous information.

Related to the idea of how readers guide their focus of attention during language comprehension is the research on the impact of situation models on memory retrieval. A series of experiments (Radvansky and Zacks, 1991; Radvansky et al., 1993; Sohn et al., 2004) used a so-called *fan effect* paradigm (Anderson, 1974) to demonstrate that response times are increased as a function of the number of associations with a concept stored in memory. Among the most popular of such experiments are those in which participants have to memorize sentences that describe objects in either a single-location condition or a multiple-location condition. A fan effect (i.e., an increase in retrieval time) is usually observed in a speeded-recognition test when a single object is described as being in several locations (e.g., “The painting is in the hotel,” “The painting is in the store,” “The painting is in the store”) than when different objects are described as being in the same location (e.g., “The painting is in the hotel,” “The wardrobe is in the hotel,” “The bed is in the hotel”). Such results line up with arguments that in a multiple-location condition different situation models are activated that interfere with a comprehender’s ability to retrieve the desired mental representation (Radvansky et al., 1998).

Although the role of the aforementioned inhibitory processes has been at the focus of research in cognitive psychology for quite some time (see Radvansky, 1999, for discussion), significant issues remain to be addressed. The first question concerns the extent to which focusing comprehenders’ attention on a particular property of the entity affects the accessibility of objects mentioned in the sentence. As discussed before, previous empirical research has mostly examined how the accessibility of information is affected by the entity’s single property (e.g., emotion or action). While this research has deepened our understanding of the specifics of each individual property, it is somewhat unclear how multiple properties of the same entity determine how facts are integrated into a situation model. For example, if the sentences (1) “John *noticed* the garbage on the beach,” (2) “John *got upset* with the garbage on the beach,” and (3) “John *picked up* the garbage on the beach” are processed, a comprehender is likely to represent these differently. This is the case because all of these sentences share the same concept (i.e., garbage on the beach) with one fundamental difference: the sentence (1) – places emphasis on the environmental issue, the sentence (2) – on the protagonist’s emotional state and the sentence (3) – on the protagonist’s action. Therefore, the critical research question is: will properties of the model that are currently at focus receive high activation, so that information about them will impair accessibility to the objects (i.e., garbage) mentioned in the sentence? This is a fairly straightforward extension of previously discussed research showing that readers infer information related to emotion and action, thus making it highly accessible in memory.

The second way in which we hope to advance our understanding of the relation between situation model construction and language processing is by examining whether the participants’ attitudes and sensitivity toward the issues described in the sentence can modulate comprehension

processes. While previous research related to situation model construction specified how events occur across space and time, involve protagonists and objects, and have intentional and causal structures, few empirical attempts have been made to explore how language processing mechanisms interact with the characteristics of the comprehender (see Knoeferle, 2019, for an in-depth discussion). This is unfortunate as without integrating this kind of information into a situation model there is no way to know whether language processing can vary as a function of comprehender characteristics. Consider, for example, a situation in which environmental problems such as air pollution, water pollution, and global warming are discussed. There is a robust and well-established literature showing that people considerably differ in their pro-environmental attitudes. For some people, environmental awareness is important to the extent that they feel a responsibility to act (Liu and Leiserowitz, 2009). For others, in contrast, environmental awareness is less important, so that they do not think there is a need to minimize the negative impact of people's actions on the surrounding environment (Kennedy et al., 2009). Given this difference in attitudes, it is reasonable to suggest that a comprehender may adopt a somewhat different interpretation of a sentence depending on the level of his/her environmental awareness. The experiments to be reported test this hypothesis, specifically examining the influence of the comprehender's environmental awareness upon the accessibility of environment-related information in different conditions of attentional focus (i.e., object-focus vs. emotional state vs. action state) during sentence comprehension.

Documenting the contributions of participants' attitudes in situation model construction could provide a more comprehensive understanding of the language comprehension processes. However, social cognition research suggests that it could be useful to distinguish between tasks that tap into unconscious automatic responses and tasks that tap into conscious intentionally edited responses when assessing the mediating role of attitudes (Fazio and Olson, 2003; Hofmann et al., 2005). Meissner et al. (2019) argued that tests of implicit measures of attitudes (i.e., tests that obscure the content of measurement) like Implicit Association Test (Greenwald et al., 1998) aim to measure evaluation ("liking") instead of motivation ("wanting"), thus suggesting that the superiority of these tests over self-report measures is questionable. Similarly, Wilson et al. (2000) suggested that explicit and implicit measures oftentimes reflect different evaluations of the same object, given that implicit tests measure highly stable, old representations and explicit tests measure more recently acquired, new representations. This suggests that previously formed representations may still be accessible in memory when new contradicting information about a concept is acquired. As noted by Gawronski and De Houwer (2014), the consequence of this is that people may have two distinct attitudes toward the same concept: the previously acquired "implicit" attitude that gets activated automatically upon encountering a relevant stimulus; and a more recently acquired "explicit" attitude that requires some controlled processes (conscious effort) to be successfully retrieved from memory. The latter point is of particular interest because it is possible that task demands may determine whether the effects

of comprehender characteristics and, more generally, of the socially interpreted context are detected: if people consider themselves to be pro-environmentally committed but still find it hard to live up to their ideals, then there remains a possibility that only direct measurement procedures (i.e., self-reports) may capture people's pro-environmental attitudes during language comprehension. Thus, one of the aims of this research was to test this possibility by using two different tasks: a sentence-picture verification task (akin to "implicit" measure) and a self-report (akin to "explicit" measure).

THE PRESENT RESEARCH

On the basis of considerations outlined in the previous section, it seems plausible that comprehender characteristics may be a modulatory factor in language comprehension depending on whether the task taps into "explicit" or "implicit" attitudes. Therefore, in the present research we assessed the impact of attitudes on the accessibility of objects in different conditions of attentional focus while participants were performing two different tasks. In Experiment 1, we used a direct self-report measurement procedure, where participants had to read a sentence describing a certain environmental problem and then rate the seriousness and the frequency of the problem mentioned in the just-read sentence on a 10-point Likert scale. The focus of participants' attention in the critical sentences was manipulated by the critical verb used to describe an event (e.g., "John *noticed* the garbage on the beach," "John *got upset* with the garbage on the beach," "John *picked up* the garbage on the beach"). In Experiment 2, in contrast, we used an indirect measurement procedure, where participants read the same sentences as in Experiment 1, except that their task was to decide as quickly as possibly whether or not the subsequently presented pictured object (e.g., plastic garbage) was mentioned in the sentence. The just-mentioned sentence-picture verification paradigm from Experiment 2 should considerably reduce participants' ability to control their responses given social desirability, and hence the impact of attitudes on participants' responses may be considered resource-independent and unconscious (Gawronski and De Houwer, 2014). At the end of both Experiments 1 and 2, we assessed participants pro-environmental attitudes by asking them to respond to a 30-item Environmental Attitudes Inventory (Milfont and Duckitt, 2010), which was validated for Portuguese population (Domingues and Gonçalves, 2018).

EXPERIMENT 1

In Experiment 1, we measured participants' ratings of the seriousness and frequency of the problem mentioned in the sentence on 10-point Likert scale (1 = *Not serious at all* and 10 = *Very serious*). We predicted that the ratings will be higher in the "object-focus" condition (e.g., "John *noticed* the garbage on the beach") than in the "emotion-focus" ("John *got upset* with the garbage on the beach") and "action-focus" ("John *picked up* the garbage on the beach") conditions. Such a prediction

follows from previous studies showing that readers assign high priority to the protagonist's mental and physical states (see Theriault and Rinck, 2007; Bower, 2008, for a discussion), which, as a consequence, may interfere with the retrieval of relevant information. Rinck and Bower (2004) proposed that attentional focus during reading may be compared to a spotlight shining into a dollhouse, where the situation model is an inner stage that comprehenders construct in working memory. By using this analogy, when a target object is in focus (i.e., environmental problem), it is as if the spotlight is shining on it, thus increasing its availability in memory relative to when a protagonist's emotion or action are in focus. Consequently, when information is at a higher level of availability (as is the case in the "object-focus" condition), participants' ratings of the seriousness and the frequency of the problem mentioned in the sentence should increase. Furthermore, if comprehenders' attitudes mediate language comprehension processes, then we would expect to see an interaction such that the ratings in the "object-focus" condition should be higher only when taking into account the scores of the participants who are more environmentally concerned. Experiment 1 was designed to test these predictions.

Method

Sample Size and Ethical Requirements

We conducted a Power analysis in G*Power to determine the necessary number of observations for both Experiments 1 and 2. A power analysis was done using an effect size ($d = 0.31$, alpha level of 0.05, and a power of 0.80) from the study of Zwaan et al. (2002), which used the same sentence-picture verification paradigm as that used in Experiment 2 of the present research. According to Brysbaert and Stevens (2018), the typical effect size in many psycholinguistic experiments is even smaller ($d = 0.10$ or $d = 0.20$). Therefore, we determined our sample size by running a power analysis on a repeated measures ANOVA, a power of 0.80, an alpha level of 0.05, and a small ($\eta_p^2 = 0.02$) effect size. The results of this analysis suggested that we would need 99 participants to find an effect if it existed. Of note, this sample size is comparable to the thematically related studies of Theriault et al. (2006) and Bailey et al. (2017), which studied the influence of dimensional focus during situation model construction using different methods. To account for low accuracy, compliance with the task requirements, or data saving errors, we always attempted to collect data from at least 110 participants.

The study was carried out in accordance with the World Medical Association's Declaration of Helsinki. In line with the ethical guidelines of the host institution, participants from both Experiments 1 and 2 gave informed consent prior to participation and were fully debriefed about the purpose of the study upon completion.

Participants

One hundred and ten native Portuguese-speaking participants ($M_{\text{age}} = 27.90$, $SD_{\text{age}} = 10.47$; 49 males) were recruited via Prolific Academic (Palan and Schitter, 2018) by using the following prescreening criteria: Country of Birth = Portugal; Country of Residence = Portugal, and First (Native) Language = Portuguese.

Participants were compensated at a rate of £5.05 (British pounds) per hour.

Materials

Seventy-two sentences were created: 18 experimental sentence triads and 18 filler sentences. The experimental sentence triads "invited" participants to attend their attention to the environmental issue (e.g., "John *noticed* the garbage on the beach"); the protagonist's emotional state (e.g., "John *got upset* with the garbage on the beach"); and the protagonist's action (e.g., "John *picked up* the garbage on the beach"). Thus, we varied the proximity of target objects to the focus of attention by using verb information. Filler sentences were of the same format as experimental sentences with a fundamental difference: they described less environmentally serious events (e.g., "John brushed his teeth with the bamboo toothbrush"), positively laden events (e.g., "John liked his new bottle made of recycled glass"), and emotionally neutral ("John examined his bicycle") events. The purpose of these filler sentences was to discourage participants from selectively paying attention to more serious environmental problems and, consequently, regulate their responses in line with social desirability (Gawronski and De Houwer, 2014). The list of critical sentences used in Experiments 1 and 2 is provided in **Supplementary Appendix A**.

A short 36-item version of the Environmental Attitudes Inventory was used (Milfont and Duckitt, 2010) to calculate a mean score of participants' environmental awareness. This set of 7-point scales assesses environmental attitudes that underlie individual's behavior toward the environment on such dimensions as preservation and utilization. Domingues and Gonçalves (2018) and Domingues et al. (2019) assessed the structure of the Portuguese version of the Environmental Attitudes Inventory of Milfont and Duckitt (2010) and found that scales 5 (confidence in science and technology) and 12 (support for population growth policies) did not reflect participants' attitudes toward the environment in Portugal. Thus, we removed these scales and calculated the mean score of environmental awareness from the remaining 30 items (see **Supplementary Appendix B**, for the list of all items used).

Design and Procedure

Three lists of stimuli were created to counterbalance items and conditions, so that the same items that appeared in one sentence condition for some participants were in the different sentence conditions for other participants. Each participant was randomly assigned to one of the lists. This produced a 3 (sentence condition: object-focus, emotion-focus, action-focus) \times 3 (list) design, with sentences being a within-participants factor and list a between-participants factor. As list was not something that was actually manipulated, we did not include it as a factor in the reporting of statistical analyses. Each participant was exposed to 18 experimental sentences (i.e., six sentences for each sentence condition) and 18 filler sentences, and then provided their responses to the 30 items of the Portuguese version of the Environmental Attitudes Inventory.

Stimulus presentation was controlled in Qualtrics Survey Software. In the first part of the experiment, participants read

sentences, which were presented in a random order, about some environmental problem (one sentence at a time) and rated the seriousness and the frequency of the problem mentioned in the just-read sentence on a 10-point scale from 1 (*Not serious/frequent at all*) to 10 (*Very serious/frequent*). In the second part of the experiment, participants indicated their level of agreement with statements from the Portuguese version of the Environmental Attitudes Inventory on a 7-point scale from 1 (*Completely disagree*) to 7 (*Absolutely agree*).

Statistical Analysis

All statistical analyses were performed within the R programming environment version 4.0.0 (R Core Team, 2020) and several R packages. R Markdown files were used to generate code and the analyses were “knit” into html files. We fitted an ordinal mixed-effects model (cumulative link mixed model) with random effects of participants and items using an “ordinal” R package (Christensen, 2019). This model is optimal for Likert-type data as it allows to predict an ordinal dependent variable given one or more independent variables. Furthermore, this model permits to simultaneously account for two random variables in our design (participants and items), which is more powerful than separate by-participants (usually denoted as F_1) and by-items (usually denoted as F_2) analyses (Brysbaert and Stevens, 2018). The full or “maximal” (Barr et al., 2013) model contained sentence condition, environmental concern score, as well as the interaction between the two as fixed effects; a by-participants random slope for sentence condition; and an intercept for participants and items. No varying slopes were considered for items or environmental awareness as the test stimuli for these two factors were never repeated (Barr et al., 2013). To make interpretation of parameter estimates easier when testing an interaction between a continuous variable (environmental concern score) and a categorical variable (sentence type), environmental concern scores were standardized by subtracting the mean and dividing by the standard deviation for analysis. We used the default R “treatment” coding scheme, where each level of the categorical variable is contrasted to a specified reference level. In the present research, the “object-focus” condition was set as a reference category. Given that the interpretation of main effects is affected by the presence of an interaction term when fitting models using treatment contrasts (i.e., lower order terms reflect the effect of one variable at the specific level of the other independent variable; see Singmann and Kellen, 2020, for discussion), in both Experiments 1 and 2 we approached the analysis of data in the following way. First, we used a likelihood ratio test that compared the likelihood of a model with the interaction term to the likelihood of a model without it to determine whether the overall interaction between variables was significant. If a likelihood ratio test was significant, then we reported the results of the model involving an interaction term, which contained two fixed effects (i.e., sentence condition, environmental concern) and two interaction terms comparing each of the non-referent levels (“emotion-focus” condition, “action-focus” condition) to the referent level (“object-focus” condition). If a likelihood ratio test was not significant (i.e., the presence of an interaction was not established), then we removed the non-significant interaction

term from the model (to facilitate the interpretation of lower-order terms) and reran the analysis with fixed effects only (i.e., sentence condition, environmental concern). Following the guidelines of Meteyard and Davies (2020), detailed results of the final models from Experiments 1 and 2 are provided in **Supplementary Appendices C and D**, respectively.

RESULTS AND DISCUSSION

The data of major interest are provided in **Figure 1**¹. With regards to the ratings of seriousness, a likelihood ratio test of the “maximal” model with fixed effects of sentence condition, environmental concern, as well as the interaction between the two against the model with fixed effects of sentence condition and environmental concern revealed a significant difference between the models [$\chi^2(2) = 9.10$, $p = 0.01$], thus suggesting that the interaction between sentence condition and environmental concern was significant. The results of the “maximal” model showed a significant effect of “action-focus” condition (estimate = -0.79 , $SE = 0.11$, $z = -7.00$, $p < 0.001$, odds ratio = 0.45), reflecting the fact that ratings in this condition were significantly lower ($M = 5.49$, $SD = 3.38$) than in the “object-focus” condition ($M = 6.65$, $SD = 3.15$). There was no significant effect of “emotion-focus” condition (estimate = -0.14 , $SE = 0.12$, $z = -1.17$, $p = 0.244$, odds ratio = 0.87) despite the fact that ratings in this condition were also lower ($M = 6.42$, $SD = 3.24$) than in the reference (i.e., “object-focus”) condition. Furthermore, as expected, there was a significant effect of participants’ environmental concern (estimate = 0.52, $SE = 0.15$, $z = 3.56$, $p < 0.001$, odds ratio = 1.68), which reflects the fact that participants with high environmental concern rated environmental problems as more serious than those with low environmental concern. Finally, as illustrated in **Figure 1A**, there was a significant interaction between environmental concern and “action-focus” condition (estimate = -0.24 , $SE = 0.11$, $z = -2.13$, $p = 0.033$, odds ratio = 0.79), but no significant interaction between environmental concern and “emotion-focus” condition (estimate = 0.09, $SE = 0.12$, $z = 0.72$, $p = 0.470$, odds ratio = 1.09). This result suggests that action-focused sentences received lower ratings of seriousness than object-focused sentences only from participants with higher concern over environmental issues.

With regards to the rating of frequency, a likelihood ratio test of the “maximal” model with fixed effects of sentence condition, environmental concern, as well as the interaction between the two against the model with fixed effects of sentence condition and environmental concern showed no significant difference between the models [$\chi^2(2) = 1.44$, $p = 0.487$], thus suggesting that the interaction between sentence condition and environmental concern was not significant. Therefore, the “simplified” model without an interaction term was used in the reporting of results. The results of the “simplified” model revealed a significant effect of “action-focus” (estimate = -1.17 ,

¹Ratings for one item (about organic farming) were removed from the analysis as the average item rating was extremely low (2 out of 10). This indicates that participants did not believe that the sentence described a serious environmental problem.

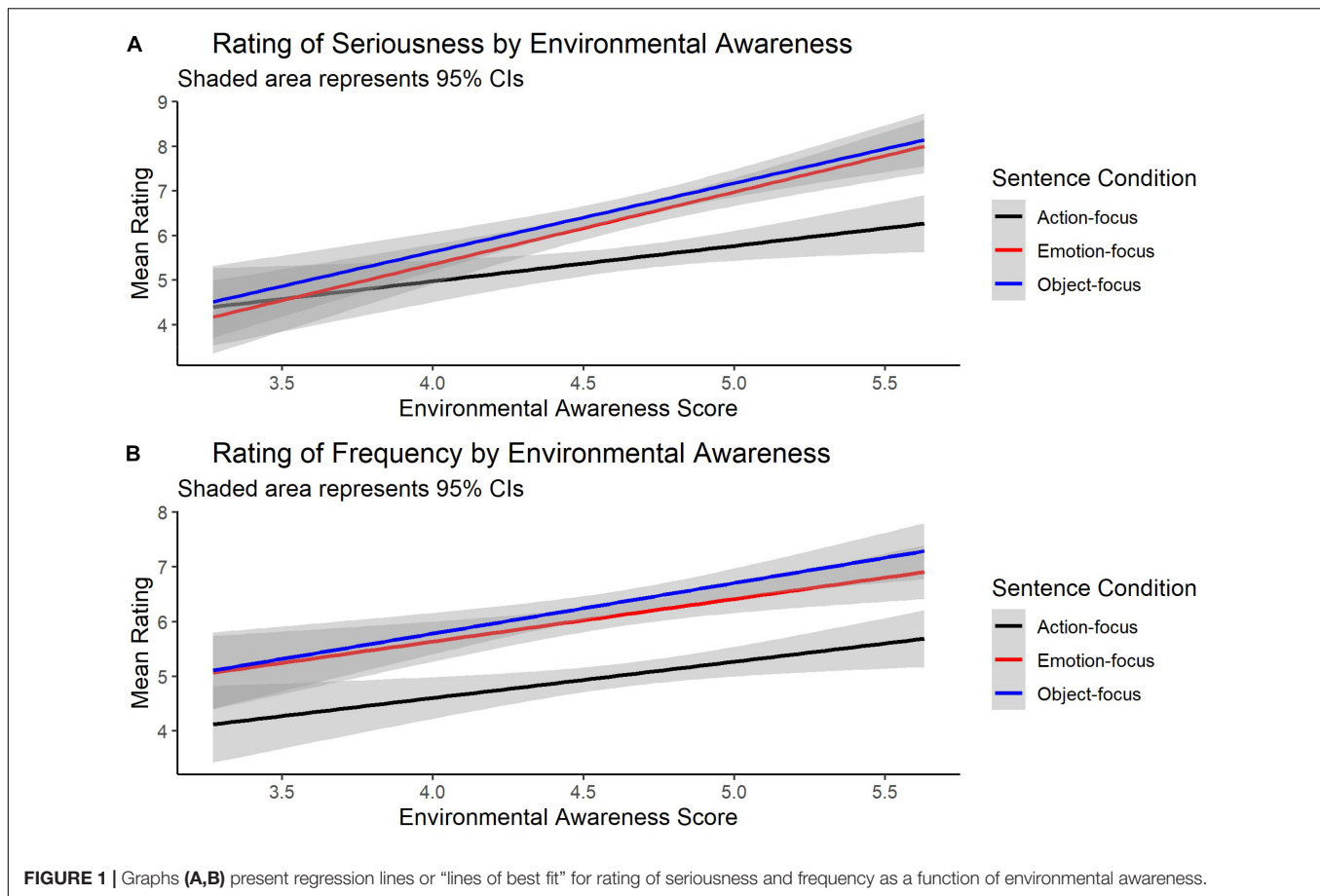


FIGURE 1 | Graphs (A,B) present regression lines or “lines of best fit” for rating of seriousness and frequency as a function of environmental awareness.

$SE = 0.11$, $z = -10.39$, $p < 0.001$, odds ratio = 0.31) and “emotion-focus” (estimate = -0.23 , $SE = 0.10$, $z = -2.24$, $p = 0.025$, odds ratio = 0.80) conditions. That is, ratings in the “action-focus” ($M = 5.04$, $SD = 2.75$) and “emotion-focus” ($M = 6.15$, $SD = 2.62$) conditions were lower than in the referent “object-focus” ($M = 6.39$, $SD = 2.69$) condition. There was also a significant effect of participants’ environmental concern (estimate = 0.27 , $SE = 0.12$, $z = 2.33$, $p = 0.020$, odds ratio = 1.31), which suggests that, as expected, more environmentally concerned participants rated the problems described in the sentence as more frequent than less environmentally concerned participants.

To summarize, the results from Experiment 1 suggest that participants’ responses depended on two independent variables: attentional focus and environmental concern. Ratings of the seriousness and frequency of an environment issue (e.g., plastic rubbish, raised water level, mud from rains) were lower when sentences focused on the protagonist’s emotion or action rather than the sentence object (environment issue). Furthermore, participants with higher concern over environmental issues (from a test of attitudes) gave higher ratings. Thus, both protagonist and comprehender characteristics affected the way participants focused their attention during sentence comprehension. Contrary to our prediction, however, ratings for object-focused sentences were not always higher for participants with greater environmental concern. With regards to

ratings of seriousness, only half of our prediction was validated: object-focused sentences produced higher ratings for those with greater environmental concern when compared to the action-focused sentences. However, this pattern was not repeated when ratings for object-focused sentences were contrasted with ratings for emotion-focused sentences. That is, the differences in ratings between object-focused and emotion-focused sentences were almost identical when taking into account the ratings of participants with both lower and greater environmental concern (Figure 1A). With regards to ratings of frequency, there was very little evidence of one variable (environmental concern) depending on the other (sentence condition): object-focused sentences always produced the highest ratings, whereas action-focused sentences produced the lowest ratings, with emotion-focused sentences (almost always) falling roughly in between (Figure 1B). Furthermore, in contrast to ratings of seriousness, object-focused sentences did produce significantly higher ratings when compared to emotion-focused sentences for the ratings of frequency. When put alongside evidence for the causal role of comprehenders’ characteristics for the rating of seriousness (as evidenced by a significant interaction between action-focused sentences as compared to the object-focused sentences), it appears that unique comprehenders’ characteristics, such as environmental attitudes, affect attentional focus to the extent to which a self-report measure asks participants to explicitly

evaluate their attitudes to the problem described in the sentence. Presumably the effect of comprehenders' attitudes and emotional sensitivity to the topic of environmental issues is less evident for the rating of frequency because this self-report measure obscures the content of measurement (i.e., pro-environmental awareness) to a much greater extent than the more explicit self-report measure of seriousness.

EXPERIMENT 2

The goal of Experiment 2 was to determine whether a moderating effect of environmental attitude found in Experiment 1 will stand up to empirical scrutiny in an experimental task that taps into automatic (rather than controlled) processes. To this end, we used a sentence-picture verification task in which participants read the same sentences as in Experiment 1, but this time their task was to decide whether the object shown in the subsequently presented picture was mentioned in the sentence. If automatic components of attitudes also affect language processing, then, similar to Experiment 1, the comprehender's environmental awareness should moderate comprehension. Consequently, we should observe a significant Sentence Condition \times Environmental Awareness interaction and/or a significant effect of Environmental Awareness.

Nonetheless, as discussed before, research on attitude formation suggests that measures of "explicit" and "implicit" attitudes often diverge in their results. Some studies found effects only on explicit measures (Gregg et al., 2006) while other studies found effects only on implicit measures (Olson and Fazio, 2006). Still others found effects on both explicit and implicit measures (Whitfield and Jordan, 2009). Such divergence between implicit and explicit tests in attitude formation research hints at the possibility that the results from a more "implicit" sentence-picture verification paradigm may also diverge from the results of a more "explicit" self-report questionnaire in the context of language comprehension research. One of the goals of Experiment 2 was thus to address this possibility.

Method

Participants

One hundred and thirty-five native Portuguese-speaking university students took part in Experiment 2 in exchange for course credit. Because of the coronavirus disease 2019 (COVID-19), students signed up for a study online through the cloud-based Participant Management Software SONA. Stimulus presentation was controlled by a web-based service PsyToolkit, which was designed for setting up, running, and analyzing online reaction-time (RT) experiments (Stoet, 2010, 2017). The responses of seven participants were discarded for having accuracy $<80\%$ on the main task. Overall, the results of Experiment 2 are based on data from 128 participants ($M_{age} = 24.47$, $SD_{age} = 7.04$), of whom 93 were females.

Materials and Design

The sentences and Environmental Attitudes Inventory were the same as in Experiment 1. Thirty-six colorful pictures were created

to accompany the sentences. In order to ensure that the pictures depicted the environmental problems we intended them to, prior to experiment three independent raters determined whether the shown pictures matched the environmental problems described in the sentences. All raters stated that our pictures matched the sentences, thus ensuring that the pictured stimuli we used actually depicted the environmental problems/environmentally related objects we wanted them to depict. All pictures were of the same size (385×385 pixels) and depicted the environmental problem described in the preceding sentence on a gray background (see Figure 2)².

Design and Procedure

The design was similar to Experiment 1, except that 36 pictures were added. Each participant saw 18 experimental sentence-picture pairs requiring "yes" responses and 18 filler sentence-picture pairs requiring "no" responses. Both experimental and filler sentences were identical in their format, thus making the potential strategy of selectively paying attention only to certain sentences suboptimal at best.

The procedure was similar to Experiment 1, except for the following important differences. First, the experimental flow was programmed in PsyToolkit web-based software (Stoet, 2010) that "forced" full screen mode on participants' computers. Participants could not do the task using any kind of keyboardless device (e.g., a smartphone, a tablet, etc.) and a Safari web browser. Kim et al. (2019) experimentally tested the reliability of this online service in comparison to a lab-based service E-Prime 3.0 in a complex psycholinguistic task and found that the results from both software programs were similar. Second, the experiment began with six practice trials to ensure that participants understood the instructions of the sentence-picture verification paradigm. Each trial started with a fixation cross in the middle of a screen for 1000 ms. Then a sentence appeared at the center of the screen until participants pressed the Spacebar, thus indicating that they finished reading the sentence. After a spacebar press, the sentence was replaced by a fixation cross for 500 ms, immediately followed by a picture of an object that was either mentioned or not in the preceding sentence. Participants indicated their decision by pressing an "S" button for a "yes" response and an "N" button for a "no" response. Third, in the final part of the experiment participants indicated their level of agreement with 30 statements from the Portuguese version of the Environmental Attitudes Inventory on a 7-point scale from 1 (*Completely disagree*) to 7 (*Absolutely agree*).

²To allow for accurate comparison of results between Experiments 1 and 2, we removed the same item (i.e., about organic farming) as in Experiment 1 from the analysis.



FIGURE 2 | Samples of critical pictures used in Experiment 2.

Data Treatment

Prior to analysis, incorrect responses, filler items, and the data of participants with an overall accuracy <80% on the main task were excluded. Second, response times (RTs) were checked for normality using histograms with normal curve and Q-Q plots. RTs were positively skewed, and thus log10 transformation was applied to get normal distributions (Baayen, 2008). Finally, responses exceeding ± 3 median absolute deviations (MAD) from the condition's median were removed. ± 3 MAD is a robust method of outlier treatment that is not affected by extremely high or extremely low values, and hence eliminates the need to define “arbitrary” lower and upper cutoff points (see Leys et al., 2013, for more information).

Statistical Analysis

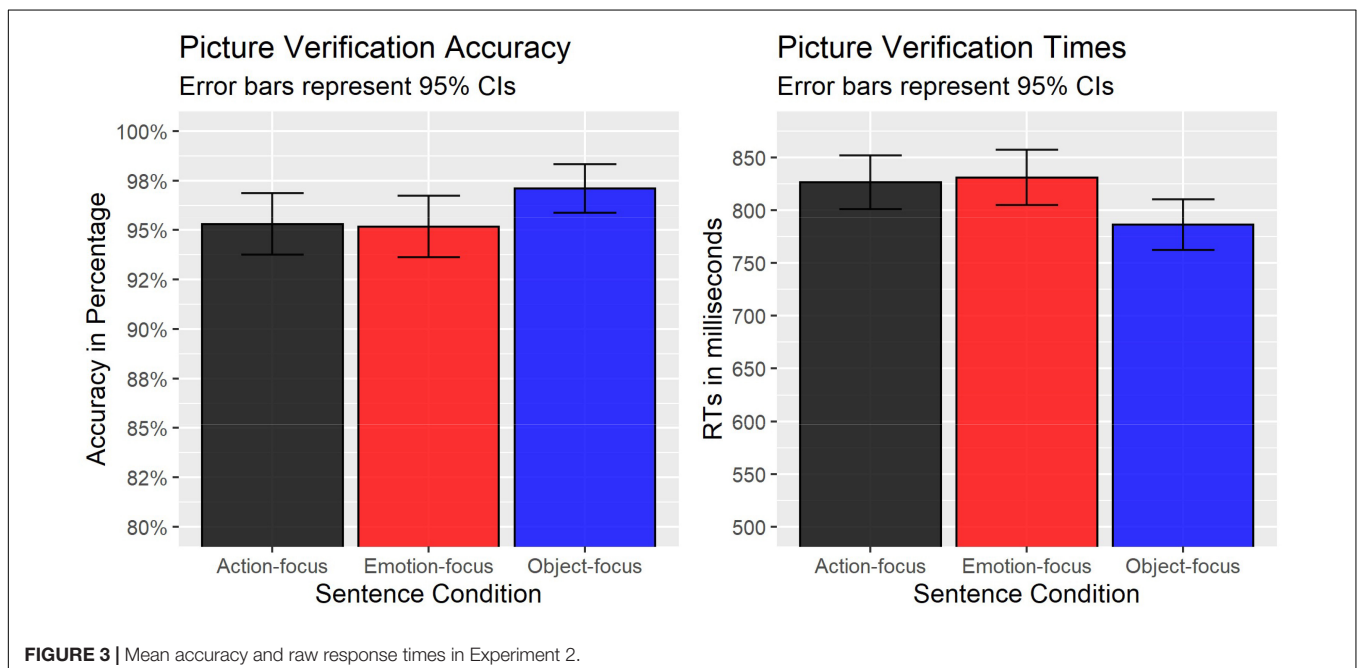
All statistical analyses were performed in R by using the lme4 package (Bates et al., 2015) and lmerTest package (Kuznetsova et al., 2017) to obtain *p*-values. Mixed-effects logistic regression and linear mixed-effects models with random effects of participants and items were run on accuracy scores and RT data, respectively. For both accuracy and response times analyses, we first always fitted the full variance-covariance structure of random effects (the so-called “maximal” model; Barr et al., 2013). However, if the random effects structure was not supported by the data, we removed the most complex part of the random effects structure (see Matuschek et al., 2017, for a discussion of “model selection criterion”). The “maximal” model (Barr et al., 2013) for the present research contained sentence condition, environmental awareness score, as well as the interaction between the two as fixed effects; a by-participants random slope for sentence condition; and an intercept for participants and items. For the same reasons as in Experiment 1, no random slopes were specified for items and environmental awareness scores. If the

“maximal” model failed to converge or produced a warning (e.g., a singular fit warning, which suggests that the model is overfitted) regarding the random effects structure, we first checked whether the model converges with a random effects structure for which no slope-intercept correlation term is specified. If this did not help, then we dropped a random slope in order not to (mistakenly) attribute all variability to the slope per participant instead of to the intercept (Brysbaert and Stevens, 2018). Finally, if one of the random intercepts was still found to be a redundant factor, then it was removed. With these considerations for random effects in mind, in the present research the best converging model for accuracy contained a random intercept for items and the best converging model for response times contained a random intercept for participants and items.

Results and Discussion

Accuracy

Participants' mean accuracy was high (95.9%). A likelihood ratio test of the best converging model (without warnings) with fixed effects of sentence condition, environmental concern, as well as the interaction between the two against the model with fixed effects of sentence condition and environmental concern showed no significant difference between the models [$\chi^2(2) = 0.39$, $p = 0.821$], thus suggesting that there was no evidence for the interaction between sentence condition and environmental concern. Therefore, the “simplified” model without an interaction term was used in the reporting of results. The results of the “simplified” model (Figure 3, left graph) showed a significant effect of “emotion-focus” (estimate = -0.58 , $SE = 0.28$, $z = -2.04$, $p = 0.042$, odds ratio = 0.56) condition and a trending effect of “action-focus” (estimate = -0.52 , $SE = 0.29$, $z = -1.81$, $p = 0.070$, odds ratio = 0.60) condition, reflecting the fact that accuracy in the “action-focus” (95%) and “emotion-focus”



(95%) conditions was lower than in the referent “object-focus” (97%) condition. Finally, there was no effect of participants’ environmental concern (estimate = 0.11, $z = 1.19$, $p = 0.234$, odds ratio = 1.14), which suggests that more environmentally concerned participants were not significantly more accurate in their responses than less environmentally concerned participants.

Response Times (RTs)

Similar to accuracy scores, a likelihood ratio test showed no significant difference between the models with the interaction term and without the interaction term [$\chi^2(2) = 2.39$, $p = 0.303$], thus suggesting that the interaction between sentence condition and environmental concern was not significant. Thus, the “simplified” model without the interaction term was used in the reporting of results. The analyses demonstrated that participants’ response times were positively correlated with corresponding effects in the accuracy scores, thus precluding speed-accuracy tradeoffs. More specifically, as shown in **Figure 3** (right graph), there was a significant effect of “emotion-focus” (estimate = 0.02, $SE = 0.01$, $t = 2.81$, $p = 0.005$) and “action-focus” (estimate = 0.02, $SE = 0.01$, $t = 2.80$, $p = 0.005$) conditions, reflecting the fact that RTs in the “action-focus” and “emotion-focus” conditions were slower than in the referent “object-focus” condition (see **Figure 3**). Finally, similar to accuracy analysis, there was no effect of participants’ environmental concern (estimate = 0.01, $SE = 0.01$, $t = 0.84$, $p = 0.404$) on RTs.

Thus, this pattern of results accords with the possibility outlined earlier: environmental problems are kept highly accessible in the “object-focus” condition (similar to results from Experiment 1), but comprehenders’ implicit (as compared to explicit) attitudes seem to have no significant effect on situation model construction and language comprehension processes.

GENERAL DISCUSSION

The present research was motivated by three goals. The first goal was to examine whether focusing readers’ attention differently on the protagonist dimension would affect the accessibility of environmental problems described in the sentence. The second goal was to begin documenting the contributions of such unique comprehender characteristics as pro-environmental attitudes to interpretation of linguistic input. The third goal was to explore in what task situations comprehenders’ pro-environmental awareness may modulate language comprehension processes. To this end, we asked participants to read sentences about environmental issues that focused on different mental and physical attributes (i.e., emotional state, action state, etc.) of the same entities (protagonists and objects) and then self-report their own thoughts, feelings, and behaviors on the topic of environment by responding to the items from the Environmental Attitudes Inventory. Importantly, we manipulated the task requirements across two experiments by administering a self-report task (Experiment 1), which required the participants to rate the seriousness and the frequency of the problem mentioned in a sentence; and administering a sentence-picture verification paradigm (Experiment 2), which required the participants to

merely indicate if the object depicted in the picture (related to a certain environmental problem) was mentioned in the preceding sentence. The results of these experiments suggest that the focus of a sentence on the environmental problem (i.e., “object-focus” condition) rather than the protagonist’s state (i.e., “emotion-focus” and “action-focus” conditions) enhances accessibility of the information related to environmental issues (e.g., plastic garbage); that the comprehender’s level of environmental awareness influences one’s attention during language processing; and that comprehender characteristics significantly modulate comprehension processes only when the measure taps into conscious representations.

The findings reported in this article add to the empirical evidence that comprehenders track various dimensions of situation models during language comprehension (e.g., Gernsbacher et al., 1992; Zwaan et al., 1995a,b; de Vega, 1996; Rinck and Bower, 2000; Rinck et al., 2003; Theriault et al., 2006; Kang et al., 2019); and that the comprehenders’ situational models capture information about a character’s emotional and actions states (e.g., de Vega, 1996; Borghi, 2004; Horchak et al., 2016; Horchak and Garrido, 2020a). Whereas most previous studies focused their efforts on understanding how multiple dimensions of the situation model (e.g., protagonists, intentionality, causation, etc.) are constructed and updated during language processing (Magliano and Schleich, 2000; Rich and Taylor, 2000; Rapp et al., 2001; Rinck and Weber, 2003), the present research assessed the extent to which focusing participants’ attention on the entity’s physical and mental attributes (i.e., a protagonist dimension) influences situation model construction. Our results show that when a sentence focused readers’ attention on protagonists’ emotional and action states (compared to when the attention was on the target environmental problem), then participants’ ratings were lower (Experiment 1) and response times to the picture probes (Experiment 2) were longer. These additive effects on ratings and response times are exactly what one would expect to observe in support of the hypothesis that readers direct their focus of attention to those aspects of the event that they consider to be in the spotlight for the current situation model. Our explanation of these results assumes that participants were building situation models organized around (1) a target environmental problem in the “object-focus” condition, (2) a protagonist’s mental state in the “emotion-focus” condition, and (3) a protagonist’s physical state in the “action-focus” condition. The experiments required from participants to either evaluate the seriousness and the frequency of environmental problem or verify if the environmentally related object (e.g., plastic bottle) was mentioned in the sentence, and therefore the greater was the number of “irrelevant” facts associated with a given problem, the more difficult it was for participants to retrieve the relevant fact. Thus, comprehenders “follow” not only the major character and objects mentioned in the sentence, but also their mental and physical attributes. The accessibility of the objects mentioned in the critical sentence “fades” as the focus moves on to other aspects of the described event.

The present research represents a significant departure from traditional research on situation model construction as it also

assessed the relevance of such comprehender characteristics as pro-environmental attitudes for sentence processing. Although some accounts of conceptual processing and language comprehension addressed how visual and action contexts affect language processing mechanisms (Zwaan, 2004; Knoeferle and Crocker, 2006; Altmann and Kamide, 2007; Altmann and Ekves, 2019), they have not tended to focus on how social evaluation may guide comprehension processes. Our findings fill this gap by integrating the insights from social psychological research on automatic and controlled components of attitudes with cognitive psychology research on situation model construction. More specifically, we asked in what task situations comprehenders' environmental awareness may mediate attentional focus during situation model construction. Our data suggest a more complex relation between language processing mechanisms and the characteristics of the comprehender than one could have predicted with confidence. Interestingly, the influence of comprehender characteristics on language processing seems to depend more on whether the measure aims to capture automatic or controlled components of attitudes than the measure itself. Such a conclusion follows from the results showing no moderating effect of comprehenders' pro-environmental attitudes on the ratings of frequency ("explicit" task in Experiment 1) and response times to picture probes ("implicit" task in Experiment 2). That is, while ratings of frequency and response times to picture probes are radically different types of tests, what both have in common is that they attempt to measure pro-environmental attitudes in a more automatic fashion, thus considerably reducing participants' ability to control their responses in line with social desirability. Indeed, participants' pro-environmental attitudes only moderated the more explicit rating of seriousness ("explicit" task in Experiment 1): the information about the environmental problem was equally accessible in all sentence conditions for participants with low environmental awareness, but not for participants with high environmental awareness. These findings thus support a conclusion that the influence of comprehenders' pro-environmental attitudes on language processing depends on whether automatic or controlled factors affect social evaluation and not the directness or indirectness of the test itself (see Ranganath et al., 2008, for a related discussion).

Given the pattern of the results observed, the obvious question is why automatic components of attitudes did not exhibit moderating effects on language comprehension processes in our research. This question is of great empirical and theoretical importance as most psycholinguistic tasks do not require introspection (i.e., the examination of one's own conscious thoughts and feelings) for the assessment of psychological attributes. Indeed, at this point it is difficult to say with any precision in what situations implicit attitudes moderate comprehension processes, but what seems to have occurred in the present research is that participants' unconscious reaction toward environmental issues was lagging behind their conscious desire for environmental improvement, which is consistent with a model of dual attitudes (Wilson et al., 2000). The central idea underlying this model is that previously formed representations cannot be easily erased from memory when people learn new

facts about something (e.g., environment is a really big problem) and then integrate them with old inconsistent information (e.g., environment is as important as many other problems) that reflects what they previously believed in. As nicely put by Gregg et al. (2006), highly stable, old representations may be compared to a credit card debt and excess calories that are difficult to cast aside. When people are faced with a certain stimulus, their conscious interpretation of it is supplemented by an automatic reaction. However, once the attitude toward a stimulus is formed through multiple direct experiences, attempts to subsequently override this attitude with a newly formed one will be successful to the extent to which the recently acquired knowledge is learned.

The present research has at least two limitations. The first limitation is that we assessed only those attitudes that are related to environmental awareness. It is thus unclear if the results would change if we considered attitudes that are more strongly related to socially sensitive topics (e.g., racial discrimination). For example, a meta-analysis of Greenwald et al. (2009) revealed that there is a considerable body of research showing an impact of old and recent experiences on both explicit and implicit measures, especially with regards to domains of stereotyping and prejudice. Furthermore, the effects of such characteristics of the comprehenders as age, education level, and knowledge of foreign languages, were successfully detected in language comprehension tasks using more implicit measurement methods, such as eye tracking (Huettig et al., 2011; Mishra et al., 2012; Carminati and Knoeferle, 2013, 2016; Ito et al., 2018). The second limitation is that the paradigms used in the present research do not provide a strong test between a propositional network view (e.g., Bower and Rinck, 2001) and a situation model view (e.g., Radvansky et al., 1998) that may explain the nature of the mental representation used to perform the tasks. For example, it was demonstrated that a sentence-picture verification task reveals the contents of mental representations that are activated as comprehenders read sentences, but does not provide a strong test for the claim that sensorimotor processes contribute functionally to language comprehension (e.g., Ostarek et al., 2019; Horchak and Garrido, 2020b). Therefore, it is not clear whether symbolic representations somehow interacted with sensorimotor (embodied) representations when participants were processing the test sentences in the current research (see Horchak et al., 2014, for a review of literature on contribution of symbolic and embodied representations to language comprehension). Although the present studies were not designed to address these issues, we believe that there are good reasons to believe that the propositional information was, at the very least, complemented with situational information beyond that provided by the sentence. Why would the focus on protagonists' mental attributes, for instance, make a difference relative to the focus on the target object in a sentence? Recall that emotion-focused sentences received similar ratings of seriousness when compared to the object-focused sentences. Perhaps it is because participants processed not only an explicitly mentioned emotional state of the protagonist (e.g., "John *got upset* with the garbage on the beach"), but also a few other automatic inferences, such as "Environment is important for John," "John must be sad right now," "John must not like people polluting the beach," etc.

These kinds of inferences likely caused mental attributes of the protagonist (on the one hand) and the environmentally related target object (on the other hand) compete for status as the concept by which the situation model is organized.

An important qualification of the present research is that it does not constitute direct evidence for the claim that mental representations of comprehenders were always organized in terms of real-world situations while they were reading the test sentences. It could be argued, for example, that a target word's meaning was simply represented in more detail when it was focused in a sentence's information structure (see Birch and Rayner, 1997; Sturt et al., 2004; Spalek et al., 2014; Gotzner et al., 2016; and Yang et al., 2019, for empirical evidence on linguistic focus hypothesis). Indeed, all our sentences were of the structure "Subject-Verb-Object" and participants' assessments of an environment issue (e.g., plastic rubbish, raised water level, mud from rains) could merely depend on whether the subject (i.e., a protagonist that experienced a particular action or emotion) or the object (i.e., an environmental issue) was in the focus of the sentence as defined by verb information. That is, focused information could have a privileged memory representation as a result of the governance of linguistic constructions. Although this scenario is consistent with the results of Experiment 2, our data from Experiment 1 have also shown that participants did not consider this as the only way to organize the information. If linguistic factor was the only one to guide attention to different aspects of the referential situation, then we should not have observed the following results: (1) participants with higher concern over environmental issues (from a test of attitudes) gave overall higher ratings; (2) object-focused sentences received similar ratings of seriousness when compared to the emotion-focused sentences; and (3) action-focused sentences received lower ratings of seriousness than object-focused sentences only with regards to the participants with higher concern over environmental issues. Given the above evidence, our findings constitute evidence that situation-based representations, at the very least, had to complement gist representations for adequate comprehension.

The remaining discussion will be focused on the following two aspects. First, we will discuss how the "Dynamic Text Comprehension (DTC)" framework of Rapp and van den Broek (2005) relates to the theoretical position advocated by this paper, namely that comprehenders' attention to events can be driven by task instructions and comprehenders' goals. Second, based on this discussion, we will provide a putative explanation of how comprehenders' could organize their representations around situation principles while reading the test sentences used in the present research.

According to the DTH framework, comprehending a text is tantamount to the construction of a situation model, whereby readers are able to not just understand the exact content of the text, but also infer implicit information (i.e., that was not directly stated in the text). Importantly, however, DTH proposes that successful comprehension arises from the interactive contributions of three factors: a text (e.g., text difficulty, genre, etc.), a reader (e.g., prior knowledge, cognitive abilities, etc.), and a task (e.g., instructions, task difficulty, etc.).

To demonstrate this, Rapp and Kendeou (2007) explored how particular methodologies used in "online" and "offline" comprehension tasks differentially modulate readers' attention. More specifically, participants were asked to read short stories implying a character's potential traits (e.g., "Albert's room is messy" ► Albert is sloppy) that ended with (1) a *simple refutation* of that trait (e.g., "Albert cared about the condition of his room" ► Albert is organized); (2) a more *explanatory refutation* explaining why an inferred trait might be wrong (e.g., "Albert cared about the condition of his room, but had only moved in to the house yesterday" ► Albert is organized); or (3) further *support* for that trait (e.g., "Albert's room is messy. He has some dirty laundry" ► Albert is sloppy). The researchers administered two types of tasks. In the "online" comprehension task participants read each sentence, one at a time, with reading times recorded for the outcome of the story (i.e., a part that either supported or refuted the initially described character trait). In the "offline" comprehension task participants were asked to explicitly indicate whether they agreed or disagreed with the outcome of the story. A key research question was whether comprehenders would update a character trait as a function of the information provided by the refutation. Rapp and Kendeou (2007) found that for *simple* refutations participants took longer to read trait-inconsistent than consistent outcomes, but for *explanatory* refutations participants took longer to read trait-consistent than inconsistent outcomes. They interpreted this finding as providing support for the claim that readers no longer expected characters to behave in trait-consistent ways when sufficient information (i.e., a condition of *explanatory* refutation) for that trait was provided. Perhaps even more interestingly, the results from "offline" task showed that the updating of trait information was observed for both types of refutations (i.e., participants always agreed that a previously encoded trait was incorrect), thus suggesting that "online" and "offline" tasks may encourage participants to focus on different aspects of a scene as a function of methodologies. In support of this is also other empirical evidence showing, for instance, that recall tasks encourage comprehenders to focus on the task at hand while ignoring prior knowledge, but judgments tasks encourage careful examination of the accuracy of information based on both what was read and what was previously experienced (e.g., Egidi and Gerrig, 2006).

The aforementioned theoretical and empirical evidence suggests that language comprehension arises not only from what a text contains, but also from task instructions and a comprehender's interest in a topic (as defined by world knowledge or beliefs). If we accept that the interaction of these factors may indeed encourage different profiles of a comprehender's attention, then it is reasonable to assume that the assumptions from DTH framework can also be extrapolated to the present research. More precisely, for the current experiments we may consider focus condition as a major sentence factor; the methodology as a major task factor; and a comprehender's level of environmental concern as a major reader factor. On this account, differences in methodologies between a questionnaire and a sentence-picture verification paradigm, either explicitly

or implicitly, could encourage readers to process information presented in a sentence differently. By using a self-report measure (a questionnaire) of seriousness, participants were explicitly asked to evaluate whether a sentence described a serious situation related to environment. It should come as no surprise then that the more the sentence aligned with participants' beliefs, the more predisposed they were to consider the sentence information carefully. Indeed, comprehenders particularly sensitive to environmental issues could focus not only on what was within linguistic focus (e.g., environmental issues), but also on what was within the focus of their own feelings or beliefs, such as the feeling that they should be doing more to help the environment, thus lingering on content such as "cried about" or "hated." Perhaps, then, it was precisely because of this why emotion-focused sentences received similar ratings of seriousness when compared to the object-focused sentences.

By using a self-report measure of frequency, in contrast, the true intent of a question (i.e., participants' *real* level of environmental concern) was obfuscated. Yet participants had enough time to carefully evaluate their level of agreement or disagreement with the information in the sentence. Notably, a self-report measure of frequency was effective at discouraging responses in line with social desirability because participants' evaluations of the environmental problems were now consistent with our predictions: ratings were lower when sentences focused on protagonist emotion or action, rather than the sentence object. But it is worth noting that the evidence of comprehenders' concern over environmental issues (i.e., higher concern = higher ratings; lower concern = lower ratings) was clearly observed for both ratings of seriousness and frequency.

Finally, by using a sentence-picture verification paradigm, participants were almost entirely discouraged from consciously evaluating the information in any particular way (e.g., if it matches their beliefs), and hence could easily adopt a strategy to focus on the most important information in a sentence, precisely because it was more relevant for completing the task (see Rapp and Mensink, 2011, for further discussion). As such task did not foster careful evaluation, it remains possible that participants primarily relied on the governance of linguistic constructions, which maintained focus on either the protagonist or sentence object (i.e., environmental problem). Whenever linguistic focus was on the environmental problem in a sentence (i.e., sentence object), participants' verification times of environment-related objects were faster.

Admittedly, our explanation of the observed pattern of results requires additional empirical support to further scrutinize how unique comprehenders' characteristics and task requirements influence sentence processing. It is our hope that our research will contribute to an agenda of items that merit discussion and future investigation to help us further develop theoretical accounts that assess the role of speaker and comprehender characteristics for situated language comprehension (e.g., Social Coordinated Interplay Account; Münster and Knoeferle, 2018).

According to our present analysis, unique comprehenders' characteristics such as attitudes help predict the variability of context effects during sentence processing. However, it is unlikely

that such characteristics may affect all kinds of information. The present research supports this claim in light of the results showing that the action-focused sentences were not so heavily moderated by attitudes. At this point in time it is hard to say with any precision why this happened. Findings from the literature on the action-attitude gap in environmental psychology provide some clues in this regard. More specifically, they suggest that the lack of a significant effect may be explained by constraints of behavioral control (Koger and Winter, 2010). According to this framework, while people report having sustainable behaviors toward the environment, their behavioral intention to actually act "ecologically" may lag behind the belief that this would only have a minimal impact on the environment. Thus, whereas it may be intuitive that the seriousness of the described environmental problem should be accessible in all sentence conditions (albeit to varying degrees), the focus of a sentence on the action might lead to questioning the validity of the facts (e.g., cleaning garbage on the beach is a waste of time as environmentally irresponsible behaviors outweigh sustainable behaviors). It remains to be seen, however, whether these predictions hold true in the task used to study language comprehension processes.

To conclude, the present research suggests that comprehenders' attitudes may alter how they focus on the major character and track his/her physical and mental attributes during sentence comprehension. Comprehenders' implicit attitudes (as compared to explicit ones) may create a stumbling block for investigating the role of comprehender characteristics during language comprehension, and hence the use of varied measurement procedures is warranted. Future research can explore to a much greater extent how attitudes related to more sensitive topics (e.g., prejudice and discrimination) affect language comprehension processes over the course of processing the sentence (e.g., eye tracking or word-by-word sentence design) to be able to better understand the functional mechanisms behind the obtained results.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/qbgzd/?view_only=d6cd8cc712aa4c69bcf3ea6ecb06128f.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

OH and MG idealized the study. OH and MG designed the sentence stimuli. OH designed the picture stimuli and drafted the manuscript. OH and MG involved in acquisition of data, analysis

and interpretation of data, critical revision of the manuscript, and approval of the submitted version for publication. Both authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Portuguese Foundation for Science and Technology by a grant awarded to the first author (SFRH/BPD/115533/2016).

REFERENCES

- Albrecht, J. E., and O'Brien, E. J. (1993). Updating a mental model: maintaining both local and global coherence. *J. Exp. Psychol.* 19, 1061–1070. doi: 10.1037/0278-7393.19.5.1061
- Altmann, G. T. M., and Ekves, Z. (2019). Events as intersecting object histories: a new theory of event representation. *Psychol. Rev.* 126, 817–840. doi: 10.1037/rev0000154
- Altmann, G. T. M., and Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: linking anticipatory (and other) eye movements to linguistic processing. *J. Mem. Lang.* 57, 502–518. doi: 10.1016/j.jml.2006.12.004
- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cogn. Psychol.* 6, 451–474. doi: 10.1016/0010-0285(74)90021-8
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge Core. Cambridge: Cambridge University Press.
- Bailey, H. R., Kurby, C. A., Sargent, J. Q., and Zacks, J. M. (2017). Attentional focus affects how events are segmented and updated in narrative reading. *Mem. Cogn.* 45, 940–955. doi: 10.3758/s13421-017-0707-2
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68:1. doi: 10.1016/j.jml.2012.11.001
- Bates, D. M., Mäechler, 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
- Birch, S., and Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Mem. Cogn.* 25, 653–660. doi: 10.3758/BF03211306
- Borghi, A. M. (2004). Object concepts and action: extracting affordances from objects parts. *Acta Psychol.* 115, 69–96. doi: 10.1016/j.actpsy.2003.11.004
- Bower, G. H. (2008). The evolution of a cognitive psychologist: a journey from simple behaviors to complex mental acts. *Annu. Rev. Psychol.* 59, 1–27. doi: 10.1146/annurev.psych.59.103006.093722
- Bower, G. H., and Rinck, M. (2001). Selecting one among many referents in spatial situation models. *J. Exp. Psychol.* 27, 81–98. doi: 10.1037/0278-7393.27.1.81
- Brysbaert, M., and Stevens, M. (2018). Power analysis and effect size in mixed effects models: a tutorial. *J. Cogn.* 1:9. doi: 10.5334/joc.10
- Carminati, M. N., and Knoeferle, P. (2013). Effects of speaker emotional facial expression and listener age on incremental sentence processing. *PLoS One* 8:e72559. doi: 10.1371/journal.pone.0072559
- Carminati, M. N., and Knoeferle, P. (2016). Priming younger and older adults' sentence comprehension: insights from dynamic emotional facial expressions and pupil size measures. *Open Psychol. J.* 9, 129–148. doi: 10.2174/1874350101609010129
- Christensen, R. H. B. (2019). *ordinal - Regression Models for Ordinal Data*. R Package Version 2019.12-10.
- Cook, A. E., Halleran, J. G., and O'Brien, E. J. (1998). What is readily available during reading? A memory-based view of text processing. *Discourse Process.* 26, 109–129. doi: 10.1080/01638539809545041
- de Vega, M. (1996). The representation of changing emotions in reading comprehension. *Cogn. Emot.* 10, 303–322. doi: 10.1080/026999396380268
- Domingues, R. B., Fernandes, J., Cabral, M., and Gonçalves, G. (2019). Psychometric properties of the portuguese environmental attitudes inventory / propriedades psicométricas do inventário português de atitudes ambientais. *PsyEcology* 10, 153–183. doi: 10.1080/21711976.2018.1554020

ACKNOWLEDGMENTS

We thank Bruno Moura for help with data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.583814/full#supplementary-material>

- Domingues, R. B., and Gonçalves, G. (2018). Assessing environmental attitudes in Portugal using a new short version of the environmental attitudes inventory. *Curr. Psychol.* 39, 629–639. doi: 10.1007/s12144-018-9786-x
- Egidi, G., and Gerrig, R. J. (2006). Readers' experiences of characters' goals and actions. *J. Exp. Psychol.* 32:1322. doi: 10.1037/0278-7393.32.6.1322
- Fazio, R. H., and Olson, M. A. (2003). Implicit measures in social cognition. research: their meaning and use. *Annu. Rev. Psychol.* 54, 297–327. doi: 10.1146/annurev.psych.54.101601.145225
- Gawronski, B., and De Houwer, J. (2014). "Implicit measures in social and personality psychology," in *Handbook of Research Methods in Social and Personality Psychology*, 2nd Edn, Eds H. Reis and C. M. Judd (Cambridge, MA: Cambridge University Press), 283–310. doi: 10.1017/cbo9780511996481.016
- Gernsbacher, M. A., Goldsmith, H. H., and Robertson, R. R. W. (1992). Do readers mentally represent characters' emotional states? *Cogn. Emot.* 6, 89–111. doi: 10.1080/02699939208411061
- Glenberg, A. M., Meyer, M., and Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *J. Mem. Lang.* 26, 69–83. doi: 10.1016/0749-596X(87)90063-5
- Gotzner, N., Wartenburger, I., and Spalek, K. (2016). The impact of focus particles on the recognition and rejection of contrastive alternatives. *Lang. Cogn.* 8, 59–95. doi: 10.1017/langcog.2015.25
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* 97, 17–41. doi: 10.1037/a0015575
- Gregg, A. P., Seibt, B., and Banaji, M. R. (2006). Easier done than undone: asymmetry in the malleability of implicit preferences. *J. Pers. Soc. Psychol.* 90, 1–20. doi: 10.1037/0022-3514.90.1.1
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., and Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Pers. Soc. Psychol. Bull.* 31, 1369–1385. doi: 10.1177/0146167205275613
- Horchak, O. V., and Garrido, M. V. (2020a). Dropping bowling balls on tomatoes: representations of object state-changes during sentence processing. *J. Exp. Psychol.* [Epub ahead of print]. doi: 10.1037/xlm0000980
- Horchak, O. V., and Garrido, M. V. (2020b). Is complex visual information implicated during language comprehension? The case of cast shadows. *Cogn. Sci.* 44:e12870.
- Horchak, O. V., Giger, J.-C., Cabral, M., and Pochwatko, G. (2014). From demonstration to theory in embodied language comprehension: a review. *Cogn. Syst. Res.* 29–30, 66–85. doi: 10.1016/j.cogsys.2013.09.002
- Horchak, O. V., Giger, J. C., and Garrido, M. V. (2016). Action contribution to competence judgments: the use of the journey schema. *Front. Psychol.* 7:448. doi: 10.3389/fpsyg.2016.00448
- Huettig, F., Singh, N., and Mishra, R. K. (2011). Language-mediated visual orienting behavior in low and high literates. *Front. Psychol.* 2:285. doi: 10.3389/fpsyg.2011.00285
- Ito, A., Pickering, M. J., and Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: a visual world eye-tracking study. *J. Mem. Lang.* 98, 1–11. doi: 10.1016/j.jml.2017.09.002

- Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.
- Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., and Altmann, G. T. M. (2019). The influence of state change on object representations in language comprehension. *Mem. Cogn.* 48, 390–399. doi: 10.3758/s13421-019-00977-7
- Kennedy, E. H., Beckley, T. M., McFarlane, B. L., and Nadeau, S. (2009). Why we don't "walk the talk": understanding the environmental values/behaviour gap in Canada. *Hum. Ecol. Rev.* 16, 151–160.
- Kim, J., Gabriel, U., and Gygax, P. (2019). Testing the effectiveness of the Internet-based instrument PsyToolkit: a comparison between web-based (PsyToolkit) and lab-based (E-Prime 3.0) measurements of response choice and response time in a complex psycholinguistic task. *PLoS One* 14:e0221802. doi: 10.1371/journal.pone.0221802
- Kintsch, W., and van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychol. Rev.* 85, 363–394. doi: 10.1037/0033-295X.85.5.363
- Knoeferle, P. (2019). Predicting (variability of) context effects in language comprehension. *J. Cult. Cogn. Sci.* 3, 141–158. doi: 10.1007/s41809-019-00025-5
- Knoeferle, P., and Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cogn. Sci.* 30, 481–529. doi: 10.1207/s15516709cog0000_65
- Koger, S., and Winter, D. D. N. (2010). *The Psychology of Environmental Problems*. New York, NY: Psychology Press.
- Komeda, H., and Kusumi, T. (2006). The effect of a protagonist's emotional shift on situation model construction. *Mem. Cogn.* 34, 1548–1556. doi: 10.3758/BF03195918
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi: 10.1016/j.jesp.2013.03.013
- Liu, J. C.-E., and Leiserowitz, A. A. (2009). From red to green? *Environ. Sci. Policy Sustain. Dev.* 51, 32–45. doi: 10.3200/ENV.51.4.32-45
- Long, D. L., Golding, J. M., and Graesser, A. C. (1992). A test of the on-line status of goal-related inferences. *J. Mem. Lang.* 31, 634–647. doi: 10.1016/0749-596X(92)90032-S
- Lutz, M. F., and Radvansky, G. A. (1997). The fate of completed goal information in narrative comprehension. *J. Mem. Lang.* 36, 293–310. doi: 10.1006/jmla.1996.2491
- Magliano, J. P., and Schleich, M. C. (2000). Verb aspect and situation models. *Discourse Process.* 29, 83–112. doi: 10.1207/S15326950dp2902_1
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001
- Meissner, F., Grigutsch, L. A., Koranyi, N., Müller, F., and Rothermund, K. (2019). Predicting behavior with implicit measures: disillusioning findings, reasonable explanations, and sophisticated solutions. *Front. Psychol.* 10:2483. doi: 10.3389/fpsyg.2019.02483
- Meteyard, L., and Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *J. Mem. Lang.* 112:104092. doi: 10.1016/j.jml.2020.104092
- Milfont, T. L., and Duckitt, J. (2010). The environmental attitudes inventory: a valid and reliable measure to assess the structure of environmental attitudes. *J. Environ. Psychol.* 30, 80–94. doi: 10.1016/j.jenvp.2009.09.001
- Mishra, R. K., Singh, N., Pandey, A., and Huettig, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability—Evidence from Indian low and high literates. *J. Eye Mov. Res.* 5:1. doi: 10.16910/jemr.5.1.3
- Morrow, D. G., Bower, G. H., and Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *J. Mem. Lang.* 28, 292–312. doi: 10.1016/0749-596X(89)90035-1
- Mulcahy, M., and Gouldthorp, B. (2016). Positioning the reader: the effect of narrative point-of-view and familiarity of experience on situation model construction. *Lang. Cogn.* 8, 96–123. doi: 10.1017/langcog.2014.45
- Münster, K., and Knoeferle, P. (2018). Extending situated language comprehension (accounts) with speaker and comprehender characteristics: toward socially situated interpretation. *Front. Psychol.* 8:2267. doi: 10.3389/fpsyg.2017.02267
- Myers, J. L., O'Brien, E. J., Albrecht, J. E., and Mason, R. A. (1994). Maintaining global coherence during reading. *J. Exp. Psychol.* 20, 876–886. doi: 10.1037/0278-7393.20.4.876
- Olson, M. A., and Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Pers. Soc. Psychol. Bull.* 32, 421–433. doi: 10.1177/0146167205284004
- Ostarek, M., Joosen, D., Ishag, A., de Nijs, M., and Huettig, F. (2019). Are visual processes causally involved in "perceptual simulation" effects in the sentence-picture verification task? *Cognition* 182, 84–94. doi: 10.1016/j.cognition.2018.08.017
- Palan, S., and Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *J. Behav. Exp. Finance* 17, 22–27. doi: 10.1016/j.jbef.2017.12.004
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Radvansky, G. A. (1999). Memory retrieval and suppression: the inhibition of situation models. *J. Exp. Psychol. Gen.* 128, 563–579. doi: 10.1037/0096-3445.128.4.563
- Radvansky, G. A., Spieler, D. H., and Zacks, R. T. (1993). Mental model organization. *J. Exp. Psychol.* 19, 95–114. doi: 10.1037/0278-7393.19.1.95
- Radvansky, G. A., and Zacks, R. T. (1991). Mental models and the fan effect. *J. Exp. Psychol.* 17, 940–953. doi: 10.1037/0278-7393.17.5.940
- Radvansky, G. A., Zwaan, R. A., Federico, T., and Franklin, N. (1998). Retrieval from temporally organized situation models. *J. Exp. Psychol.* 24, 1224–1237. doi: 10.1037/0278-7393.24.5.1224
- Ranganath, K. A., Smith, C. T., and Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *J. Exp. Soc. Psychol.* 44, 386–396. doi: 10.1016/j.jesp.2006.12.008
- Rapp, D. N., Gerrig, R. J., and Prentice, D. A. (2001). Readers' trait-based models of characters in narrative comprehension. *J. Mem. Lang.* 45, 737–750. doi: 10.1006/jmla.2000.2789
- Rapp, D. N., and Kendeou, P. (2007). Revising what readers know: updating text representations during narrative comprehension. *Mem. Cogn.* 35, 2019–2032. doi: 10.3758/bf03192934
- Rapp, D. N., and Mensink, M. C. (2011). "Focusing effects from online and offline reading tasks," in *Text relevance and learning from text*, eds M. T. McCrudden, J. P. Magliano, and G. J. Schraw (Greenwich, CT: Information Age), 141–164.
- Rapp, D. N., and van den Broek, P. (2005). Dynamic text comprehension: an integrative view of reading. *Curr. Direct. Psychol. Sci.* 14, 276–279. doi: 10.1111/j.0963-7214.2005.00380.x
- Rich, S. S., and Taylor, H. A. (2000). Not all narrative shifts function equally. *Mem. Cogn.* 28, 1257–1266. doi: 10.3758/BF03211825
- Rinck, M., and Bower, G. H. (2000). Temporal and spatial distance in situation models. *Mem. Cogn.* 28, 1310–1320. doi: 10.3758/BF03211832
- Rinck, M., and Bower, G. H. (2004). Goal-based accessibility of entities within situation models. *Psychol. Learn. Motiv.* 44, 1–33. doi: 10.1016/s0079-7421(03)44001-2
- Rinck, M., Gámez, E., Díaz, J. M., and De Vega, M. (2003). Processing of temporal information: evidence from eye movements. *Mem. Cogn.* 31, 77–86. doi: 10.3758/BF03196084
- Rinck, M., and Weber, U. (2003). Who when where: an experimental test of the event-indexing model. *Mem. Cogn.* 31, 1284–1292. doi: 10.3758/BF03195811
- Singmann, H., and Kellen, D. (2020). "An Introduction to Mixed Models for Experimental Psychology," in *New Methods in Cognitive Psychology*, eds D. H. Spieler and E. Schumacher (New York, NY: Routledge), 4–31. doi: 10.4324/9780429318405-2
- Sohn, M.-H., Anderson, J. R., Reder, L. M., and Goode, A. (2004). Differential fan effect and attentional focus. *Psychon. Bull. Rev.* 11, 729–734. doi: 10.3758/BF03196627
- Spalek, K., Gotzner, N., and Wartenburger, I. (2014). Not only the apples: focus sensitive particles improve memory for information-structural alternatives. *J. Mem. Lang.* 70, 68–84. doi: 10.1016/j.jml.2013.09.001
- Stoet, G. (2010). PsyToolkit: a software package for programming psychological experiments using Linux. *Behav. Res. Methods* 42, 1096–1104. doi: 10.3758/BRM.42.4.1096
- Stoet, G. (2017). PsyToolkit: a novel web-based method for running online questionnaires and reaction-time experiments. *Teach. Psychol.* 44, 24–31. doi: 10.1177/0098628316677643

- Sturt, P., Sanford, A. J., Stewart, A., and Dawydiak, E. (2004). Linguistic focus and good-enough representations: an application of the change-detection paradigm. *Psychon. Bull. Rev.* 11, 882–888. doi: 10.3758/BF03196716
- Therriault, D. J., and Rinck, M. (2007). “Multidimensional situation models,” in *Higher Level Language Processes in the Brain: Inference and Comprehension Processes*, eds F. Schmalhofer and C. A. Perfetti (New Jersey: Lawrence Erlbaum Associates Publishers), 311–327.
- Therriault, D. J., Rinck, M., and Zwaan, R. A. (2006). Assessing the influence of dimensional focus during situation model construction. *Mem. Cogn.* 34, 78–89.
- Trabasso, T., and Sperry, L. L. (1985). Causal relatedness and importance of story events. *J. Mem. Lang.* 24, 595–611. doi: 10.1016/0749-596X(85)90048-8
- Trabasso, T., and Suh, S. (1993). Understanding text: achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Process.* 16, 3–34. doi: 10.1080/01638539309544827
- Whitfield, M., and Jordan, C. H. (2009). Mutual influence of implicit and explicit attitudes. *J. Exp. Soc. Psychol.* 45, 748–759. doi: 10.1016/j.jesp.2009.04.006
- Wilson, T. D., Lindsey, S., and Schooler, T. Y. (2000). A model of dual attitudes. *Psychol. Rev.* 107, 101–126. doi: 10.1037/0033-295X.107.1.101
- Yang, C. L., Zhang, H., Duan, H., and Pan, H. (2019). Linguistic focus promotes the ease of discourse integration processes in reading comprehension: evidence from event-related potentials. *Front. Psychol.* 9:2718.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., and Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychol. Bull.* 133, 273–293. doi: 10.1037/0033-2909.133.2.273
- Zwaan, R. A. (2004). The immersed experiencer: toward an embodied theory of language comprehension. *Psychol. Learn. Motiv.* 44, 35–62. doi: 10.1016/s0079-7421(03)44002-4
- Zwaan, R. A., Langston, M. C., and Graesser, A. C. (1995a). The construction of situation models in narrative comprehension: an event-indexing model. *Psychol. Sci.* 6, 292–297. doi: 10.1111/j.1467-9280.1995.tb00513.x
- Zwaan, R. A., Magliano, J. P., and Graesser, A. C. (1995b). Dimensions of situation model construction in narrative comprehension. *J. Exp. Psychol.* 21, 386–397. doi: 10.1037/0278-7393.21.2.386
- Zwaan, R. A., Stanfield, R. A., and Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychol. Sci.* 13, 168–171. doi: 10.1111/1467-9280.00430
- Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185. doi: 10.1037/0033-2909.123.2.162

Conflict of Interest: 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 © 2020 Horchak and Garrido. 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 Context in Language Development for Children With Autism Spectrum Disorder

Patricia Sánchez Pérez^{1*†}, Anders Nordahl-Hansen^{2†} and Anett Kaale^{1,3†}

¹ Department of Special Needs Education, Faculty of Educational Sciences, University of Oslo, Oslo, Norway, ² Faculty of Education, Østfold University College, Halden, Norway, ³ Norwegian Centre of Expertise for Neurodevelopmental Disorders and Hypersomnias, Oslo University Hospital, Oslo, Norway

OPEN ACCESS

Edited by:

Marcela Pena,
Pontificia Universidad Católica
de Chile, Chile

Reviewed by:

Laura Barca,
National Research Council (CNR), Italy
Ernesto Guerra,
University of Chile, Chile

*Correspondence:

Patricia Sánchez Pérez
sanchezperezpatri@gmail.com

†ORCID:

Patricia Sánchez Pérez
orcid.org/0000-0003-2917-5835
Anders Nordahl-Hansen
orcid.org/0000-0002-6411-3122
Anett Kaale
orcid.org/0000-0002-2680-3384

Specialty section:

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

Received: 20 May 2020

Accepted: 30 November 2020

Published: 23 December 2020

Citation:

Sánchez Pérez P,
Nordahl-Hansen A and Kaale A
(2020) The Role of Context in
Language Development for Children
With Autism Spectrum Disorder.
Front. Psychol. 11:563925.
doi: 10.3389/fpsyg.2020.563925

Parent and preschool teacher ratings of the 10 noun categories of MacArthur-Bates Communication Development Inventory (CDI) were used to study expressive language in 2–4-year-old children with autism spectrum disorder (ASD) ($N = 58$) across the home and preschool context. There was no significant difference in the total number of words the children said in the two contexts, but the children said significantly more words in the noun categories “Furniture and rooms” and “People” at home. Only one third of the words the children said were said both at home and in the preschool, while the other two thirds were said only at home or only in preschool. This suggests that what words the children use across contexts differ substantially and that their vocabulary is larger than it seems when measured only in one context. This novel study highlights the importance of assessing the language in children with ASD in multiple contexts in order to better measure their vocabulary and to design appropriate language interventions.

Keywords: autism spectrum disorder, expressive language, context, home, preschool

INTRODUCTION

There is growing evidence that development of expressive language in children with autism spectrum disorder (ASD) follows a qualitatively similar, but delayed pattern compared to children with typical development (TD) (Charman et al., 2003; Luyster et al., 2007; Weismer et al., 2010). Still, little is known about factors influencing early language in children with ASD. In typical development, variables such as the socioeconomic status of the main caregiver (e.g., Hoff, 2003; Rowe, 2012; Schwab and Lew-Williams, 2016) influence the early language of the child. Also, caregiver’s child-directed speech (e.g., Huttenlocher et al., 2010) and their diverse use of semantic categories (Jones and Rowland, 2017) show a strong impact on early language development. Tamis-LeMonda et al. (2019) observed how objects associated with what the child was doing during different home routines determined the semantic content of the child-directed speech of the mothers. They found that mothers were more likely to use words for toys during play with their child, words for food and utensils while feeding them, and words for body parts while grooming them. Even though decontextualized language emerges in typically developing children around 2 years of age (Uccelli et al., 2019), this might be delayed in children with ASD because these

children often show delays in their general language development (e.g., Tager-Flusberg et al., 2005) and also have difficulties in generalizing from one context to another (Plaisted, 2015).

Most children with ASD attend preschool, and they will encounter many similar, but also some different objects, persons, and places across the home and preschool context. This object variation will influence what words parents and teachers use during their interaction with the child, which may subsequently affect the words used by young children with ASD. The aim of this study was to examine how expressive language of children with ASD might be different in various contexts. First, we compared the amount of words said by children with ASD at home and in preschool across 10 categories listing various types of objects, places, animals and persons that we assumed to be typical for one or the other context. Second, we investigated the degree of overlap between the words reported to be used by the children at home and in preschool. The results of this research may provide valuable insight into the role of context in the development of expressive language in young children with ASD.

METHOD

Participants

Fifty-eight 2–4 year old children ($M = 48.8$ months, $SD = 8$), 47 (81%) boys, with a confirmed ICD-10 diagnosis of childhood autism and their parents and preschool teachers participated in this study (Table 1). Children's cognitive and language skills were assessed with the Mullen Scales of Early Learning (MSEL, Mullen, 1995) and the Norwegian translation of Reynell Developmental Language Scales (RDLS, Hagtvet and Lillestøen, 1985). The children had a mean mental age of 27.9 months ($SD = 11.4$), a mean receptive language age of 23.2 months ($SD = 11.1$), and a mean expressive language age of 21.1 months ($SD = 11.9$). Demographic data and information about the preschools was collected with questionnaires to parents and preschool teachers. Almost half of the mothers (26/46%) and fathers (23/44%) had a higher education degree. Forty-one (72%) were exclusively Norwegian-speaking homes, five (9%) were non-Norwegian-speaking homes and 11 (19%) spoke Norwegian and another language at home. The majority of children were Caucasian (40/69%), while the others had other backgrounds (8/14% Asian, 5/9% other/mixed, and 3/5% African). Most of the children (51/88%) attended a public mainstream preschool, while three (5%) attended a public preschool for children with ASD and four (7%) attended a unit for children with ASD in a public mainstream preschool. Their mean attendance in preschool was 37.3 h per week ($SD = 5.1$). The data used in the present study was a subset of the data collected for the baseline assessment in a previous study (Kaale et al., 2012). The Norwegian National Committee for Research Ethics approved the study. Participants provided written informed consent.

Measures

The Communicative Development Inventories "Words and Gestures" form (CDI-WG; Fenson et al., 1994), completed by

TABLE 1 | Sample characteristics.

	Mean/No. (%)	SD	Range
Children			
Chronological age, overall	48, 8	8, 0	30–60
Mental age ^{1,2} , overall	27, 9	11, 4	9–59
2-year-olds ($n = 6$)	14, 2	4, 4	10–21
3-year-olds ($n = 14$)	26, 1	6, 7	18–43
4-year-olds ($n = 37$)	29, 5	12, 2	9–59
Receptive language age ³ , overall	23, 2	11, 1	6–60
2-year-olds ($n = 6$)	11, 2	7, 3	9–24
3-year-olds ($n = 14$)	22, 8	7, 7	9–36
4-year-olds ($n = 38$)	24, 2	11, 9	6–60
Expressive language age ³ , overall	21, 1	11, 9	3–60
2-year-olds ($n = 6$)	12, 6	6, 5	8–24
3-year-olds ($n = 14$)	19, 4	5, 8	10–30
4-year-olds ($n = 38$)	22, 3	13, 1	3–60
Gender			
Female	11 (19%)		
Male	47 (81%)		
Hours in preschool per week ⁴	37, 3	5, 1	20–45
Parents			
Mother's educational level ⁵			
Primary education	8 (14%)		
Secondary education	22 (40%)		
University/College	26 (46%)		
Father's educational level ⁶			
Primary education	5 (9%)		
Secondary education	24 (45%)		
University/College	23 (44%)		
Language spoken at home ⁷			
Norwegian only	41 (72%)		
Norwegian and another	11 (19%)		
Other than Norwegian	5 (9%)		

¹Mullen Scale of Early Learning (MSLE).

²Missing data from one child.

³Reynell Developmental Language Scale (RDLS); for scores <4 stanine for 1.5 years language age was based on MSLE.

⁴Missing data from two children.

⁵Missing data from two mothers.

⁶Missing data from six fathers.

⁷Missing data from five fathers and one mother.

parents and preschool teachers, were used to measure the words the children said at home and in the preschool. The CDI forms were sent separately to parents and preschool teachers, along with instructions on how to fill them in, and they were collected upon arrival the day of cognitive and language assessments. CDI includes a checklist with 396 vocabulary items across 19 different semantic categories including nouns, verbs, adjectives, pronouns, prepositions and quantifiers. In the present study, only the 10 categories containing exclusively nouns were used (e.g., "Toys," "Clothes," "Rooms and Furniture," and "Small things in the household"). Objects, persons and places listed in most of the categories are equally typical for home and preschool, but some of the noun categories are more typical for one context than the other. Words from the categories "People" (e.g., aunt and babysitter), "Rooms and furniture" (e.g., bedroom and

living room) and “Small things in the household” (e.g., towel, vacuum cleaner) are more typical for the home context. The amount of words varied from eight to 36 in the noun categories, and across the 10 noun categories there were a total of 228 words. Based on the CDI data from parents and preschool teachers we calculated the number of words said by the child at home and in the preschool across the 10 noun categories and for each category. We also calculated “matching” variables defined as the number of words the children said both at home and in the preschool in each of the 10 noun categories. In addition, we calculated the number of “unique words” the children said across the two contexts. This was computed based on the number of words reported by parents plus the number of words reported by preschool teachers minus the number of words reported by both of them (the “matching”). Last, we computed the percentage for the “matching” variables (i.e., number of words reported by parents plus preschool teachers divided by the number of “unique words” multiplied by 100). The CDI has previously shown high concurrent validity with direct assessments (Nordahl-Hansen et al., 2014) and high inter-rater reliability (Nordahl-Hansen et al., 2013) when used with children with ASD.

Statistical Analyses

Along with descriptive statistics, paired sample *t*-tests were performed to compare the differences in the number of words the children were reported to say at home and in the preschool both overall and across the 10 noun categories. In addition, two paired sample *t*-tests were conducted to compare the number of words reported to be said at home and in preschool combined with the number of words reported to be said only by parents and only by preschool teachers, respectively. We also ran descriptive analyses on the “matching” variable (i.e., mean percentages). The software IBM SPSS version 25 and Microsoft Excel 2016 were used to analyze the data.

RESULTS

There was no significant difference in the overall number of words the children said at home and in the preschool ($M_{home} = 78.1$, $SD = 78.4$; $M_{preschool} = 70.5$, $SD = 75.9$, and $p = 0.07$) (Table 2). The same was true for most of the 10 noun categories, except for “Furniture and rooms” ($M_{home} = 7.4$, $SD = 8.8$; $M_{preschool} = 5.4$, $SD = 7.4$, and $p = 0.00$) and “People” ($M_{home} = 6.0$, $SD = 5.5$; $M_{preschool} = 5.0$, $SD = 5.8$, and $p = 0.04$), which were significantly different. Parents and preschool teachers reported that the children said one third of the 228 words listed in the 10 noun categories at home ($M_{home} = 34\%$, $SD = 34\%$) and in the preschool ($M_{preschool} = 31\%$, $SD = 33\%$). The highest percentages of listed words reported to be said by both parents and preschool teachers were in the noun categories “Vehicles (real or toy),” “Food and drinks,” and “Toys.”

After analyzing the “matching” variables, we found that among the words parents and preschool teachers reported that the children said, 38% was said both at home and in the preschool (ranging from 26 to 47% across the 10 categories/ Figure 1). The match was lowest for “Furniture and rooms” (26%) and “Small things in the household” (27%), and highest for “Vehicles (real or toy)” (47%) and “Food and drinks” (44%). For example, the children were reported to say on average 12.2 and 11.1 words in the category “Food and drinks” at home and in the preschool, respectively, but only 44% (5.2) of these words were reported to be said in both contexts. This suggests that the children said 7.0 “Food and drinks” words at home that they did not say in the preschool and 5.9 words in preschool that they did not say at home. This means that the actual number of “Food and drinks” words the children said was closer to 18 (5.2 words said both at home and in preschool + 7.0 words said only at home + 5.9 words said only in preschool = 18.1 words). This pattern was true for all 10 noun categories. In fact, the children said a total of 91, 7 words ($SD = 86, 8$) if adding together words said in both contexts, words said only at home and words said only in preschool.

TABLE 2 | Number of words parents and preschool teachers report that the children ($N = 58$) say at home and in preschool across all 10 noun categories and for each category.

Children Development Inventory (CDI) categories	Home mean (SD)/% ²	Preschool mean (SD)/%	Mean diff. (SD)/%	t(df), p
Across all the 10 noun categories (228) ¹	78.1 (78.4)/34	70.5 (75.9)/31	7.6 (31.5)/3	1.8 (57), $p = 0.07$
Animals (real or toy) (36)	11.6 (12.3)/32	11.5 (12.3)/32	0.1 (6.4)/0	0.3 (57), $p = 0.87$
Vehicles (real or toy) (9)	4.4 (3.6)/49	4.1 (3.8)/46	0.3 (1.6)/3	1.5 (57), $p = 0.15$
Toys (8)	3.2 (3.2)/40	3.3 (3.2)/41	-0.1 (1.4)/-1	-0.5 (57), $p = 0.65$
Food and drinks (30)	12.2 (11.4)/41	11.1 (10.8)/37	1.1 (4.6)/4	1.8 (57), $p = 0.09$
Clothes (19)	6.3 (6.9)/33	6.0 (6.8)/31	0.3 (3.5)/2	0.6 (57), $p = 0.53$
Body parts (20)	7.7 (8.0)/38	7.2 (8.1)/36	0.4 (3.9)/2	0.9 (57), $p = 0.40$
Furniture and rooms (24)	7.4 (8.8)/31	5.4 (7.4)/23	2.0 (4.7)/8	3.2 (57), $p = 0.00$
Small things in the household (36)	11.4 (12.7)/32	9.7 (12.0)/27	1.6 (6.3)/5	2.0 (57), $p = 0.06$
Outside things and places to go (26)	8.1 (9.0)/30	7.2 (8.7)/27	0.9 (4.4)/3	1.5 (57), $p = 0.15$
People (20)	6.0 (5.5)/30	5.0 (5.8)/25	1.0 (3.8)/5	2.1 (57), $p = 0.04$

¹Number in brackets in this column indicates total amount of words in that semantic category.

²Percentage of total possible words the parents/preschool teachers report that the child say (i.e., number of words reported \times 100/total number of words in the category). The values in bold represent statistical significance.

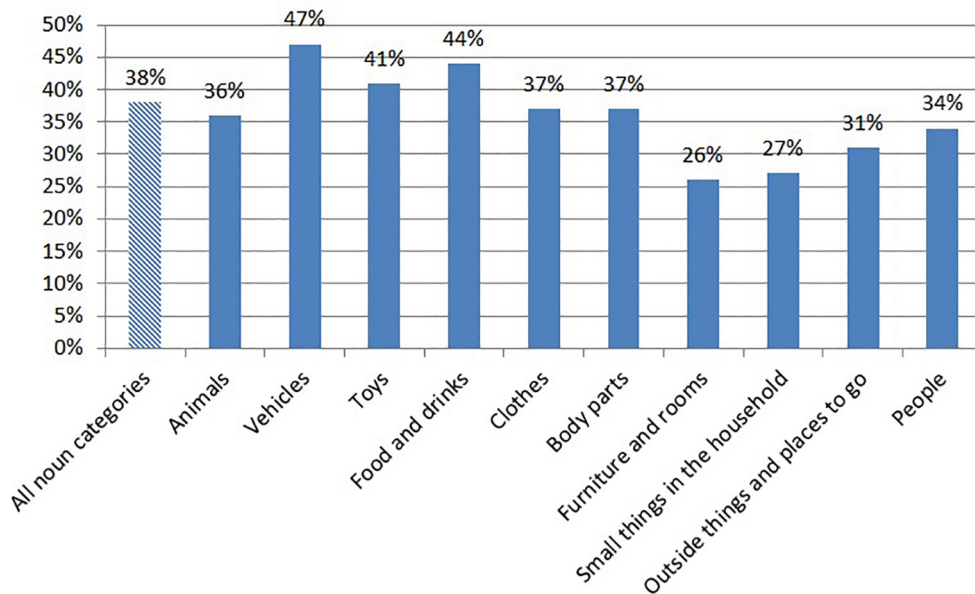


FIGURE 1 | Percentages of words said both at home and in the preschool (match) across all 10 noun categories and for each category.

There was a significant difference between the total number of words reported at home and preschool combined and both the number of words reported only by parents [$M_{\text{difference}} = 13.5$, $SD = 16.5$, $t(57) = 6.23$, and $p < 0.000$] and the number of words reported only by preschool teachers [$M_{\text{difference}} = 21.1$, $SD = 27.4$, $t(57) = 5.89$, and $p < 0.000$]. Thus, the children's vocabulary was larger than what was captured by relying on reports only from parents or preschool teachers.

DISCUSSION

This study aimed to investigate how expressive language of young children with ASD might be different in various contexts using CDI completed by parent and preschool teachers. We found that the overall number of words said at home and in the preschool was quite similar. Both parents and preschool teachers reported that the children said most words in the categories "Toys," "Vehicles (real or toy)," and "Food and drinks." This could be explained by the fact that play and food are two important areas in the first years of life, and objects, and therefore words related to these areas, are probably of high frequency in the learning contexts of these children. We also found that the children used more words from the "Furniture and Rooms" and "People" categories at home compared to preschool. This finding is in line with our expectations as we assumed that objects belonging to these two categories are more prevalent at home, and thus will affect the caregiver's child-directed speech, which next supports the children's use of these words (Huttenlocher et al., 2010). Interestingly, we found that only one third of the words the children said were used both

at home and in the preschool and two thirds only in one or the other context. This suggests that language during the first 4 years of life in children with ASD is context-dependent, which is similar to what is found in younger children with typical development (Uccelli et al., 2019). This interpretation is further supported by the fact that the highest proportions of words said across both contexts were in the categories "Toys," "Food and Drinks," and "Vehicles (real or toy)," which are essential objects both at home and in the preschool. This supports Hills et al. (2009) notion of "preferential acquisition" as the working principle behind language development: words that are better interconnected in the learning context, rather than in the child's internal semantic network, are learned earlier in development. When adding together all the words the children were reported to say both at home and in preschool, only at home and only in preschool, we found that the children vocabulary was significantly larger compared to measuring their vocabulary only at home or only in preschool. This suggests that the vocabulary of young children with ASD may be larger than what is revealed by investigating only one context, which is currently the most common way to collect information about everyday language of children with ASF. In a previous study, we reported high interrater reliability between parent and preschool teacher reports (Nordahl-Hansen et al., 2013), but the reliability was then calculated using total amount of words reported by parents and teachers, not the actual words the children say across the home and preschool contexts. The finding of the present study suggests that a cumulative CDI score from multiple sources such as parents and preschool teachers combined will give a better picture of children's language abilities.

Limitations

The expressive language of children with ASD has previously been studied with the CDI (e.g., Charman et al., 2003), but this is the first study to use this instrument to measure expressive language across contexts by comparing data from two sources for each child. Still, the study has some limitations. First, we did not include a comparison group of children with typical development or children with developmental language disorder. Therefore we do not know if the findings are unique to children with ASD, although based on previous research (Charman et al., 2003; Luyster et al., 2007; Luyster and Arunachalam, 2018) we expect that the same pattern will be evident in younger children with TD. Second, it might be that parents and preschool teachers are more prone to report that children say words related to objects in their environment, and that the findings do not reflect the words the children say, but a recall bias in parents and teachers. A more narrow age range among the study participants (e.g., only 4-year-olds versus 2–4-year-olds as in the present study) might have helped making the results more precise. However, as children with ASD are known to be very heterogeneous regarding their language development recruiting only 4-year-olds would pose some of the same challenges compared to TD samples. Also, although the CDI has shown high concurrent validity compared to both direct structured tests and language samples (Dale, 1991; Nordahl-Hansen et al., 2014), observational data of factual word use in both contexts would have strengthened the findings. Last, the findings are based on parents and teachers reports of the words the children produce only at one time point so we do not know how stable the results are.

Future Directions

Future research should include a comparison group of children with typical development who are matched both on chronological and mental age. Other interesting aspects would be to investigate which specific words within the 10 categories are typically used in one of these two contexts but not in the other (there might be even a pattern), the influence of the language spoken at home and the use of words across the two contexts in subgroups of children with ASD (e.g., those who are just starting to speak and those with a more developed vocabulary).

REFERENCES

- Charman, T., Drew, A., Baird, C., and Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *J. Child Language* 30, 213–236. doi: 10.1017/s0305000902005482
- Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *J. Speech Language Hearing Res.* 34, 565–571. doi: 10.1044/jshr.3403.565
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., et al. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, 1–173. doi: 10.4324/9781315111322-1
- Hagtvet, B., and Lillestøen, R. (1985). Reynell språktest.
- Hills, T. T., Maouene, M., Maouene, J., Sheya, A., and Smith, L. (2009). Categorical structure among shared features in networks of early-learned nouns. *Cognition* 112, 381–396. doi: 10.1016/j.cognition.2009.06.002

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Not available due to restrictions related to ethical regulations. Requests to access these datasets should be directed to AK, anett.kaale@isp.uio.no.

ETHICS STATEMENT

This study was reviewed and approved by the Norwegian National Committee for Research Ethics. Written informed consent was provided by parents and preschool teachers.

AUTHOR CONTRIBUTIONS

The data used in this study was originally gathered for a RCT lead by AK. All authors contributed in writing the manuscript and took part in the design and analyses of this study.

FUNDING

This research was supported by grants (2005069 and 2016066) from South-Eastern Norway Regional Health Authority, Regional Centre for Child and Adolescent Mental Health, East and South Norway, Regional Resource Centre for Autism, AD/HD, Tourette Syndrome and Narcolepsy at Oslo University Hospital, Norwegian Centre of Expertise for Neurodevelopmental Disorders and Hypersomnias at Oslo University Hospital, and the research group Communicative Processes (ComPros) at University of Oslo.

ACKNOWLEDGMENTS

The authors would like to thank the participating parents, children, and preschool teachers for their time and willingness. The authors gratefully acknowledge research assistants who helped to collecting the data.

- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Dev.* 74, 1368–1378. doi: 10.1111/1467-8624.00612
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., and Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognit. Psychol.* 61, 343–365. doi: 10.1016/j.cogpsych.2010.08.002
- Jones, G., and Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognit. Psychol.* 98, 1–21. doi: 10.1016/j.cogpsych.2017.07.002
- Kaale, A., Smith, L., and Sponheim, E. (2012). A randomized controlled trial of preschool-based joint attention intervention for children with autism. *J. Child Psychol. Psychiatr.* 53, 97–105. doi: 10.1111/j.1469-7610.2011.02450.x
- Luyster, R. J., and Arunachalam, S. (2018). Brief Report: Learning Language Through Overhearing in Children with ASD. *J. Autism Dev. Disord.* 50, 2616–2624. doi: 10.1007/s10803-018-3672-0

- Luyster, R., Qiu, S., Lopez, K., and Lord, C. (2007). Predicting Outcomes of Children Referred for Autism Using the MacArthur-Bates Communicative Development Inventory. *J. Speech Language Hearing Res.* 50:667. doi: 10.1044/1092-4388(2007/047)
- Mullen, E. M. (1995). *Mullen scales of early learning*. Minnesota, MN: AGS Circle Pines.
- Nordahl-Hansen, A., Kaale, A., and Ulvund, S. E. (2013). Inter-rater reliability of parent and preschool teacher ratings of language in children with autism. *Res. Autism Spectr. Disord.* 7, 1391–1396. doi: 10.1016/j.rasd.2013.08.006
- Nordahl-Hansen, A., Kaale, A., and Ulvund, S. E. (2014). Language assessment in children with autism spectrum disorder: Concurrent validity between report-based assessments and direct tests. *Res. Autism Spectr. Disord.* 8, 1100–1106. doi: 10.1016/j.rasd.2014.05.017
- Plaisted, K. C. (2015). Reduced generalization in autism: an alternative to weak central coherence. Available online at: <https://www.repository.cam.ac.uk/handle/1810/248652>
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Dev.* 83, 1762–1774. doi: 10.1111/j.1467-8624.2012.01805.x
- Schwab, J. F., and Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. *Wiley Interdiscipl. Rev.* 7, 264–275. doi: 10.1002/wcs.1393
- Tager-Flusberg, H., Paul, R., and Lord, C. (2005). Language and communication in autism. *Handbook Autism Pervasive Dev. Disord.* 1, 335–364.
- Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., and Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child Dev.* 90, 2135–2152. doi: 10.1111/cdev.13089
- Uccelli, P., Demir-Lira, ÖE., Rowe, M. L., Levine, S., and Goldin-Meadow, S. (2019). Children's early decontextualized talk predicts academic language proficiency in midadolescence. *Child Dev.* 90, 1650–1663. doi: 10.1111/cdev.13034
- Weismer, S. E., Lord, C., and Esler, A. (2010). Early Language Patterns of Toddlers on the Autism Spectrum Compared to Toddlers with Developmental Delay. *J. Autism Dev. Disord.* 40, 1259–1273. doi: 10.1007/s10803-010-0983

Conflict of Interest: 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 © 2020 Sánchez Pérez, Nordahl-Hansen and Kaale. 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.



Do the Eyes Have It? A Systematic Review on the Role of Eye Gaze in Infant Language Development

Melis Çetinçelik^{1*}, Caroline F. Rowland^{1,2} and Tineke M. Snijders^{1,2}

¹ Language Development Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, ² Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

OPEN ACCESS

Edited by:

Ramesh Kumar Mishra,
University of Hyderabad, India

Reviewed by:

Nicole Altwater-Mackensen,
Johannes Gutenberg University
Mainz, Germany
Marilyn Vihman,
University of York, United Kingdom

*Correspondence:

Melis Çetinçelik
melis.cetincelik@mpi.nl

Specialty section:

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

Received: 30 July 2020

Accepted: 25 November 2020

Published: 08 January 2021

Citation:

Çetinçelik M, Rowland CF and
Snijders TM (2021) Do the Eyes Have
It? A Systematic Review on the Role
of Eye Gaze in Infant Language
Development.
Front. Psychol. 11:589096.
doi: 10.3389/fpsyg.2020.589096

Eye gaze is a ubiquitous cue in child–caregiver interactions, and infants are highly attentive to eye gaze from very early on. However, the question of why infants show gaze-sensitive behavior, and what role this sensitivity to gaze plays in their language development, is not yet well-understood. To gain a better understanding of the role of eye gaze in infants' language learning, we conducted a broad systematic review of the developmental literature for all studies that investigate the role of eye gaze in infants' language development. Across 77 peer-reviewed articles containing data from typically developing human infants (0–24 months) in the domain of language development, we identified two broad themes. The first tracked the effect of eye gaze on four developmental domains: (1) vocabulary development, (2) word–object mapping, (3) object processing, and (4) speech processing. Overall, there is considerable evidence that infants learn more about objects and are more likely to form word–object mappings in the presence of eye gaze cues, both of which are necessary for learning words. In addition, there is good evidence for longitudinal relationships between infants' gaze following abilities and later receptive and expressive vocabulary. However, many domains (e.g., speech processing) are understudied; further work is needed to decide whether gaze effects are specific to tasks, such as word–object mapping or whether they reflect a general learning enhancement mechanism. The second theme explored the reasons why eye gaze might be facilitative for learning, addressing the question of whether eye gaze is treated by infants as a specialized socio-cognitive cue. We concluded that the balance of evidence supports the idea that eye gaze facilitates infants' learning by enhancing their arousal, memory, and attentional capacities to a greater extent than other low-level attentional cues. However, as yet, there are too few studies that directly compare the effect of eye gaze cues and non-social, attentional cues for strong conclusions to be drawn. We also suggest that there might be a developmental effect, with eye gaze, over the course of the first 2 years of life, developing into a truly ostensive cue that enhances language learning across the board.

Keywords: eye contact, gaze following, language development, word acquisition, object processing

INTRODUCTION

Social interaction plays a critical role in language acquisition. Children typically learn language through face-to-face interactions with their caregivers in social contexts, and face-to-face communication is inherently multimodal. The communicating social partners exchange a variety of information beyond the verbal domain, using facial expressions, gestures, and most pertinently for the present paper, eye gaze.

Eye gaze is, in fact, a central element in human communication. Gaze cues during a communicative interaction can indicate social engagement, reflect a desire to communicate, reveal the speaker, and the listener's goals and feelings, and can direct the attention of the listener to objects in the environment (Kleinke, 1986). Eye gaze can act as an ostensive cue to a speaker's intent, by specifying the addressee of the communication and signaling that the accompanying actions are communicative and meaningful rather than random acts (Csibra, 2010). This last function is especially crucial for human infants, since their limited knowledge of language means that they cannot rely on the semantic context of the speech signal to understand that communication is directed toward them. Rather, they can infer that the social partner (i.e., the adult) is addressing them by social signals in communication, such as eye gaze, infant-directed speech, and calling the infant's name (Csibra and Gergely, 2009). The current review focuses specifically on the role of eye gaze in infant language development, over and above other social cues.

Not only do adults often use such social cues when communicating with infants, but infants also show a sensitivity to, and preference for, these signals from early on. Infants display a sensitivity to eye gaze in at least two distinct ways. First, they engage in mutual eye contact with their social partner. Newborns look longer at faces with open eyes than faces with closed eyes (Batki et al., 2000). They also prefer faces with direct gaze with which they can engage in mutual eye contact, as evidenced by their preference for direct gaze only for upright and not for inverted faces (Farroni et al., 2002, 2004). This ability to detect and engage in mutual eye contact in live socially interactive settings develops further over the first 4 months of life (Vecera and Johnson, 1995). Second, infants learn to follow an interlocutor's gaze. Infants begin developing gaze-following abilities between 2 and 4 months, which become fairly stable by 6–8 months (D'Entremont et al., 1997; Gredeback et al., 2010). Orienting to gaze cues becomes almost automatic, with adult's gaze direction causing fast visual attention shifts even in infants as young as 3 months of age (Hood et al., 1998).

However, it is not yet clear to what extent infants' sensitivity to social eye gaze has a function beyond basic perception/attention. In particular, it is not yet clear whether, and in what ways, it also facilitates infants' learning in cognitive domains, such as language. There are (at least) two reasons to expect that a sensitivity to eye gaze might facilitate language development. Acting as an ostensive cue, mutual eye gaze (i.e., eye contact) can convey the communicative intent of the caregiver and can put infants in a highly receptive state for accompanying

or upcoming information (Csibra and Gergely, 2009). This is the role of eye gaze according to natural pedagogy theory (Csibra and Gergely, 2009), which holds that ostensive cues, such as eye gaze have a special status in human ontogeny. On this theory, human communication creates opportunities for a transfer of knowledge between a sender and a receiver (caregiver and infant, in this case), and these opportunities are marked by an abundance of ostensive cues, such as eye gaze. Human infants are argued to be innately specified to be sensitive to such cues, such that the presence of those cues puts them in a highly receptive state for upcoming or accompanying stimuli.

Yet, eye gaze could also act as a more basic, simple, attentional cue. In particular, sensitivity to eye gaze could allow infants to optimize the use of limited attentional resources, by directing attention to only those parts of the environment in which the other partner is interested (Niedzwiecka et al., 2018). This could be in the form of mutual gaze or gaze following. Mutual gaze draws infants' attention to the social partner and presumably to the speech signal provided by them. Gaze following directs their attention to a target location in the environment, which facilitates the learning of object properties and their names (Wu et al., 2014). On this view, eye gaze has no special social status but is simply an attentional cue. It may not be different than other low-level cues, such as movement, which equally attract infants' attention. Eye gaze provides learning opportunities for infants through attention modulation rather than serving a special communicative purpose (Szufnarowska et al., 2014).

A related issue concerns the types of language learning tasks that are facilitated by eye gaze. Most studies to date have focused on the role of gaze following in learning about objects in the environment, studies in which children follow the gaze of an interlocutor toward an object and which then test whether gaze following facilitates the encoding of object properties or object-word mappings. However, it is possible that eye gaze might have a more general learning enhancement function, as specified by the natural pedagogy theory. In this case, we might expect eye gaze to have a facilitatory effect on other language tasks (e.g., learning to process speech).

The goal of this study was thus to systematically review the literature on the role of eye gaze in early language learning in the first 2 years of life (infancy). Given the focus on infancy, the scope of the review is mainly restricted to vocabulary development (eye gaze may play a role in other areas, such as grammar and pragmatics, but these develop later in childhood). The literature searching process identified relevant work not only on vocabulary development itself but also in three subdomains that are crucial for the development of vocabulary: word-referent mapping (labeling), object processing, and speech processing. We summarize work in all four domains below, before turning to the question of why eye gaze may facilitate language learning. In particular, we discuss whether there is evidence that eye gaze is a highly specialized socio-cognitive cue that puts infants in a highly charged receptive learning state, as specified by the natural pedagogy theory, or whether it is simply a highly effective attentional cue.

METHOD

We searched the PsycInfo, PubMed, Scopus, and Web of Science databases from the beginning of database records until January 2019. In order to capture the existing literature, we used broad search terms, *infan** AND (eye contact OR gaze) AND (attention OR learning), within the title or the abstract. We focused on “attention” and “learning,” rather than narrowing down to language, because we wanted to include, at this first step, papers that assessed the role of eye gaze in aspects of cognitive development that were relevant to language learning, such as the ability to learn to identify objects in the environment, crucial for object labeling. This search yielded 2,061 papers in total, which was reduced to 1,405 entries after duplicate removal.

We then narrowed the search to the following inclusion criteria: (1) peer-reviewed articles written in English, which (2) study typically developing human infants between 0 and 24 months of age, and (3) present data for a group of participants (that meet the second inclusion criteria) in the domain of cognitive development/learning. We excluded papers that simply documented the development of infant eye gaze behavior without addressing the effect of such abilities on learning, papers that investigated the role of eye gaze in socio-emotional development (e.g., understanding of facial emotional expression) and motor development, and papers that investigated the role of different types of cue on infant sensitivity to eye gaze (e.g., infant temperament, maternal depression) where such studies did not also include an element of learning or processing. We also excluded papers investigating eye gaze behaviors in children with autism or other developmental disabilities, as these focused on different questions (e.g., how to characterize the socio-cognitive abilities of children with autism).

We first screened the 1,405 entries based on their titles and abstracts with regards to the inclusion criteria. We identified 91 papers as eligible for full-text review. An additional seven papers were identified through hand searching the reference lists of the retained articles and were added to the review, resulting in 98 papers. In the second stage, we retrieved the full text of each paper and reviewed them for inclusion, which resulted in 77 papers included in the final review. **Figure 1** illustrates the literature search process.

RESULTS

The papers identified in the review can be broadly divided into two main themes. The first set of papers documented the effect of eye gaze on four developmental domains: the effect on (1) vocabulary development in general, and then, its effect in three subdomains that are crucial for the development of vocabulary; (2) word-referent mapping (labeling); (3) object processing; and (4) speech processing. The second set discussed theories that aim to explain why eye gaze might be facilitative for learning. We discuss each here in turn. **Supplementary Table 1** summarizes the main information of all studies included.

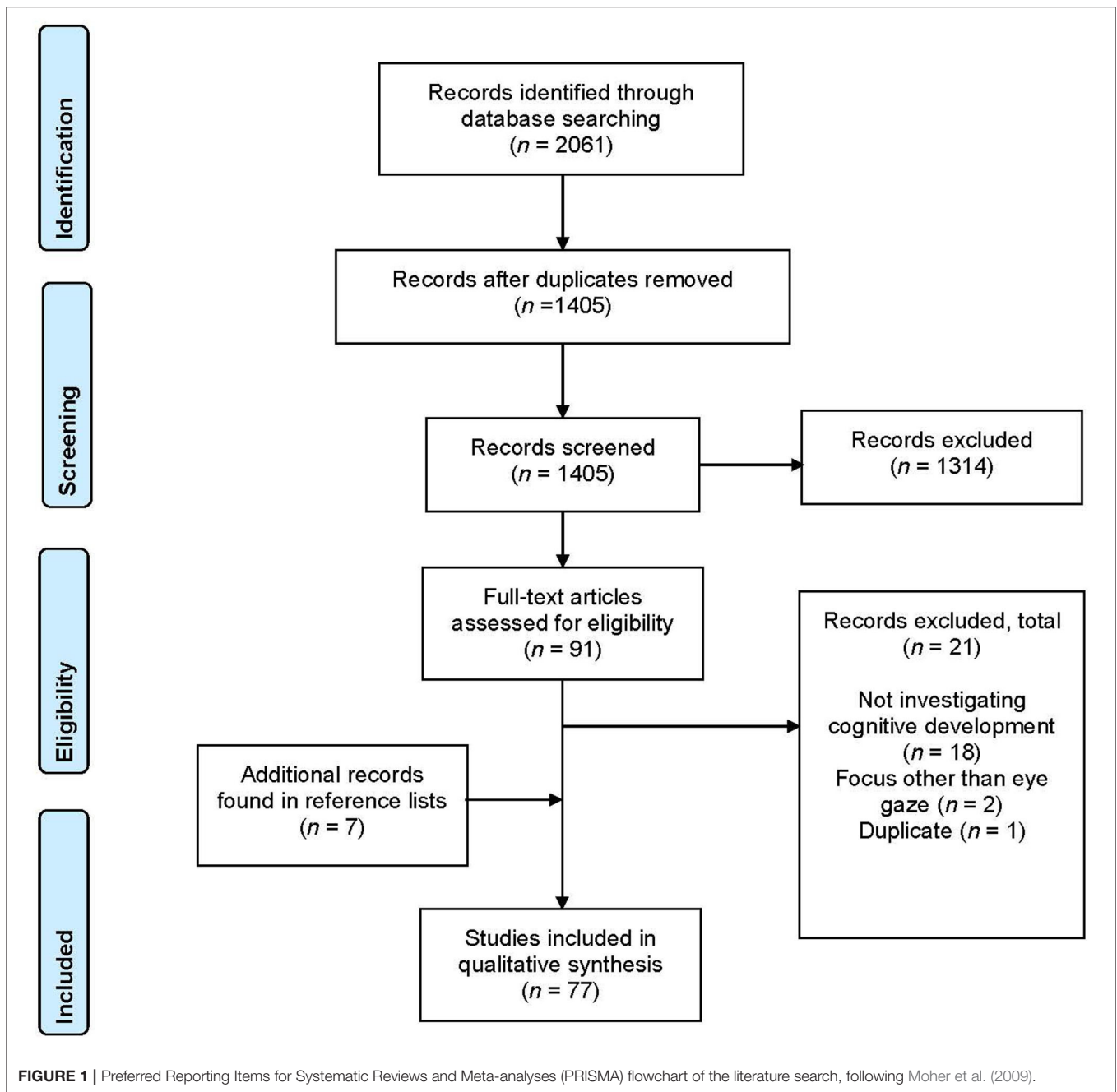
Effects of Eye Gaze on Infant Learning Vocabulary Development: Do Infants With Better Joint Attention Skills Later Develop Better Vocabulary?

Joint attention refers to the caregiver–infant dyads’ shared attention to an object or an event while maintaining their attention to each other. It encompasses a set of socio-cognitive skills that develop in the first year of life, such as pointing, referencing, and gaze following. Gaze following is of utmost importance in studying joint attention in the context of language development, as most of the studies that investigated links between joint attention and vocabulary development use children’s gaze direction and following as a measure of the understanding of the shared (joint) attentional focus (Akhtar and Gernsbacher, 2007). Joint attention involves establishing mutual eye contact with the social partner and then alternating gaze between the partner and the object (either in the presence or absence of other cues, such as pointing or verbal referencing).

Gaze following is an important social skill that develops in the first year of life. Following the gaze of an interactional partner enables infants to engage in joint attentional periods with that partner, and this provides potential learning opportunities for the young language learner. By engaging in joint attention with the adult and following their gaze, infants can selectively attend to a single source (e.g., an object). Hereby, they can direct their attentional resources (which are otherwise naturally limited) to the properties of that object, allowing them to disambiguate the speaker’s likely referent. Thus, the ability to follow a partner’s gaze, and engage in joint attention, is likely to promote vocabulary development.

Our review identified 17 studies that link gaze following and vocabulary development. In most of these studies, children’s understanding of the shared (joint) attentional focus was measured by testing whether the child followed the adult gaze to a target location or object when the adult turned toward the target and then whether they alternated their gaze between the adult and the target. To do this, many studies used standardized measures, such as the Early Social Communication Scales (Mundy et al., 2003a) that assess children’s non-verbal communication skills, including initiating and responding to joint attention. Subsequent vocabulary growth was assessed by parental reports using standardized language tests (such as the MacArthur-Bates Communicative Developmental Inventories; Fenson et al., 2007) at later ages, with the respective subtests for comprehensive and productive vocabulary, and in some studies, complemented with laboratory observation data using standardized language measures. As far as possible, we discuss relations with receptive and expressive vocabulary separately, but there is some overlap here, as the reviewed papers often test both and do not always distinguish between them.

Most of the studies reviewed identified positive links between joint attentional periods between infants and parents and infants’ later vocabulary development (Carpenter et al., 1998; Morales et al., 1998; Brooks and Meltzoff, 2005, 2015; Beuker et al., 2013). In terms of receptive vocabulary, individual differences in responding to joint attention, indexed by infants’ gaze following



scores, were found to be meaningful at as early as 6 months and contributed to differences in receptive vocabulary scores at 12 months (Morales et al., 2000b). Gaze-following behavior at 10–11 months predicted receptive vocabulary at both 14 and 18 months (Brooks and Meltzoff, 2005). Further, full-term infants' responsiveness to gaze alternations in triadic interactions at 9 months and initiating triadic interactions at 14 months were positively correlated with later language, such that infants with more responsivity to gaze shifts had better receptive and expressive language scores at 30 months (De Schuymer et al., 2011). One study, however, did not find a link between infants'

responding to joint attention skills at 12 months and their receptive vocabulary at 12 and 18 months, although it did report correlations with expressive vocabulary at 18, 21, and 24 months (Markus et al., 2000).

Many studies also found infants' joint attentional skills to be predictive of expressive vocabulary when tested at later ages. Studies have reported that (1) individual differences in infants' gaze following abilities at 6 months were positively linked to their receptive vocabulary at 12 months and subsequent expressive vocabulary at 18, 21, and 24 months (Morales et al., 1998); (2) infants' responding to joint attention at 6, 8, 10, 12, and

18 months was positively correlated with expressive vocabulary size at 30 months (although individual differences at 21 and 24 months did not predict language development; Morales et al., 2000a); (3) individual differences in responding to joint attention at 9 and 12 months and initiating joint attention at 18 months predicted 24-months expressive language (Mundy et al., 2007); (4) responding to joint attention at 14 months predicted 24-months expressive language, when controlling for general cognitive development (Mundy et al., 2003b); and that (5) responding to joint attention at 16 months was related to receptive language at the time of test and both later receptive and expressive vocabulary (Mundy and Gomes, 1998). Brooks and Meltzoff reported that infants with higher gaze-following scores at 10.5 months were able to produce more mental-state words at 2.5 years, reflecting an effect both on their vocabulary and theory of mind capacities (Brooks and Meltzoff, 2015). One study did not report a significant link: Morales et al. (2000b) did not find links between gaze-following scores at 6 months and expressive vocabulary at 12 months, but note that this is perhaps not surprising since there is very little variation in expressive vocabulary at 12 months. Overall, then, we conclude that the literature overwhelmingly supports the idea of a meaningful relationship between joint attentional abilities and receptive and expressive vocabulary development in infancy.

Interestingly, the findings of one study (Tenenbaum et al., 2015) suggest that eye gaze cues might enhance learning by focusing infants' attention to linguistically relevant information as well as to their referents in the environment. This study tested 12-months-old infants' gaze following to an object at the point at which the adult was describing the object, as well as their looking toward the speaker's mouth, rather than simply testing gaze following ability *per se*. Infants' gaze performance in this task predicted later expressive vocabulary at 18 and 24 months. These findings are in line with the literature suggesting a developmental shift in infants' attention from the speaker's eyes to their mouth between 4 and 8 months, with a shift back to eyes starting to emerge around 12 months (Lewkowicz and Hansen-Tift, 2012).

On a related note, some studies suggest that later language development is better predicted when we consider multiple pre-linguistic communicative behaviors together, rather than focusing only on gaze following. Brooks and Meltzoff (2008), using growth curve modeling, reported that, together, pointing, gaze following, and looking time (duration of looks at the target object) at 10–11 months predicted more of the variance in the speed of vocabulary growth at 2 years than the individual predictors alone. Importantly, infants' pointing and gaze-following scores did not correlate significantly with each other, suggesting that they are tapping different abilities (but also see Carpenter et al., 1998). This finding suggests that different pre-linguistic abilities may serve different functions; for example, while pointing helps infants initiate triadic attention with the parent, following the parent's gaze may enable infants to understand the referent of words. In a similar vein, D'Odorico et al. (1997) reported that the coordinated development of two communicative behaviors between 12 and 20 months, gaze and vocalizations, was a sign of conversational competence, which then predicted language production at 20 months. Although

care must be taken when interpreting this result as the sample size was small ($N = 13$), the findings are supported by recent studies that suggest a key role of infant vocalizations in the pre-linguistic period (McGillion et al., 2017; Donnellan et al., 2020). Interestingly, coordinated gaze-vocalization behavior at 11 months, which may signal infants' communicative intent, was the strongest predictor of expressive vocabulary at multiple time points up to 24 months (Donnellan et al., 2020).

However, while many studies reported positive relationships between gaze following or gaze alternation skills and language development, some have suggested that these effects are mediated by other factors. For instance, a recent study measuring both parent–infant joint attention and infant sustained attention in naturalistic interactions found that both joint attention and infant sustained attention at 9 months predicted language development at 12 and 15 months, but joint attention by itself was not a strong predictor (Yu et al., 2019). Joint attention periods that did not coincide with sustained infant attention to the target object did not predict later vocabulary scores, while infant sustained attention to the object did, regardless of the joint attentional state at the time of the naming event (Yu et al., 2019). Further, one study did not find any links between joint attention abilities, indexed by gaze alternation between the adult and object, at 20 months and later language, although it reported associations with theory of mind abilities (note though that the sample size was low: 13 infants; Charman et al., 2000).

In summary, with a few exceptions, the results generally support a positive association between joint attention skills and vocabulary development. However, it is not possible to directly establish cause and effect from such studies, since they are observational, and in the main, correlational. In the next three sections, we review the evidence from three subdomains linked to vocabulary acquisition that might help explain why there is such a positive association: evidence that eye gaze (a) facilitates the learning of word–object mappings, (b) facilitates the learning of object properties, and (c) facilitates speech processing.

Word–Object Mapping

In joint attentional contexts, adult and infant attend to the same object while maintaining shared attention with each other, established by mutual gaze (Carpenter et al., 1998). In such instances, eye gaze can direct infants' attention to a specific object in their environment, thereby creating an ideal window for them to learn more about properties of that object, including the label used to refer to it. Hence, gaze following can be a reliable strategy for infants to map words onto objects. In this section, we review evidence for the role of gaze (following) in forming these word–object mappings, based on the 12 papers identified by our systematic review.

As early as 6–8 months of age, the frequency of infants' gaze switches between their mother and an object that occur just at the point at which the mother names an object can affect word learning; two studies reported that infants who switched their gaze frequently were more likely to learn word–object mappings in naturalistic interactions (Gogate et al., 2006; Matatyaho and Gogate, 2008). However, evidence for a sophisticated ability to use gaze cues to form word–object associations does not seem to

appear until the second year of life. In particular, the evidence suggests that it is only toward the end of the second year that infants can use gaze cues to map labels onto the object in situations where perceptual salience cues conflict with social (eye gaze) cues (Moore et al., 1999). For instance, Moore et al. (1999) demonstrated that, when presented with a novel label, 24-months-old, but not 18-months-old, picked the toy to which the adult had directed their gaze during familiarization. This was true even when the saliency of the other object was higher (saliency was manipulated by illuminating and rotating one of the two possible referents of the label). Eighteen-months-old infants only correctly matched the object to its label when both referential and saliency cues reference the same object (Moore et al., 1999). Similarly, Hollich et al. (2000) reported that 24-months-old infants chose to follow adult's gaze direction to map words to objects, rather than using other salient but non-referential cues.

It is possible that younger children can use eye gaze cues to learn word-object associations but only in certain situations. In one study, infants aged between 12 and 18 months followed the gaze of a stranger as well as that of their caregiver but only formed word-object associations when following the caregiver's gaze (Barry-Anwar et al., 2017), although note that in other studies, infants of this age did learn from a stranger (e.g., Moore et al., 1999; Hirotani et al., 2009). The "social" nature of the agent also seems to be important for younger children to establish word-object mappings; 18-months-old learned the names of novel objects when the labels were provided by a human, but not a robot, although they did follow the robot's gaze (O'Connell et al., 2009). Even older infants' ability to learn from eye gaze cues can be derailed. For example, Graham et al. (2010) reported that 24-months-old infants' learning from gaze cues was also affected by "default" biases, such as mutual exclusivity (the assumption that a single object has only one label), with infants relying on mutual exclusivity when gaze cues offered conflicting information. Optimal learning only occurred when eye gaze and mutual exclusivity provided coinciding information (Graham et al., 2010).

There also seems to be some evidence that infants only treat gaze cues as referential cues (i.e., cues to object labels rather than simply low-level attentional cues) from about 24 months of age. Graham et al. (2011) tested 24-months-old infants' word learning both when the experimenter's gaze cued the location of the object and when the gaze cued the object itself. In the location condition, the experimenter looked at the target location, provided the label, and then placed the object in that location. In the object condition, the object was presented in the target location before the label was given. The results showed that, although infants followed gaze equally well in both conditions, they formed word-object mappings only when the experimenter's gaze cued the object already in the location (the object condition). The authors argued that this was because the infants treated gaze as a signal of referential intent; they expected to find an object to which the word can be mapped in the cued-at location and only learned the label in the condition in which this expectation was satisfied. Interestingly, one study suggests that even infants' preferred social cue preferences may change with age: Paulus and Fikkert (2014) reported that young infants (14-months-old) relied more

on eye gaze cues when learning word-object mappings, but adults and older infants (24-months-old) relied more on pointing cues.

In summary, it seems that eye gaze influences infants' learning of word-object pairings. There exists a clear developmental trend in increasing sophistication over the first 2 years of life, supported by infants' developing attention, memory, and information processing capacities (Yurovsky and Frank, 2017). However, it is not always clear from such studies whether it is eye gaze *per se* that is driving the effect because it is usually difficult to disentangle eye gaze and other joint attentional cues. For instance, Hirotani et al. (2009) used an event-related potential (ERP) paradigm to investigate the effects of joint attention on infants' word learning at 18–21 months. Infants were taught novel word-object combinations in either a joint attention context (eye contact, positive tone of voice) or non-joint attention context (eye contact averted, neutral tone of voice). For words learned in the joint attention context, incongruent object-word pairs resulted in a late negativity, similar to a N400 effect, reflecting a failure in semantic integration. However, as both eye gaze and other social cues, such as the speaker's tone of voice were jointly manipulated, it is difficult to conclude which cue drove the effect.

There is one study, however, that provides evidence that infants can use eye gaze cues under more tightly controlled experimental settings. Houston-Price et al. (2006) used prerecorded videos of an experimenter turning her head (with gaze following) to one of the two objects placed on her right and left, while a label was provided over the loudspeaker. The use of a prerecorded video allowed the authors to control for the presence of other (covert) joint attentional cues. Fifteen-months-old successfully learned the word-object pairings in this context. Although this is only one study that needs to be replicated, such findings in controlled settings suggest that learning can indeed be driven by the presence of the intended social cue (gaze direction) and is not a result of additional, covert social cues that might occur when infants are interacting with a live experimenter (Houston-Price et al., 2006).

Object Processing

Another explanation for why infants' sensitivity to gaze cues might affect vocabulary development concerns the role of gaze cues in object processing. Given that a necessary precursor to learning to map words onto objects is learning to identify objects themselves, it may be that eye gaze facilitates object processing directly, which then indirectly facilitates word learning. By 4 months, infants start to follow an adult's gaze to a location and use this cue to switch their attention to that location. It has been suggested that this facilitates processing of the properties of the target object rather than other competing stimuli in the environment (Reid and Striano, 2005). In other words, infants not only are more likely to pay attention to a specific object as a result of the adult's gaze direction but also will be more likely to process that object and thus learn more about it. This facilitatory effect is likely to contribute to infant's ability to form word-object mappings by enhancing successful encoding of object properties, to which words are then mapped, and thus language development. In fact, object processing has been found to be a

mediator of the relationship between gaze following and later vocabulary development (Okumura et al., 2017).

Our systematic review identified a large number of studies ($n = 28$) demonstrating a role for eye gaze cueing in infants' object processing. Many of these ($n = 14$) used variations on the behavioral novelty preference paradigm. In this paradigm, infants are first familiarized with novel objects using simple, prerecorded images or videos that either depict a person looking toward (cued) or away from the object (uncued) or that depict a person first establishing eye contact and then turning their head and gaze to one of two objects placed on either side of the face. In a subsequent test phase, the object(s) are presented without the face, and infant looking times to the object(s) are measured. Differences in infants' novelty preference in the test phase is taken to reflect differential processing of cued and uncued objects in the familiarization phase, possibly due to more attention to the cued object. The assumption here is that the previously uncued object will be perceived as more novel, thus resulting in longer looking times due to infants' novelty preference. In general, such studies have reported the expected novelty preference for the uncued object (see, e.g., Reid and Striano, 2005, who found this effect with 4-months-old). Similarly, social cues, such as the face and eyes turning toward the target object have been shown to enable 9-months-old infants in making inferences about object properties, even when distracting cues are present (Wu et al., 2011).

In an experimental setting using live joint attention interactions, joint attention has been shown to facilitate object processing for 9-months-old, but not 4-months-old, compared to a non-joint attentional condition in which the experimenter alternated gaze between object and the ceiling (Cleveland and Striano, 2007). A similar study found effects for 7-months-old but not 5-months-old (Cleveland et al., 2007). These results point at a developmental shift during the first year in how infants understand and make use of eye gaze cues in joint attentional settings to learn about objects. This shift occurs at about the same time as, or slightly earlier than the time that, children are starting to learn to understand, and perhaps even produce, their first word. Although 4–5-months-old infants might not be able to process the triadic interaction facilitated by eye gaze in complex interactional settings, it looks like there is a gradual shift toward more sophisticated understanding of joint attentional cueing, starting at 5 months (Cleveland et al., 2007).

However, infants' reliance on joint attention in object-processing tasks may be affected quite substantially by the nature, and in particular complexity, of the task. For example, Striano et al. (2006a) compared infant looking times in two conditions: (a) one in which the experimenter alternated their gaze between a toy and the infant while speaking about the toy during familiarization and (b) one in which the experimenter switched their gaze between a spot on the ceiling and the toy, without looking at the infant during familiarization. In the test phase, when the familiar and a novel toy were presented together, 12-months-old infants looked equally at the novel toy irrespective of the presentation condition, whereas 9-months-old looked at the novel toy only after the joint attention condition. This suggests perhaps that the reliance of infants on social cues depends on how challenging the task is at the developmental stage they are

in. Object processing may be a challenging task for 9-months-old infants, who may thus rely on the presence of joint attentional cues that simplify the task by directing their attention to a specific location. By 12 months, infants may be able to parse more of their environment more easily and thus may no longer depend on such cues for simple tasks, such as processing basic properties of a single object. However, they may still heavily make use of joint attentional cues in more complex, cognitively demanding settings, such as in the presence of multiple objects, more challenging object properties, or multimodal input.

That said, infants seem to clearly understand the referential nature of the gaze following action by 12 months. Twelve- and 18-months-old infants can follow an experimenter's gaze behind barriers where an object is hidden (Moll and Tomasello, 2004). Similarly, 8- and 12-months-old show surprise reactions when objects are not at the expected location when the barrier is removed (as indicated by longer looking times; Csibra and Volein, 2008). This suggests that they expect gaze to convey information about the object's location (i.e., referential information). Further evidence that infants expect human eye gaze to convey referential information comes from studies comparing infants' reactions to human and robot gaze. In a study comparing infants' gaze following behavior of a human and a robot agent, 12-months-old infants reliably followed both human and robot gaze, but they demonstrated reliable prediction of an object at the target location only when it was cued by human gaze (Okumura et al., 2013b). Moreover, their learning about objects seemed to be affected by the humanness of their interlocutor, as they only showed enhanced processing of the object when it was gazed at by a human agent (Okumura et al., 2013a). Interestingly, children only 2 months younger, at 10th month of age, failed to predict the appearance of the object when cued by a robot or a human agent (Okumura et al., 2013b; note that although this finding may seem to contradict previous studies showing learning of object properties after gaze following in younger infants, in fact the task in this study was more challenging, as the infants had to anticipate the location of the object in order to show successful learning).

The literature reviewed above suggests a role for eye gaze cueing in facilitating infants' object processing that develops during the first year of life; infant's novelty preference for uncued objects in the test phase is taken to reflect enhanced processing of cued objects during familiarization. However, this might not have long-term learning effects. For example, in one study, although 12-months-old infants followed the experimenter's gaze to the cued object, they only displayed a novelty preference for the uncued item during the first test trial, and not during later trials (Theuring et al., 2007). This raises the possibility that gaze cueing may have only short-term effects on 12-months-old infants' processing of objects (Reid and Striano, 2005).

It is also important to note that infants might learn from non-social cues as well. Barry et al. (2015) reported that 9-months-old successfully used both social (a person's eye gaze) and non-social (a rectangle) cues to learn about statistical object regularities. However, recent electroencephalogram (EEG) studies with 4-months-old infants comparing the effect of social and non-social cues in learning object properties provided

contrasting evidence, suggesting that infants' object processing was facilitated specifically by social cues (Wahl et al., 2013, 2019; Michel et al., 2019).

Another intriguing finding is that the presence of mutual gaze, possibly signaling communicative intent, might affect what infants learn about an object. Yoon et al. (2008) showed 9-months-old infants videos of an adult that either (a) pointed at an object while making eye contact with the infants and greeting them in an infant-directed manner (communicative context) or (b) reached for an object, without making eye contact or addressing the infants (non-communicative context). Infants retained information about the object identity, but not its location, when objects were presented in a communicative social context, and retained location but not identity information in the non-communicative setting. The authors suggested that ostensive cues in communication, such as eye contact, pointing, and infant-directed speech, may bias infants to encode generalizable features that support learning about object kinds. Variable information, such as the spatiotemporal features of an object, is deemed non-generalizable and thus is not retained. Note that a conceptual replication study by Okumura et al. (2016) only partially replicated the results. The authors reported an object identity bias in the communicative context but no location bias in the non-communicative context. Thus, the hypothesis that eye gaze might modulate what infants learn from interactions remains an important issue for future research.

In summary, a large number of studies report a facilitative role for eye gaze in infants' object processing. However, it is important to note that most, if not all, of the studies described above provide only indirect evidence that infants detect differences between cued and uncued objects because they are reliant on interpreting novelty preferences (they interpret novelty preference to uncued objects as indicating greater stimulus encoding or processing of the object presented in a prior cued phase). Neuroimaging studies, however, can provide more unambiguous evidence for differential processing in eye gaze cue vs. other conditions. Our review identified a number of studies ($n = 14$) that use neuroimaging paradigms to address these issues. The focus of these studies is mainly on identifying the neural mechanisms that underlie infants' enhanced processing and learning to understand why infants learn better in the presence of eye gaze cues, which we will review in detail in *Why Does Eye Gaze Facilitate Learning?* below. However, it is worth noting that a number of these studies provide direct evidence for the role of eye gaze during object processing, as they show that infants' neural responses to objects differ as a function of eye gaze during the object familiarization or test period (Reid et al., 2004; Reid and Striano, 2005; Striano et al., 2006b; Hoehl et al., 2008b, 2012; Parise et al., 2008; Kopp and Lindenberger, 2011, 2012; Wahl et al., 2013, 2019; Hutman et al., 2016; Michel et al., 2019). Thus, we conclude that the balance of evidence suggests that infants can reliably use adults' gaze to facilitate attention to a location by 4–5 months of age. Gaze cues also seem to lead to enhanced object processing in infants as young as 4 months. However, again, care must be taken in interpreting the overall results, as many studies may conflate eye gaze with other ostensive, joint attentional cues.

Speech Processing

The review identified a small number of studies ($n = 3$) showing that the direction of eye gaze, signaling whether the infant is addressed as the receiver of the communication, also modulated infants' neural responses to speech. In one study, mutual gaze (direct vs. averted) as well as object-directed gaze (referential vs. averted) influenced the ERP response to forward compared to backward speech in 4–5-months-old infants, both at early stages of processing (the Nc, for mutual gaze only) and at later latencies (slow wave, for both mutual and referential gaze; Parise et al., 2011). Similarly, a functional near-infrared spectroscopy (fNIRS) study by Lloyd-Fox et al. (2015) that used a naturalistic interaction design revealed that 6-months-old infants' cortical responses to infant-directed speech (and gestures) were enhanced in inferior frontal, anterior temporal, and temporoparietal regions when speech was presented with direct eye contact. These regions were found to be involved in the processing of communicative cues in previous studies (Grossmann et al., 2008). Interestingly, in this study, the facilitatory effect of eye gaze was only observed in combination with infant-directed speech, which is not surprising since this is the register caregivers generally use when talking to their infants (and which also may be processed as an ostensive cue). Besides these two studies reporting gaze effects on the neural processing of speech, one study also reported gaze effects on the discrimination of phonemic boundaries from speech. Conboy et al. (2015) examined 9.5–10.5-months-old English infants' joint attention with Spanish-speaking interlocutors in a live interactive setting in which the interlocutor described objects to the infants and read picture books to them. They found that infants' gaze shifts between the objects and the interlocutor, an index of their joint attention, predicted their perception of Spanish phonemes when tested at 11 months, such that infants with greater gaze shifts showed better neural discrimination of Spanish phoneme contrasts.

These findings suggest that eye gaze cues provided by the social partner, as well as the degree to which infants make use of them, might influence how infants process and learn from speech. In particular, gaze shifts may reflect infants' information processing abilities and signal attention to the information provided by the social partner, thereby increasing the opportunities for learning. However, as these conclusions come from only three studies, further work is needed to understand what drives this learning effect and the mechanisms that support a connection between social behavior and speech perception.

Why Does Eye Gaze Facilitate Learning?

In the previous sections, we presented evidence showing that infants tend to learn more in the presence of gaze cues compared to the absence of such cues. However, the discussion so far has not provided an answer to why infants learn better in the presence of gaze cues. Our systematic review process identified 32 studies that either directly addressed this question or that present evidence that speaks to this question. In this section, we first discuss the neurocognitive mechanisms by which eye gaze might have a facilitatory effect in infant's language learning. Then, we present

the evidence for the different theories that aim to specify the status of eye gaze in infants' learning: do infants learn better simply because gaze is an attention-grabbing cue or does gaze hold special meaning for infants, signaling the referential and communicative intent of the adult?

Neurocognitive mechanisms of the facilitatory role of eye gaze

A number of the studies identified in our review demonstrated an early specialization of the cortical regions that are involved in the processing of face-to-face communication cues, such as eye gaze perception, showing that even very young infants (at 4 months) show adult-like responses to eye gaze and facial communication cues (Grossmann et al., 2007, 2008) and display similar behavior even when presented with schematic gaze cues (Farroni et al., 2006). For example, in a neuroimaging study with 4-months-old infants, infants' gamma oscillatory activity was different for direct compared to averted gaze, in right frontotemporal regions, similar to adults (Grossmann et al., 2007). Furthermore, mutual gaze and eyebrow raise together with a smile (when mutual gaze was established) elicited similar neural activations, and the eyebrow raise with a smile led to this activation only when it was preceded by mutual gaze, so only when the infant was directly addressed. Possibly, this activation was only elicited when the cue was interpreted as ostensive (and communicative; Grossmann et al., 2008).

Eye gaze may function by facilitating infants' general attention and arousal, thereby increasing their receptivity in social interactions that foster learning. Extensive neuroimaging work has pointed at multiple neural correlates that differ as a function of eye gaze cues and which index attentional processes. For instance, Reid et al. (2004) showed that 4-months-old infants had enhanced positive slow wave (PSW) responses in their ERP signals to objects that were previously not cued with the experimenter's gaze, compared to cued objects. The PSW component is related to memory processes and stimulus encoding and has been found to be larger for novel objects and faces compared to already processed items (de Haan and Nelson, 1999). The authors, thus, argued that the infants needed to perform additional memory updating for the uncued objects, giving rise to the enhanced PSW. This effect was further modulated by the nature of the social cue (Michel et al., 2019) and familiarity of the adult, since in one study with 4-months-old infants, an enhanced PSW was only observed for uncued objects after objects were presented by the caregiver (Hoehl et al., 2012). This finding suggests that cues used by the caregiver might result in enhanced learning, possibly because of an additional increase in processing capacity and/or an increase in attention when interacting with a familiar adult. It should also be noted that older infants might benefit from caregivers' and strangers' eye gaze cues to a similar extent, since it has also been demonstrated that infants between 4 and 6 months show a stranger preference when following gaze (Gredeback et al., 2010).

Further, the negative central (Nc) component, which is taken to reflect attentional arousal and attentional orienting to salient stimuli (Richards et al., 2010), was found to be enhanced in response to objects that were previously cued by the adult's eye gaze and when joint attention was established by mutual

gaze before directing gaze to the object. For example, Parise et al. (2008) reported that 5-months-old had significantly larger Nc components in the left frontocentral regions in response to objects that were presented with joint attention (alternating gaze between infant and object after sharing mutual gaze with the infant) compared to the non-joint attention condition (no mutual gaze, looking at object only) during the familiarization phase (Parise et al., 2008). Hoehl et al. (2008b) presented similar results with 3-months-old, showing increased Nc for objects presented with direct gaze and a fearful expression. Findings of Wu et al. (2014) suggest that 8-months-old did learn more about the location of multimodal objects when ostensive cues, such as a video showing a person with direct eye gaze (while also verbally addressing the infant) preceded non-social attentional cues (flashing squares) in the training phase, even when the face did not turn toward the cued location. Similarly, a small sample EEG study suggested that "joint engagement," which presumably entails more than gaze cues (e.g., gestures and facial expression), led to a larger frontal positive component for objects presented with joint engagement and a larger Nc for objects presented without, indicating more familiarity for objects presented with joint engagement (Hutman et al., 2016). However, establishing a causal role for eye gaze in this observed enhancement effect is not possible in these studies due to coinciding ostensive cues (eye gaze presented together with verbal cues or gestures).

The studies we identified demonstrated that infants show differential brain states during object processing and social interaction with or without joint attention, involving direct eye contact, with an adult (Senju et al., 2006; Hoehl et al., 2014a,b; Michel et al., 2015; Urakawa et al., 2015). For example, Striano et al. (2006b) demonstrated that eye contact established before joint attentional periods during object viewing led to enhanced Nc in 9-months-old, reflecting attentional orienting or attentional arousal. This can then lead to more successful information encoding due to the channeling of limited attentional resources to the relevant aspects of information. Similarly, enhanced Ncs for objects not previously cued with adult's eye gaze shift or head direction were observed in 4-months-old (Hoehl et al., 2014b). Further, 9-months-old infants showed desynchronization of alpha oscillatory activity when viewing objects with an adult, only when the adult engaged in direct eye contact with the infant prior to orienting to the object (Hoehl et al., 2014a), similar to findings of joint attention studies with adults (Lachat et al., 2012). Similar results were observed for 4- and 9-months-old oscillatory activity for object-directed gaze (Michel et al., 2015), which was interpreted as a reflection of infants' developing executive attention control networks. The desynchronization of alpha-band activity, in the context of joint attention, is taken to reflect cortical excitation, attentional suppression of external input in order to focus on relevant information (Ward, 2003), and interestingly, an activation of a generic semantic knowledge system in adults (Klimesch, 2012). Thus, Michel et al. (2015) tentatively concluded that this desynchronization effect could reflect infants' enhanced receptive state of semantic knowledge transmission, which was activated by the use of ostensive gaze cues, thereby offering an interpretation of the attentional arousal

effect in terms of natural pedagogy. This proposal requires further investigation.

The difference in neural responses to objects could also arise from infants' differential neural processing of the adult's gaze in relation to the object. Infants processed the experimenter's gaze differentially when her gaze was directed to an object compared to when gaze was averted from an object (Hoehl et al., 2008a). Object-directed gaze led to an enhanced positive slow wave (PSW), while object-averted gaze elicited a more enhanced Nc, with its peak occurring significantly later compared to object-directed gaze (Hoehl et al., 2008a; Wahl et al., 2019). These results suggest that object-directed gaze might be encoded faster and require less attentional resources as reflected by the latency and amplitude of the Nc and may promote better memory encoding as reflected by the enhanced PSW. These could create opportunities for better processing of consequent environmental stimuli. Gaze cues in the context of joint attention (as in Striano et al., 2006b) might also affect the long-term retention of information about objects. Nine-months-old positive components (Pb; positive deflection between 200 and 400 ms, possibly reflecting contextual processing and expectation of an event) differed as a function of whether infants were familiarized with objects in joint attention or non-joint attention contexts. Similar effects were observed in another study both immediately and 1 week after familiarization (Kopp and Lindenberger, 2011), although in this study, two ostensive cues, direct eye contact and infant-directed speech, were conflated in the joint attention context.

Eye gaze in infants' learning: special or "just" attention?

Overall, studies testing infants' processing of eye gaze cues, mostly in relation to objects, suggest that eye gaze cues might facilitate learning by enhancing attention and memory encoding. However, while these studies provide a basis for interpreting eye gaze as a highly salient and advantageous social cue in infants' social communication and learning, they cannot provide a concrete answer to whether such ostensive cues have a special state for infants (i.e., whether they convey meaning over and above other attentional cues). This is because these studies do not directly compare eye gaze to other non-ostensive attentional cues. Our search procedure identified only a small number of studies that directly investigated whether it is the enhanced processing elicited by the social nature of such cues or their (low-level) attention-grabbing features, such as movement, that contribute to learning (Farroni et al., 2000). The results of these studies are, overall, inconsistent. Some report results that support the natural pedagogy theory (i.e., that there is enhanced processing associated with the ostensive nature of eye gaze cues), but others conclude that eye gaze is not more facilitative than other low-level attentional cues.

In support of the natural pedagogy theory, Senju and Csibra (2008) reported differences in how infants responded to the ostensive and non-ostensive cues that preceded an adult's head-turn/gaze switch toward an object. They demonstrated that 6-months-old infants followed an adult's gaze when the gaze switch was preceded by an ostensive cue, such as direct eye contact or infant-directed speech. The 6-months-old infants, however,

did not reliably follow gaze when the gaze switch followed a non-ostensive, attention-grabbing cue. Similarly, 8-months-old infants performed more successfully in learning the location of cues in multimodal events when ostensive cues (a face addressing the infants with direct eye contact, accompanied by infant-directed speech) preceded non-social attentional cues (flashing squares). This was true even when the ostensive cue itself did not orient toward the cued location. These results suggest that the ostensive cue helped infants learn from other non-ostensive cues (Wu et al., 2014).

There is also evidence that, during object processing, 4-months-old infants showed sensitivity to eye gaze cues but not to non-social attentional cues, as shown by their enhanced positive slow wave ERP responses to uncued objects (Michel et al., 2019). Further evidence is provided by Parise and Csibra (2013) who illustrated that 5-months-old infants' had overlapping electrophysiological responses to infant-directed speech and direct eye gaze in (pre)frontal regions, similar to adults. As direct eye gaze and infant-directed speech occur in different modalities, they do not have any common low-level physical properties; the overlapping brain activity must thus be due to another mechanism than the perception of low-level stimulus features. The authors hypothesized that if the observed activity in these regions was driven by increased attention, the combination of the two signals should produce a greater activity; however, the two signals gave rise to the same activity, with an early latency, as either signal in isolation. The authors took this obligatory response with an early latency as indicating infants' "fast and rudimentary interpretation" of the stimuli as ostensive, rather than resulting from the stimuli's low-level attention-grabbing features (Parise and Csibra, 2013). Interestingly, the combination of one ostensive and one non-ostensive signal, such as infant-directed speech (IDS) and averted gaze, did not cancel out the effects, but this might be due to the fact that the infants were too young to inhibit the early response to one ostensive signal, even if the accompanying cue in the other modality did not corroborate its ostensive nature.

Consistent with these results, 4-months-old infants' object processing was influenced by social cues (Wahl et al., 2013). Here, the effects were compared of directionally cueing objects with either an inanimate object (e.g., a car) or a human face. When the human face provided the cues, infants showed increased attention to, and processing of, uncued objects compared to the cued ones. This was indicated by increased looking times and enhanced Nc amplitudes for the uncued object, suggesting that the cued object was processed more efficiently (Wahl et al., 2013). When the cues were provided by the car, there were no looking time differences and only marginally significant ERP effects. However, in a later study, the authors raised concerns about the perceptual similarities between cars and human faces (features of the car stimuli that could be interpreted as face-like features by infants, such as side mirrors) and infants' possible familiarity with cars. Instead, they used a box with either a checkerboard pattern or with eyes as the central cue (Michel et al., 2019). Their results revealed a more robust enhanced PSW in response to uncued objects when cued with eyes (although the Nc component did not differ between conditions). This suggests that

social cues (even schematic patterns thereof) might play a specific role in infants' learning about objects, over and above other attention-grabbing, dynamic cues. Interestingly, another study found increased looking times and an enhanced Nc in response to objects that were previously not cued by isolated eyes gazing at the object (without a face). An enhanced slow-wave positivity was found in response to the object-directed (vs. averted) isolated eyes cue, suggesting that eyes only might be sufficient to facilitate object encoding (Wahl et al., 2019). This might also depend on the contrast polarity of the schematic images (black circles on white background vs. white circles on black background; Michel et al., 2017 also see Jessen and Grossmann, 2014).

The studies summarized above have been interpreted as support for natural pedagogy theory by many, since they seem to show differential (and sometimes enhanced) reactions or learning in the presence of ostensive cues. These reactions or learning, accordingly, do not result from the presence of low-level attentional, non-ostensive cues only (Csibra and Gergely, 2011). However, there are also studies that report no difference in differential gaze following or learning preceded by ostensive and non-ostensive conditions. These usually conclude that eye gaze is simply an attentional cue, which is highly salient for the infants from an early age on as evident by their automatic-like orientation toward its direction. On this view, eye gaze acts as a powerful attention modulator because it highlights to the infant where to attend in the noisy environment and which relevant information pieces are available in the environment to learn. On this view, eye contact may enhance infants' overall social attention to the environment and communication partner. This facilitates learning (Szufnarowska et al., 2014), but it does not necessarily hold a unique (ostensive) meaning for the infant to the extent that they treat it differently from other non-social attentional cues. Moreover, gaze following does not necessarily signal that infants understand the interlocutor's communicative intent.

Our review identified a number of studies ($n = 5$) showing that ostensive cues (such as eye contact but also infant-directed speech) do not necessarily need to be present for infants to follow an adult's gaze to a particular part of the environment. For instance, de Bordes et al. (2013) showed that 20-months-old infants followed the adult's gaze equally well after eye contact was established as when adult's eyes were made salient by placing colorful moving dots over them but no direct eye contact was present (although note that the adult's gaze was still directed at the infant, even when it was covered by the blinking dots, so infants might still have interpreted this condition as direct eye contact). In addition, these infants were substantially older, and thus capable of more sophisticated gaze cuing, than the children in many other studies). Similarly, it has been suggested that 6-months-old infants follow gaze in different ostensive and non-ostensive contexts, when the adult's action preceding the gaze orienting head turn was attention grabbing for the infant, irrespective of whether this action was ostensive or not (Szufnarowska et al., 2014; Gredebäck et al., 2018). Moreover, recent evidence has shown that infants between 11 and 24 months and their parents can coordinate visual attention without gaze following, by relying on the coordination of eyes and hands

in naturalistic, complex settings (Yu and Smith, 2017). These results suggest that it might be domain-general attention-based mechanisms, rather than the special status of the eyes, that explains why infants follow, and learn from, adults' gaze. Such domain-general accounts are also used to explain infants' ability at 9 months to learn from non-social cues, such as shapes, as well as from social cues when learning about object statistics (Barry et al., 2015). Other related theories have also proposed that infants acquire sensitivity to eye gaze through reinforcement learning by 9 months, without assigning a privileged status to eye gaze (Moore et al., 1997). Finally, one study found that 9-months-old infants learned object sequences equally well from social and non-social cues (Barry et al., 2015).

In sum, taken together, there exists a considerable body of literature suggesting that gaze is a highly attention-grabbing cue, to which infants show early sensitivity. The literature reviewed in the first part of this section presents quite convincing indirect evidence for differential processing and learning as a result of eye gaze cues compared to non-social attentional cues. However, the studies presented in the second half, which directly compared infants' tendency to follow the adult's gaze in ostensive and non-ostensive conditions, provided mixed evidence about the question of whether eye gaze is more than simply a high attention-grabbing cue. Thus, we only tentatively conclude that learning is especially enhanced when infants are addressed by ostensive signals, which may support the hypothesis that gaze cues facilitate infants' attention, arousal, and memory mechanisms in a way that other attentional cues do not.

GENERAL DISCUSSION

Summary of Results

Our review identified studies assessing the role of eye gaze in infants' language learning in four different domains: (1) vocabulary development, (2) word-object mappings, (3) object processing, and (4) speech processing. We then discussed the mechanisms by which eye gaze might play a role in infants' learning in these domains. With regards to vocabulary development, it appears that there is a strong association between infants' pre-linguistic communicative skills, such as following an adult's gaze direction, and their later receptive and expressive vocabulary. We suggest that this could be a cumulative result of enhanced processing due to eye gaze in the other domains we discussed, namely, word-object mappings, object processing, and speech processing. Eye gaze seems to facilitate the formation and retention of word-object mappings, as shown by both behavioral and neuroimaging studies, although the presence of other social cues coinciding with eye gaze in many of the studies makes it difficult to interpret whether the facilitation is indeed due to eye gaze. Similarly, object processing was found to be enhanced by eye gaze cues, although caution must be applied as, here too, some studies did not manipulate eye gaze in isolation. Finally, the limited evidence with regards to infants' speech processing suggests that infants might process speech sounds differently when accompanied by ostensive cues, such as eye gaze and that infants who shifted gaze learned more from the speech stream, as indexed by their phonemic discrimination. In general, therefore,

it seems that eye gaze can act as a powerful social cue in guiding infants' learning in different cognitive domains that are linked to language development.

There are two types of eye gaze cues in communication. The first is gaze alternation (the speaker alternates their gaze between the listener and the object being referred to), which invites the social partner to gaze follow. The second is simply establishing mutual eye contact with the partner. Many of the studies identified by our review focused on the first of these—gaze following. Such studies treat eye gaze as a spatiotemporal referential cue that signals the listener where to attend in the environment, so that referent–label associations may be formed. In that sense, the observed effects may be limited to certain domains or tasks that require a spatial referent in the environment, such as word–object mapping. However, it is also possible that the observed facilitatory effects of eye gaze reflect a general learning enhancement mechanism that is not confined to spatiotemporal mapping driven by gaze following but might follow the establishment of mutual eye contact. Our review provided support for a general enhancement mechanism. First, eye gaze in the form of gaze following seems to have a facilitatory role only when it is preceded by mutual eye contact. For instance, infants processed objects differently in live joint attentional contexts with an adult as a function of whether the adult provided eye contact or not, such that they had greater attentional mechanisms involved in object processing when the adult offered eye contact (Parise et al., 2008; Hoehl et al., 2014a). Moreover, the facilitatory effects are observed in domains that do not necessarily require spatial cues in the environment, such as speech processing, which was enhanced when the adult spoke to the infant in an infant-directed manner while providing mutual eye contact (Parise et al., 2011). Taken together, we suggest that eye gaze may have a general learning enhancing function in infants' (language) learning, such that the enhanced attentional and arousal mechanisms are observed when gaze following is accompanied by eye contact (which happens most frequently in natural interactions), and across domains. However, as few studies investigated the effects of eye gaze in domains other than object processing and word–object associations, the evidence to support our interpretation is limited.

Here, a related point that arose from our review is whether eye gaze holds a special status in infants' learning as a highly specialized socio-cognitive cue that is different from other attentional cues. Only a few studies directly assessed this question. However, once again, taking the evidence in all domains into consideration, the balance of evidence suggests that eye gaze may be a special attentional cue in that it facilitates learning to an extent that other low-level attentional cues cannot. Yet, we would argue, a developmental approach is necessary to fully understand the mechanisms by which children use eye gaze in learning. Our reading of the evidence to date is that children do not start out by treating eye gaze as an ostensive and referential cue but gradually learn to treat it as such throughout the first 2 years of life. Young infants have a preference for direct gaze, and for upright faces (Farroni et al., 2002), show early specialization of cortical regions involved in the processing of gaze cues and show mature neural responses to such cues. However, this does not mean that eye gaze

has a special status in human ontogeny from the start. Eye gaze could act as an attention-grabbing, albeit highly salient, cue early in development but not yet be treated as ostensive or referential. The development of an ostensive, referential understanding of eye gaze, instead, seems to develop between 9 and 12 months, as demonstrated by studies showing that infants follow gaze in conditions that signal referential, object-directed information by this age (Butler et al., 2000; Brooks and Meltzoff, 2002; Caron et al., 2002; Woodward, 2003; D'Entremont and Morgan, 2006; Johnson et al., 2007), and by neuroimaging studies showing that infants process referential information in an adult-like way by 9 months (Senju et al., 2006). Their ability to use eye gaze for object labeling, in an adultlike way, however, seems to come even later, at about 24 months of age.

This interpretation is also consistent with research showing that 8-months-old infants learned from social cues, whereas 4-months-old learned from non-social attention-grabbing cues, suggesting that “learning to learn” from social cues might be a skill that develops during infancy (Wu and Kirkham, 2010), and their sensitivity to social cues may develop gradually through the development of attention control, memory, and information processing networks (Yurovsky and Frank, 2017). Note, though, that a developmental explanation would not necessarily predict a linear developmental pattern. Multiple factors are likely to contribute to when and how infants make use of information provided by eye gaze cues, including the nature of the task and/or interaction they are engaged in. For example, it may be that older infants do not need to use eye gaze cues to solve simple object processing tasks (Striano et al., 2006a) but might still benefit from them in more complex settings, such as naturalistic interaction.

In fact, eye gaze as a social cue, in the form of mutual eye contact or gaze direction, rarely occurs in isolation in natural social–communicative contexts. Infant–adult social interactions are rich in a number of social signals that help infants in learning from others, and eye gaze often co-occurs with other ostensive cues, such as infant-directed speech and pointing. While this poses a problem for studying the role of eye gaze in isolation in infant learning, it also provides an important area for further research: to identify which kinds of rich communicative settings are optimal for learning. If we also consider recent evidence showing that infants show enhanced sustained attentional states during joint attention episodes (Yu et al., 2019), it will also be important to consider the role of the infants' endogenous attentional states within the context of parent–infant interactions.

In sum, eye gaze, both in the form of eye contact and gaze following, may direct and help infants in sustaining their attention and thus learn about relevant information in the environment. The involvement of attentional processes further corroborates the possibility that the use of eye gaze cues might serve infants' learning by highlighting the information to be attended and channeling their attentional resources. Further, eye gaze may, over the course of the first 2 years of life, develop into a truly ostensive, referential cue that enhances language learning across the board. However, further work is needed to fully understand the mechanisms behind the observed effects.

Limitations

In this review, we assessed the available experimental evidence on the effects of eye gaze on infants' learning and attention. Therefore, a number of other domains, such as face processing and emotion understanding were excluded from our analyses. Although this was intentionally done, it is important to acknowledge that infants' emerging social skills, such as face processing and understanding of others' emotions and intentions have effects on their cognitive abilities and language development, such as theory of mind development and mental state vocabulary. Separate reviews of these literature may throw additional light onto some of the issues discussed here.

Furthermore, as our review focuses on eye gaze, we narrowed our key search terms to include eye gaze or eye contact, rather than searching for literature on joint attention and learning. This allows us to focus on the role of gaze as intended but means we exclude literature using composite scores that include gaze as one of the components [e.g., studies that use the Early Social Communication Scales (ESCS) composite scores rather than the result for the individual gaze questions]. A future review, building on the present one, and collating information from a range of joint attentional tasks, would be a useful addition to the literature.

Another point is to note that the age at which infants were tested differed substantially across different learning domains. For example, in most of the studies on word-referent mapping (Word-Object Mapping), infants are older than in the studies on object processing (Object Processing). This is not unexpected, as children are rarely tested on their abilities to form word-object associations before 12 months and have difficulties to form these associations before 13 months (Woodward et al., 1994; Werker et al., 1998). However, the developmental differences make it difficult to compare the evidence across different domains, thus limiting our ability to draw concrete conclusions about the timescale of development.

Another important limitation is that some of the findings we report have not yet been replicated. Relatedly, we also observed that, for many studies, attrition posed a major challenge to the interpretation of the findings. For instance, Parise et al. (2008) tested 69 5-month-olds but were able to include data from only 15 infants. Fifty-two of those infants were excluded due to fussiness or for failing to reach a certain threshold that allowed for an adequate averaging of the ERP data. While the authors acknowledged this high dropout rate and argued that it was due to the relatively high task demands of their study, such high dropout rates are not unusual in the reviewed literature. This is a concern for the neuroimaging evidence in particular (e.g., Senju et al., 2006, who retained only 10/33 infants tested in the final dataset) but also for the behavioral studies. For behavioral studies, the dropout rates seem to differ with the numbers of participants recruited for the study and the task requirements (e.g., compare Cleveland et al., 2007 who retained 16/22 of the infants tested, to Gredebäck et al., 2018 where 94/95 of the infants were included in the final dataset). Thus, there are questions of generalizability

to be answered; for instance, is this evidence reflective only of a selected group of infants who seem to have better attentional spans as well as possibly better perceptual capacities?

Future Directions

Some additional themes that emerged from our search were not discussed in depth above because the literature was too sparse to draw reliable conclusions. For instance, in Speech Processing, we discussed the literature on speech processing, but the section is small because the literature is thin, leaving many unanswered questions. We need more of such work, which has the potential to address, directly, the question of whether gaze cues yield a general processing enhancement effect.

Another issue concerns the effect of live social interaction on language learning. Kuhl (2007) has suggested that social interaction is crucial for language learning, such that infants only learn to discriminate non-native phonemes in the context of live communication, not from videotaped interactions. This could be due to increased attention and arousal during live interactions compared to videotaped tutoring or to live situations being richer in social referential cues (such as eye gaze) that promote learning (although note that these two explanations are not mutually exclusive, since social cues might lead to enhanced attention). While many of the studies reported here tested language learning from audio or audiovisual stimuli presented in laboratory settings, it is possible that learning more complex linguistic information requires the presence of a live speaker who can convey the referential nature of the communication.

It is not currently possible to systematically compare the results of the studies that had a live interaction paradigm to those using prerecorded stimuli, since such studies had many methodological differences other than these variables, such as age and number of infants tested. Further research is needed to compare live interaction and classical lab studies that can assess the importance of natural interaction in different aspects of language learning. There has recently been a move toward studying social interaction in a more ecologically valid context and to consider how interpersonal communication affects information transfer, taking bidirectional influences between the partners into account. Recent dual-imaging work showed that eye gaze enhanced interpersonal brain synchrony between adult-infant pairs, in both live interaction and in a prerecorded condition (Leong et al., 2017). This provides possible explanations of mechanisms of how gaze functions to create learning opportunities for young infants during social interactions, perhaps by facilitating interpersonal synchrony through phase-resetting oscillatory activity and thereby putting children in a receptive state. Studying interpersonal neural dynamics in the context of infant learning is a fruitful area for further work.

Finally, to fully understand the socio-cognitive mechanisms that underlie the effects of eye gaze, we need more work directly testing whether eye gaze is interpreted as special, and ostensive, by infants or is treated as simply another attentional cue. Such studies must take account of the fact that there may be developmental- and task-specific differences in how children

react to eye gaze cues. It is probable that eye gaze serves different purposes at different ages for infants, starting as a salient attentional cue, and perhaps gaining a special status as infants develop. Further developmental work is required to establish the viability of these hypotheses.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

MÇ, TS, and CR contributed to the conception and design of the study. MÇ was primarily responsible for the literature search, the

study selection process, data extraction, and wrote the first draft of the manuscript. TS and CR supervised the study. All authors contributed to manuscript revision and read and approved the submitted version.

FUNDING

This work was funded by the MPI for Psycholinguistics Language Development Department Ph.D. studentship awarded to MÇ.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.589096/full#supplementary-material>

REFERENCES

- Akhtar, N., and Gernsbacher, M. A. (2007). Joint attention and vocabulary development: a critical look. *Lang. Linguist. Compass* 1, 195–207. doi: 10.1111/j.1749-818X.2007.00014.x
- Barry, R. A., Estes, K. G., and Rivera, S. M. (2015). Domain general learning: infants use social and non-social cues when learning object statistics. *Front. Psychol.* 6:551. doi: 10.3389/fpsyg.2015.00551
- Barry-Anwar, R. A., Burris, J. L., Estes, K. G., and Rivera, S. M. (2017). Caregivers and strangers: the influence of familiarity on gaze following and learning. *Infant Behav. Dev.* 46, 46–58. doi: 10.1016/j.infbeh.2016.11.005
- Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., and Ahluwalia, J. (2000). Is there an innate gaze module? *Evidence from human neonates. Infant Behav. Dev.* 23, 223–229. doi: 10.1016/S0163-6383(01)00037-6
- Beuker, K. T., Rommelse, N. N. J., Donders, R., and Buitelaar, J. K. (2013). Development of early communication skills in the first two years of life. *Infant Behav. Dev.* 36, 71–83. doi: 10.1016/j.infbeh.2012.11.001
- Brooks, R., and Meltzoff, A. N. (2002). The importance of eyes: how infants interpret adult looking behavior. *Dev. Psychol.* 38, 958–966. doi: 10.1037/0012-1649.38.6.958
- Brooks, R., and Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Dev. Sci.* 8, 535–543. doi: 10.1111/j.1467-7687.2005.00445.x
- Brooks, R., and Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *J. Child Lang.* 35, 207–220. doi: 10.1017/S030500090700829X
- Brooks, R., and Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: a longitudinal study connecting gaze following, language, and explicit theory of mind. *J. Exp. Child Psychol.* 130, 67–78. doi: 10.1016/j.jecp.2014.09.010
- Butler, S. C., Caron, A. J., and Brooks, R. (2000). Infant understanding of the referential nature of looking. *J. Cogn. Dev.* 1, 359–377. doi: 10.1207/S15327647JCD0104_01
- Caron, A. J., Butler, S., and Brooks, R. (2002). Gaze following at 12 and 14 months: do the eyes matter? *Br. J. Dev. Psychol.* 20, 225–239. doi: 10.1348/026151002166424
- Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63:i–vi, 1–143. doi: 10.2307/1166214
- Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., and Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cogn. Dev.* 15, 481–498. doi: 10.1016/S0885-2014(01)00037-5
- Cleveland, A., Schug, M., and Striano, T. (2007). Joint attention and object learning in 5- and 7-month-old infants. *Infant Child Dev.* 16, 295–306. doi: 10.1002/icd.508
- Cleveland, A., and Striano, T. (2007). The effects of joint attention on object processing in 4- and 9-month-old infants. *Infant Behav. Dev.* 30, 499–504. doi: 10.1016/j.infbeh.2006.10.009
- Conboy, B. T., Brooks, R., Meltzoff, A. N., and Kuhl, P. K. (2015). Social interaction in infants' learning of second-language phonetics: an exploration of brain-behavior relations. *Dev. Neuropsychol.* 40, 216–229. doi: 10.1080/87565641.2015.1014487
- Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind Lang.* 25, 141–168. doi: 10.1111/j.1468-0017.2009.01384.x
- Csibra, G., and Gergely, G. (2009). Natural pedagogy. *Trends Cogn. Sci.* 13, 148–153. doi: 10.1016/j.tics.2009.01.005
- Csibra, G., and Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 1149–1157. doi: 10.1098/rstb.2010.0319
- Csibra, G., and Volein, A. (2008). Infants can infer the presence of hidden objects from referential gaze information. *Br. J. Dev. Psychol.* 26, 1–11. doi: 10.1348/026151007X185987
- de Bordes, P. F., Cox, R. F. A., Hasselman, F., and Cillessen, A. H. N. (2013). Toddlers' gaze following through attention modulation: intention is in the eye of the beholder. *J. Exp. Child Psychol.* 116, 443–452. doi: 10.1016/j.jecp.2012.09.008
- de Haan, M., and Nelson, C. A. (1999). Brain activity differentiates face and object processing in 6-month-old infants. *Dev. Psychol.* 35, 1113–1121. doi: 10.1037/0012-1649.35.4.1113
- De Schuymer, L., De Groote, I., Striano, T., Stahl, D., and Roeyers, H. (2011). Dyadic and triadic skills in preterm and full term infants: a longitudinal study in the first year. *Infant Behav. Dev.* 34, 179–188. doi: 10.1016/j.infbeh.2010.12.007
- D'Entremont, B., Hains, S. M. J., and Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behav. Dev.* 20, 569–572. doi: 10.1016/S0163-6383(97)90048-5
- D'Entremont, B., and Morgan, R. (2006). Experience with visual barriers and its effects on subsequent gaze-following in 12- to 13-month-olds. *Br. J. Dev. Psychol.* 24, 465–475. doi: 10.1348/026151005X51248
- D'Odorico, L., Cassibba, R., and Salerni, N. (1997). Temporal relationships between gaze and vocal behavior in prelinguistic and linguistic communication. *J. Psycholinguist. Res.* 26, 539–556. doi: 10.1023/A:1025027830861
- Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., and Matthews, D. (2020). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: a longitudinal investigation of infants' vocalizations, gestures and word production. *Dev. Sci.* 23:e12843. doi: 10.1111/desc.12843

- Farroni, T., Csibra, G., Simion, G., and Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9602–9605. doi: 10.1073/pnas.152159999
- Farroni, T., Johnson, M. H., Brockbank, M., and Simion, F. (2000). Infants' use of gaze direction to cue attention: the importance of perceived motion. *Vis. Cogn.* 7, 705–718. doi: 10.1080/13506280050144399
- Farroni, T., Johnson, M. H., and Csibra, G. (2004). Mechanisms of eye gaze perception during infancy. *J. Cogn. Neurosci.* 16, 1320–1326. doi: 10.1162/0898929042304787
- Farroni, T., Menon, E., and Johnson, M. H. (2006). Factors influencing newborns' preference for faces with eye contact. *J. Exp. Child Psychol.* 95, 298–308. doi: 10.1016/j.jecp.2006.08.001
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual, 2nd Edn.* Oxford: Oxford Brookes University.
- Gogate, L. J., Bolzani, L. H., and Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy* 9, 259–288. doi: 10.1207/s15327078in0903_1
- Graham, S. A., Nilsen, E. S., Collins, S., and Olineck, K. (2010). The role of gaze direction and mutual exclusivity in guiding 24-month-olds' word mappings. *Br. J. Dev. Psychol.* 28, 449–465. doi: 10.1348/026151009X424565
- Graham, S. A., Nilsen, E. S., Friesen, C. K., and Johnson, J. (2011). Examining the role of attention and intention in two-year-olds' acquisition of novel words. *Enfance* 63, 311–328. doi: 10.4074/S0013754511003041
- Gredeback, G., Fikke, L., and Melinder, A. (2010). The development of joint visual attention: a longitudinal study of gaze following during interactions with mothers and strangers. *Dev. Sci.* 13, 839–848. doi: 10.1111/j.1467-7687.2009.00945.x
- Gredeback, G., Astor, K., and Fawcett, C. (2018). Gaze following is not dependent on ostensive cues: a critical test of natural pedagogy. *Child Dev.* 89, 2091–2098. doi: 10.1111/cdev.13026
- Grossmann, T., Johnson, M. H., Farroni, T., and Csibra, G. (2007). Social perception in the infant brain: gamma oscillatory activity in response to eye gaze. *Soc. Cogn. Affect. Neurosci.* 2, 284–291. doi: 10.1093/scan/nsm025
- Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., et al. (2008). Early cortical specialization for face-to-face communication in human infants. *Proc. Biol. Sci.* 275, 2803–2811. doi: 10.1098/rspb.2008.0986
- Hirotsu, M., Stets, M., Striano, T., and Friederici, A. D. (2009). Joint attention helps infants learn new words: event-related potential evidence. *Neuroreport* 20, 600–605. doi: 10.1097/WNR.0b013e3283282a0a7c
- Hoehl, S., Michel, C., Reid, V. M., Parise, E., and Striano, T. (2014a). Eye contact during live social interaction modulates infants' oscillatory brain activity. *Soc. Neurosci.* 9, 300–308. doi: 10.1080/17470919.2014.884982
- Hoehl, S., Reid, V., Mooney, J., and Striano, T. (2008a). What are you looking at? *Infants' neural processing of an adult's object-directed eye gaze.* *Dev. Sci.* 11, 10–16. doi: 10.1111/j.1467-7687.2007.00643.x
- Hoehl, S., Wahl, S., Michel, C., and Striano, T. (2012). Effects of eye gaze cues provided by the caregiver compared to a stranger on infants' object processing. *Dev. Cogn. Neurosci.* 2, 81–89. doi: 10.1016/j.dcn.2011.07.015
- Hoehl, S., Wahl, S., and Pauen, S. (2014b). Disentangling the effects of an adult model's eye gaze and head orientation on young infants' processing of a previously attended object. *Infancy* 19, 53–64. doi: 10.1111/inf.12035
- Hoehl, S., Wiese, L., and Striano, T. (2008b). Young infants' neural processing of objects is affected by eye gaze direction and emotional expression. *PLoS ONE* 3:e2389. doi: 10.1371/journal.pone.0002389
- Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., et al. (2000). Breaking the language barrier: an emergentist coalition model for the origins of word learning. *Monogr. Soc. Res. Child Dev.* 65, i–vi, 1–135.
- Hood, B. M., Willen, J. D., and Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychol. Sci.* 9, 131–134. doi: 10.1111/1467-9280.00024
- Houston-Price, C., Plunkett, K., and Duffy, H. (2006). The use of social and salience cues in early word learning. *J. Exp. Child Psychol.* 95, 27–55. doi: 10.1016/j.jecp.2006.03.006
- Hutman, T., Harrop, C., Baker, E., Elder, L., Abood, K., Soares, A., et al. (2016). Joint engagement modulates object discrimination in toddlers: a pilot electrophysiological investigation. *Soc. Neurosci.* 11, 525–530. doi: 10.1080/17470919.2015.1114966
- Jessen, S., and Grossmann, T. (2014). Unconscious discrimination of social cues from eye whites in infants. *Proc. Nat. Acad. Sci. U.S.A.* 111, 16208–16213. doi: 10.1073/pnas.1411333111
- Johnson, S. C., Ok, S.-J., and Luo, Y. (2007). The attribution of attention: 9-month-olds' interpretation of gaze as goal-directed action. *Dev. Sci.* 10, 530–537. doi: 10.1111/j.1467-7687.2007.00606.x
- Kleinke, C. L. (1986). Gaze and eye contact. A research review. *Psychol. Bull.* 100, 78–100. doi: 10.1037/0033-2909.100.1.78
- Klimesch, W. (2012). α -band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007
- Kopp, F., and Lindenberger, U. (2011). Effects of joint attention on long-term memory in 9-month-old infants: an event-related potentials study. *Dev. Sci.* 14, 660–672. doi: 10.1111/j.1467-7687.2010.01010.x
- Kopp, F., and Lindenberger, U. (2012). Social cues at encoding affect memory in 4-month-old infants. *Soc. Neurosci.* 7, 458–472. doi: 10.1080/17470919.2011.631289
- Kuhl, P. K. (2007). Is speech learning “gated” by the social brain? *Dev. Sci.* 10, 110–120. doi: 10.1111/j.1467-7687.2007.00572.x
- Lachat, F., Huguette, L., Lemarchal, J.-D., Conty, L., and George, N. (2012). Oscillatory brain correlates of live joint attention: a dual-EEG study. *Front. Hum. Neurosci.* 6:156. doi: 10.3389/fnhum.2012.00156
- Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., and Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. *Proc. Natl. Acad. Sci. U.S.A.* 114, 13290–13295. doi: 10.1073/pnas.1702493114
- Lewkowicz, D. J., and Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1431–1436. doi: 10.1073/pnas.1114783109
- Lloyd-Fox, S., Széplaki-Köllöd, B., Yin, J., and Csibra, G. (2015). Are you talking to me? Neural activations in 6-month-old infants in response to being addressed during natural interactions. *Cortex* 70, 35–48. doi: 10.1016/j.cortex.2015.02.005
- Markus, J., Mundy, P., Morales, M., Delgado, C. E. F., and Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Soc. Dev.* 9, 302–315. doi: 10.1111/1467-9507.00127
- Matatyaho, D. J., and Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word-object relations. *Infancy* 13, 172–184. doi: 10.1080/15250000701795655
- McGillion, M., Herbert, J. S., Pine, J., Vihman, M., dePaolis, R., Keren-Portnoy, T., et al. (2017). What paves the way to conventional language? *The predictive value of babble, pointing, and socioeconomic status.* *Child Dev.* 88, 156–166. doi: 10.1111/cdev.12671
- Michel, C., Pauen, S., and Hoehl, S. (2017). Schematic eye-gaze cues influence infants' object encoding dependent on their contrast polarity. *Sci. Rep.* 7:7347. doi: 10.1038/s41598-017-07445-9
- Michel, C., Stets, M., Parise, E., Reid, V. M., Striano, T., and Hoehl, S. (2015). Theta- and alpha-band EEG activity in response to eye gaze cues in early infancy. *Neuroimage* 118, 576–583. doi: 10.1016/j.neuroimage.2015.06.042
- Michel, C., Wronski, C., Pauen, S., Daum, M. M., and Hoehl, S. (2019). Infants' object processing is guided specifically by social cues. *Neuropsychologia* 126, 54–61. doi: 10.1016/j.neuropsychologia.2017.05.022
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed.1000097
- Moll, H., and Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Dev. Sci.* 7, F1–F9. doi: 10.1111/j.1467-7687.2004.00315.x
- Moore, C., Angelopoulos, M., and Bennett, P. (1997). The role of movement in the development of joint visual attention. *Infant Behav. Dev.* 20, 83–92. doi: 10.1016/S0163-6383(97)90063-1
- Moore, C., Angelopoulos, M., and Bennett, P. (1999). Word learning in the context of referential and salience cues. *Dev. Psychol.* 35, 60–68. doi: 10.1037/0012-1649.35.1.60
- Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., et al. (2000a). Responding to joint attention across the 6-through 24-month age period and early language acquisition. *J. Appl. Dev. Psychol.* 21, 283–298. doi: 10.1016/S0193-3973(99)00040-4
- Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Neal, R., and Schwartz, H. K. (2000b). Gaze following, temperament, and language development in

- 6-month-olds: a replication and extension. *Infant Behav. Dev.* 23, 231–236. doi: 10.1016/S0163-6383(01)00038-8
- Morales, M., Mundy, P., and Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behav. Dev.* 21, 373–377. doi: 10.1016/S0163-6383(98)90014-5
- Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., and Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Dev.* 78, 938–954. doi: 10.1111/j.1467-8624.2007.01042.x
- Mundy, P., Delgado, C., Goldstein, J., Parlade, M., Hogan, A., and Seibert, J. (2003a). *Early Social Communication Scales (ESCS)*. Coral Gables, FL: University of Miami.
- Mundy, P., Fox, N., and Card, J. (2003b). EEG coherence, joint attention and language development in the second year. *Dev. Sci.* 6, 48–54. doi: 10.1111/1467-7687.00253
- Mundy, P., and Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behav. Dev.* 21, 469–482. doi: 10.1016/S0163-6383(98)90020-0
- Niedzwiecka, A., Ramotowska, S., and Tomalski, P. (2018). Mutual gaze during early mother-infant interactions promotes attention control development. *Child Dev.* 89, 2230–2244. doi: 10.1111/cdev.12830
- O'Connell, L., Poulin-Dubois, D., Demke, T., and Guay, A. (2009). Can infants use a nonhuman agent's gaze direction to establish word-object relations? *Infancy* 14, 414–438. doi: 10.1080/15250000902994073
- Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., and Itakura, S. (2013a). The power of human gaze on infant learning. *Cognition* 128, 127–133. doi: 10.1016/j.cognition.2013.03.011
- Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., and Itakura, S. (2013b). Infants understand the referential nature of human gaze but not robot gaze. *J. Exp. Child Psychol.* 116, 86–95. doi: 10.1016/j.jecp.2013.02.007
- Okumura, Y., Kanakogi, Y., Kobayashi, T., and Itakura, S. (2017). Individual differences in object-processing explain the relationship between early gaze-following and later language development. *Cognition* 166, 418–424. doi: 10.1016/j.cognition.2017.06.005
- Okumura, Y., Kobayashi, T., and Itakura, S. (2016). Eye contact affects object representation in 9-month-old infants. *PLoS ONE* 11:e0165145. doi: 10.1371/journal.pone.0165145
- Parise, E., and Csibra, G. (2013). Neural responses to multimodal ostensive signals in 5-month-old infants. *PLoS ONE* 8:e72360. doi: 10.1371/journal.pone.0072360
- Parise, E., Handl, A., Palumbo, L., and Friederici, A. D. (2011). Influence of eye gaze on spoken word processing: an ERP study with infants. *Child Dev.* 82, 842–853. doi: 10.1111/j.1467-8624.2010.01573.x
- Parise, E., Reid, V. M., Stets, M., and Striano, T. (2008). Direct eye contact influences the neural processing of objects in 5-month-old infants. *Soc. Neurosci.* 3, 141–150. doi: 10.1080/17470910701865458
- Paulus, M., and Fikert, P. (2014). Conflicting social cues: fourteen- and 24-month-old infants' reliance on gaze and pointing cues in word learning. *J. Cogn. Dev.* 15, 43–59. doi: 10.1080/15248372.2012.698435
- Reid, V. M., and Striano, T. (2005). Adult gaze influences infant attention and object processing: Implications for cognitive neuroscience. *Eur. J. Neurosci.* 21, 1763–1766. doi: 10.1111/j.1460-9568.2005.03986.x
- Reid, V. M., Striano, T., Kaufman, J., and Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *Neuroreport* 15, 2553–2555. doi: 10.1097/00001756-200411150-00025
- Richards, J. E., Reynolds, G. D., and Courage, M. L. (2010). The neural bases of infant attention. *Curr. Dir. Psychol. Sci.* 19, 41–46. doi: 10.1177/0963721409360003
- Senju, A., and Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Curr. Biol.* 18, 668–671. doi: 10.1016/j.cub.2008.03.059
- Senju, A., Johnson, M. H., and Csibra, G. (2006). The development and neural basis of referential gaze perception. *Soc. Neurosci.* 1, 220–234. doi: 10.1080/17470910600989797
- Striano, T., Chen, X., Cleveland, A., and Bradshaw, S. (2006a). Joint attention social cues influence infant learning. *Eur. J. Dev. Psychol.* 3, 289–299. doi: 10.1080/17405620600879779
- Striano, T., Reid, V. M., and Hoehl, S. (2006b). Neural mechanisms of joint attention in infancy. *Eur. J. Neurosci.* 23, 2819–2823. doi: 10.1111/j.1460-9568.2006.04822.x
- Szafnarowska, J., Rohlfling, K. J., Fawcett, C., and Gredeback, G. (2014). Is ostension any more than attention? *Sci. Rep.* 4:5304. doi: 10.1038/srep05304
- Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., Morgan, J. L., and Shah, R. J. (2015). Attention to the mouth and gaze following in infancy predict language development. *J. Child Lang.* 42, 1173–1190. doi: 10.1017/S0305000914000725
- Theuring, C., Gredebäck, G., and Hauf, P. (2007). Object processing during a joint gaze following task. *Eur. J. Dev. Psychol.* 4, 65–79. doi: 10.1080/17405620601051246
- Urakawa, S., Takamoto, K., Ishikawa, A., Ono, T., and Nishijo, H. (2015). Selective medial prefrontal cortex responses during live mutual gaze interactions in human infants: an fNIRS study. *Brain Topogr.* 28, 691–701. doi: 10.1007/s10548-014-0414-2
- Vecera, S. P., and Johnson, M. H. (1995). Gaze detection and the cortical processing of faces: evidence from infants and adults. *Vis. Cogn.* 2, 59–87. doi: 10.1080/13506289508401722
- Wahl, S., Marinović, V., and Träuble, B. (2019). Gaze cues of isolated eyes facilitate the encoding and further processing of objects in 4-month-old infants. *Dev. Cogn. Neurosci.* 36:100621. doi: 10.1016/j.dcn.2019.100621
- Wahl, S., Michel, C., Pauen, S., and Hoehl, S. (2013). Head and eye movements affect object processing in 4-month-old infants more than an artificial orientation cue. *Br. J. Dev. Psychol.* 31, 212–230. doi: 10.1111/bjdp.12001
- Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn. Sci.* 7, 553–559. doi: 10.1016/j.tics.2003.10.012
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., and Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Dev. Psychol.* 34, 1289–1309. doi: 10.1037/0012-1649.34.6.1289
- Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Dev. Sci.* 6, 297–311. doi: 10.1111/1467-7687.00286
- Woodward, A. L., Markman, E. M., and Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Dev. Psychol.* 30, 553–566. doi: 10.1037/0012-1649.30.4.553
- Wu, R., Gopnik, A., Richardson, D. C., and Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Dev. Psychol.* 47, 1220–1229. doi: 10.1037/a0024023
- Wu, R., and Kirkham, N. Z. (2010). No two cues are alike: depth of learning during infancy is dependent on what orients attention. *J. Exp. Child Psychol.* 107, 118–136. doi: 10.1016/j.jecp.2010.04.014
- Wu, R., Tummelshammer, K. S., Gliga, T., and Kirkham, N. Z. (2014). Ostensive signals support learning from novel attention cues during infancy. *Front. Psychol.* 5:251. doi: 10.3389/fpsyg.2014.00251
- Yoon, J. M. D., Johnson, M. H., and Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13690–13695. doi: 10.1073/pnas.0804388105
- Yu, C., and Smith, L. B. (2017). Hand-eye coordination predicts joint attention. *Child Dev.* 88, 2060–2078. doi: 10.1111/cdev.12730
- Yu, C., Suanda, S. H., and Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Dev. Sci.* 22:e12735. doi: 10.1111/desc.12735
- Yurovsky, D., and Frank, M. C. (2017). Beyond naive cue combination: salience and social cues in early word learning. *Dev. Sci.* 20:e12349. doi: 10.1111/desc.12349

Conflict of Interest: 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 © 2021 Çetincelik, Rowland and Snijders. 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.



Mood Induction Differently Affects Early Neural Correlates of Evaluative Word Processing in L1 and L2

Johanna Kissler^{1*} and Katarzyna Bromberek-Dyzman²

¹ Department of Psychology, Bielefeld University, Bielefeld, Germany, ² Faculty of English, Adam Mickiewicz University, Poznań, Poland

OPEN ACCESS

Edited by:

Marcela Pena,
Pontificia Universidad Católica
de Chile, Chile

Reviewed by:

Ernesto Guerra,
University of Chile, Chile
Yang Zhang,
University of Minnesota Health Twin
Cities, United States

*Correspondence:

Johanna Kissler
johanna.kissler@uni-bielefeld.de

Specialty section:

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

Received: 29 July 2020

Accepted: 08 December 2020

Published: 12 January 2021

Citation:

Kissler J and
Bromberek-Dyzman K (2021) Mood
Induction Differently Affects Early
Neural Correlates of Evaluative Word
Processing in L1 and L2.
Front. Psychol. 11:588902.
doi: 10.3389/fpsyg.2020.588902

We investigate how mood inductions impact the neural processing of emotional adjectives in one's first language (L1) and a formally acquired second language (L2). Twenty-three student participants took part in an EEG experiment with two separate sessions. Happy or sad mood inductions were followed by series of individually presented positive, negative, or neutral adjectives in L1 (German) or L2 (English) and evaluative decisions had to be performed. Visual event-related potentials elicited during word processing were analyzed during N1 (125–200 ms), Early Posterior Negativities (EPN, 200–300 ms and 300–400 ms), N400 (350–450 ms), and the Late Positive Potential (LPP, 500–700 ms). Mood induction differentially impacted word processing already on the N1, with stronger left lateralization following happy than sad mood induction in L1, but not in L2. Moreover, regardless of language, early valence modulation was found following happy but not sad mood induction. Over occipital areas, happy mood elicited larger amplitudes of the mood-congruent positive words, whereas over temporal areas mood-incongruent negative words had higher amplitudes. In the EPN-windows, effects of mood and valence largely persisted, albeit with no difference between L1 and L2. N400 amplitude was larger for L2 than for L1. On the LPP, mood-incongruent adjectives elicited larger amplitudes than mood-congruent ones. Results reveal a remarkably early valence-general effect of mood induction on cortical processing, in line with previous reports of N1 as a first marker of contextual integration. Interestingly, this effect differed between L1 and L2. Moreover, mood-congruent effects were found in perceptual processing and mood-incongruent ERP amplification in higher-order evaluative stages.

Keywords: mood, emotion, language, bilingualism, word processing, context

INTRODUCTION

Bilinguals use two language systems to communicate and comprehend emotional meanings. Previous research has pointed to both differences and similarities in sensitivity to emotional content in bilinguals when they operate in their L1 and L2 (e.g., Pavlenko, 2012; Caldwell-Harris, 2015). It has indicated that linguistic systems acquired at different stages in life and with different proficiency, may vary also in the degree and depth of affective integration. Importantly, words people use to share meanings come coupled with contextual embeddings. Situational,

social and/or emotional contexts may endow single words' meanings with personal relevance, or communicative salience, and thus modify their processing. Transient feelings—moods, constitute one such communicative embedding: an emotional context against which words meanings are comprehended and interpreted. Here, we investigate whether and how moods modify the neurophysiological dynamics of word processing in the two linguistic systems of German-English bilinguals: German (L1) and English (L2).

Research on neural correlates of emotional word processing in L1 shows that valenced words, i.e., positive and negative ones, are processed more rapidly and evoke larger responses than neutral words (for reviews see, Kissler et al., 2006; Citron, 2012; Hinojosa et al., 2019). Event-related potentials (ERPs) research has shown that emotional features of words influence brain signatures at temporally distinct ERP components (e.g., Kissler and Herbert, 2013). Emotion effects for words have been most consistently reported at the early posterior negativity (EPN), peaking at around 200–300 ms post-stimulus, demonstrating higher amplitudes for emotional rather than neutral words (e.g., Kissler et al., 2007, 2009; Herbert et al., 2008; Palazova et al., 2011, 2013; Citron et al., 2013). N400 amplitudes, peaking around 400 ms, and showing smaller amplitudes for emotional than for valence-free words are also often reported (e.g., Sass et al., 2010; Palazova et al., 2013; Zhang et al., 2014). At later, integration stages emotional words typically elicit enhanced late positive potential (LPP) amplitudes peaking between 400 and 800 ms (e.g., Herbert et al., 2006, 2008; Hofmann et al., 2009; Schacht and Sommer, 2009a; Kissler and Herbert, 2013). By contrast, emotion effects at very early temporal stages are more sporadically observed (cf., Citron, 2012). Studies that detected such early effects, report amplified amplitudes on P1, peaking between 80 and 120 ms (e.g., Hofmann et al., 2009; Scott et al., 2009; Bayer et al., 2012; Schindler et al., 2019b), and/or on N1 between 100 and 200 ms post stimulus. They are typically valence-specific and often more pronounced for negative words (e.g., Scott et al., 2009; Kissler and Herbert, 2013; Yao et al., 2016; Schindler et al., 2019b), but have also been reported selectively for positive, e.g., happiness-related words (Briesemeister et al., 2014).

The visual N1 has been suggested as a first neural marker of context effects in word processing (Serenio and Rayner, 2003). Serenio et al. (1998) demonstrated the N1 to be sensitive to word-frequency effects in lexical decision. Moreover, this group also revealed that N1 amplitudes elicited by homonyms are modulated by meaning-biasing sentence context. For instance, N1 amplitude elicited by “bank” varied depending on whether the context contained “river” or “money” (Serenio et al., 2003). Scott et al. (2009) further showed an interaction of emotion with word frequency on the N1 in that the N1 was larger for high- than low-frequency negative words, whereas neutral words showed the opposite frequency modulation. Addressing neural effects of attributed social contexts, Schindler et al. (2019b) recently observed that valence effects on early brain potentials such as the N1 were elicited only when emotional trait adjectives were embedded in personally relevant communicative context—as a feedback personally targeted at the participant. When devoid of social embedding, the same emotional words elicited only late

ERP amplifications (LPP). Together, the above findings highlight the role of the N1 as an early marker of context integration in word processing, in line with cascaded interactive processing models (see also Hauk et al., 2006). Whilst, at least in L1, the N1 has been shown to be sensitive to the emotional content of words in reading as well as to some semantic and social contexts, it is presently unknown whether it responds to mood contexts.

Some ERP studies have compared the processing of emotional words in L1 and L2. A common assumption in the bilingualism literature is that bilinguals are less sensitive to the emotional aspects in L2 (e.g., Pavlenko, 2012). Yet, extant electrophysiological studies point to similarities, especially in proficient bilinguals. For instance, Opitz and Degner (2012) testing late, but highly proficient bilinguals, report similar, if latency-shifted, results in both groups of bilinguals tested—German-French and French-German. Enhanced processing of emotional compared to neutral words was reflected in a larger EPN measured between 280 and 430 ms after word onset. While the EPN effect itself did not differ in amplitude between L1 and L2, it was delayed for L2. This suggests that emotional word content in L2 is processed in a less immediate way due to delayed lexical access. Similarly, in a lexical decision study with late German-Spanish and Spanish-German bilinguals, Conrad et al. (2011) reported morphologically highly similar ERPs across L1 and L2: Larger EPN and LPP for emotional words compared to neutral words in both languages. Again, particularly EPN latencies were delayed in L2. However, specifically the patterns for negative content in L2 differed between more and less proficient participants. While in the more proficient bilinguals they observed enhanced EPN and LPP for both positive and negative words in L2, in the less proficient ones, ERP modulations were restricted to positive words. This indicates that negative, but not positive emotional words may be treated in an unemotional manner in the L2, which is in line with a recent study showing that there is a learning effect for negative words in general, such that negative emotional words tend to be acquired later than positive words (Ponari et al., 2017).

Indeed, a growing body of studies reports flattened behavioral and/or electrophysiological responses particularly to negative word valence in participants' L2. Several of these studies investigated the N400 ERP component (e.g., Wu and Thierry, 2012; Jończyk et al., 2016) which is a well-established marker of integration of words into their semantic context, particularly in sentence processing (e.g., Van Berkum et al., 1999).

Summing up, the second language research shows that, at least in proficient users, L2 should not be understood as totally unemotional, or driven by entirely different mechanisms than L1. Instead, the available evidence indicates weaker and delayed effects in L2 relative to L1, perhaps particularly regarding negative valence.

Recent interactional models of communication (e.g., Van Berkum, 2018, 2019) as well as embodiment theories (e.g., Matheson and Barsalou, 2018) emphasize that to gain insight into how individuals process and *experience* the affective content of words in communication, more attention should be paid to interactions between their respective *linguistic systems* and the accompanying *contexts*. One such context is mood, which has

recently been conceptualized as an “overarching state of mind” with pervasive influence on all aspects of cognition (Herz et al., 2020). Arguably, in communicative interactions people draw on contextual information including their somatic states: how they feel when interacting (e.g., Zajonc, 1980; Higgins, 1998) to constrain cognition and guide their actions. Accordingly, recent models of affective language comprehension (e.g., Van Berkum, 2018, 2019) posit that in order to make sense of verbal content, people rely on their moods as sources of information (cf. Clore and Huntsinger, 2009). Still, studies on word processing in bilinguals have hardly explored to what extent mood-states modify the processing of semantic and affective word content.

A recent functional magnetic resonance imaging (fMRI) study showed mood effects on language lateralization in general, revealing left lateralization of word fluency in anterior insula during happiness, and right hemisphere dominance during sadness (Costanzo et al., 2015). Bilingualism research has also found lateralization differences between L1 and L2 (e.g., D’Anselmo et al., 2013; Román et al., 2015), the combined findings suggesting that moods might differently impact neural processing of L1 and L2. So far, research targeting mood-effects in bilinguals boils down to one study testing mood and creativity (Kharkhurin and Altarriba, 2016), showing that positive mood enhanced creativity in participants’ dominant language, while negative mood boosted creativity in the non-dominant language.

Overall, it stands to reason that mood effects in either L1 or L2 should be most pronounced when emotional contents are processed, in line with the general ideas of mood-congruent processing (Bower, 1981) or affective priming (Klauer and Musch, 2003). Focusing on L1, several older behavioral studies have found mood-congruence effects for specific categories of words, but not for mood-valence agreement more broadly. That is, when happy or sad mood induction preceded lexical decisions on happiness- or sadness-related words, mood-congruent acceleration of reaction times was found (Niedenthal et al., 1994, 1997; Olafson and Ferraro, 2001; Ferraro et al., 2003), but the effect did not extend to positive or negative words in general (Niedenthal et al., 1994, 1997). Using a more extensive and more tightly controlled stimuli set than initial studies did, Sereno et al. (2015) recently observed faster reaction times in both a positive and a negative mood group compared to the control group (no mood induction). Moreover, whereas in positive mood reaction times were faster for both positive and negative words than for the neutral ones, in negative mood reaction times were fastest specifically for positive words, similar to what was found in the group without mood induction. This pattern was explained in terms of a general arousal-driven response acceleration in positive mood, in line with a motivated attention account that posits privileged processing of emotional content regardless of its valence (see also Kuperman, 2015). By contrast, automatic vigilance (Pratto and John, 1991) was suggested to operate in negative mood. Automatic vigilance refers to more pronounced attention capture, and delayed attentional disengagement from negative stimuli, therefore yielding faster reaction times for positive relative to negative words.

Kiefer et al. (2007) studied ERP correlates of mood effects on encoding of positive and negative adjectives. They specifically

hypothesized that good, but not bad mood would facilitate mood-congruent processing (Fiedler, 2001). Empirically, they observed valence differentiation only in good mood, but not necessarily always reflecting a mood-congruent pattern: Early (200–350 ms) valence-dependent ERP differences over left central scalp regions occurred only in good mood, with negative words eliciting more negativity than positive words, reflecting mood incongruence. Between 350 and 500 ms, also in good mood only, an N400-like ERP was less negative-going for positive than for negative words, suggesting facilitated processing of positive words in good mood. In the LPP-window (500–650 ms), again, valence modulated ERPs only in good mood: Negative words elicited a more positive potential than positive words, specifically over frontal sites and originating in frontal and temporal regions. Thus, valence differentiation and recruitment of language-related brain regions were stronger for good relative to bad mood, but not necessarily in a consistent mood-congruent manner. Herring et al. (2011), investigating ERP correlates of affective word priming also found slower reaction times to affectively incongruent than congruent targets, and a larger LPP to these affectively incongruent targets, whereas N400 was insensitive to evaluative prime-target congruency in that study.

Prior research has indicated that words with emotional meaning need not always evoke representations of emotional content/feelings (e.g., Niedenthal et al., 1994). It might therefore be instrumental to use a task that taps directly into emotional aspects of word meanings. Therefore, similar to Herring et al. (2011), we employed an *evaluative decision task*, which directs participants’ attention to the emotional representation of the word meaning. Unlike tasks that call for lexical access solely, an emotion evaluation task should direct participants’ attention to the emotional content, thereby potentially also enhancing somatic representations of words’ meanings, which might amplify brain responses to words’ emotional content and even facilitate mood congruence across broad valence categories.

In sum, here, we investigate whether and how happy and sad moods will impact evaluative word processing in bilinguals. Given previous evidence from lexical decisions (Sereno et al., 2015), we expect faster responses and larger amplitudes for both positive and negative-neutral words in happy mood in L1. This pattern would be in line with predictions based on motivated attention (Kuperman, 2015). In sad mood, automatic vigilance may operate, which should be reflected in delayed responses to negative words (see also Sereno et al., 2015). On the neurophysiological level, stronger valence differentiation is expected in happy relative to sad mood (Kiefer et al., 2007). We analyze N1, EPN, N400, and LPP brain potentials regarding their modulations by mood and valence in L1 compared with L2. The full sequence of ERPs is assessed to determine the theoretically important temporal stages of potential interactions between mood, language, and word valence. Previous research has pointed to N1 as the first locus of integration between content and context, suggesting it as the first time-window of interactions between mood, word valence, and language status. EPN has consistently shown higher amplitudes to emotional than to neutral words, with its peak delayed in L2. N400 is a general marker of semantic integration whose amplitude is

commonly larger for L2 (e.g., Ardal et al., 1990). N400 has also been found to be sensitive to emotional content (Herbert et al., 2008; De Pascalis et al., 2009; Holt et al., 2009; Moreno and Vázquez, 2011), its emotion modulation sometimes differing between L1 and L2 (Wu and Thierry, 2012; Martin et al., 2013; Jończyk et al., 2016). Mood-specific valence effects have also been reported on the N400 (Kiefer et al., 2007). Therefore, N400 could be another locus of integration of mood context with emotional content, which could further differ between L1 and L2. Finally, the LPP has been shown to be emotion-sensitive, including sensitivity to evaluative incongruence in priming (Herring et al., 2011) and mood sensitivity (Kiefer et al., 2007), but any differences between L1 and L2 remain to be explored. In order to specifically compare arousal- and valence-driven effects on the aforementioned components, we follow-up on any significant interactions with emotional content with pairs of linear and quadratic contrast. This allows us to distinguish between u-shaped (quadratic) effects that apply to both positive and negative content and are indicative of arousal-driven motivated attention effects, and valence-specific linear contrasts that differentiate between positive and negative contents, in line with predictions of automatic vigilance models. This strategy is commonly used in the emotion literature (e.g., Lang et al., 1993; Schindler et al., 2019a).

MATERIALS AND METHODS

Participants

Twenty-seven student participants were recruited at Bielefeld University. They provided written informed consent according to the Declaration of Helsinki and participated either in partial fulfillment of a course requirement or were independently recruited via flyers and received 20 Euros for taking part in an experiment consisting of two experimental sessions on separate days. Of the 27 participants four had to be excluded. Two did not return for the second experimental session, one of the course participants was not a German native speaker and one participant indicated a current attention deficit hyperactivity disorder (ADHD) diagnosis on medical history screening and had markedly increased depression scores on the Beck Depression Inventory (BDI, Beck et al., 2001). Thus, data from 23 participants were included in the analysis. All participants spoke German as their (L1). They reported using both German and English on an everyday basis, in both formal and informal contexts, yet with L1 being their dominant language (see **Tables 1, 2**). Our participants were late learners of English as their L2, which they learnt in formal school settings in Germany. Their proficiency level in English was assessed via an on-line LexTale test (Lemhöfer and Broersma, 2012), whose mean result indicated B2 proficiency level according to the Common European Framework of Reference for Languages (B2 cut-off > 60). In line with De Groot (2011), demographic information, and proficiency ratings, our participants are classified as upper intermediate/advanced, unbalanced, late English-German bilinguals. Due to experimenter error, LexTale scores are missing for two participants. All included participants were right-handed

and free from acute psychiatric or neurological disorder as indicated by self-report. None of the included participants exhibited elevated anxiety and depression scores as reflected on the State-Trait Anxiety Inventory (STAI, Spielberger et al., 1999) and the BDI. For full demographic information (see **Table 1**). Self-reported language history and communication skills according to the Language History Questionnaire (Li et al., 2014) are detailed in **Table 2**.

Stimuli

Word Material

Thirty six positive, negative, and neutral German adjectives were selected from the Berlin Affective Word List—Reloaded (BAWL-R, Vö et al., 2009). Adjectives appropriately differed in valence and arousal and were matched regarding concreteness, word length, word frequency, orthographic neighborhood density, and bigram frequency (see **Table 3**).

To create a corresponding English stimulus set, these words were translated into English. Corresponding values for English are given in **Table 4**.

Mood Induction

For mood induction, six different short movie excerpts with an average duration of 60 s were used. Three of these were happy and three were sad. The excerpts had been previously validated to generate the expected significantly different happy and sad moods states.

The happy clips were: “The Lion King: Final Scene (01:22:36–01:23:23; 47 s),” “The Lottery Ticket: Winning the Lottery” (00:23:53–00:25:07; 01:14 min), and “An Officer and Gentleman: Carried Away” (00:55:42–01:56:53; 01:11 min). The sad clips were: “The Lion King: Mufasa’s Death” (00:36:37–00:37:48; 01:11 min), “The Green Mile: John Coffee’s Death” (02:47:55–02:49:11; 01:16 min), and “The Champ: Final Scene” (01:53:08–01:54:05; 00:57 min). According to Gross and Levenson (1995), the final scene from “The Champ” is the most effective clip for inducing sad mood in their set. The clips were taken from the German and English versions of the movies, respectively.

Procedure

The experiment was divided into two sessions, taking place on two separate days. On the first day, upon arrival at the laboratory, participants were introduced to the EEG set-up and the aim of the study was explained to them in general terms as a study on

TABLE 1 | Demographic information for the participants.

Variable	(N = 23)
Gender female/male	18/5
Age	24.9 (19–39, 4.3)
BDI Score	5 (0–12, 3.9)
STAI trait	35.61 (24–51, 8.3)
STAI state session 1	32.7 (23–46, 5.75)
STAI state session 2	31.8 (24–40, 4.9)
LexTale score	69.5 (9.15, 48–87)

BDI, Beck Depression Inventory; STAI, State-Trait Anxiety Inventory. LexTale: Language Proficiency.

TABLE 2 | Linguistic information for the participants (Language History Questionnaire—LHQ; Li et al., 2014): self-reported information on the Age of acquisition of communicative competencies in English; Self-reported proficiency (1–7 scale), and daily time (in hours) spent using English.

Age of acquisition of English		Self-reported proficiency in English		Daily use of English (hours/day)	
Speaking	9.33 (1.77)	Speaking	5.19 (0.93)	Watching TV	0.8 (1.0)
Reading	9.52 (1.47)	Reading	5.57 (0.84)	Reading for fun	0.5 (0.5)
Writing	9.76 (1.27)	Writing	5.43 (0.87)	Interacting via the Internet	0.7 (0.6)
		Listening	5.38 (0.86)	Speaking with friends	0.5 (0.7)

emotion in language processing in their L1 and L2. While the electrodes were being attached, participants completed several questionnaires: On the first appointment, a demographic and health questionnaire, the BDI, and the STAI state questionnaire were administered. On the second appointment the STAI state and trait questionnaires as well as the LHQ and the LexTale test were given.

After electrode placement, the study was explained in more detail: Participants were told that they would see short video clips that they should watch attentively. Thereafter, they would be presented with words that they should categorize via button-press (left arrow, up-arrow or right arrow) as positive, negative, or neutral. This procedure would repeat several times after which the words would switch to a different language in a separate language block (English or German, respectively).

Words were presented in three blocks, each preceded by a short movie clip. The valence of the mood induction remained constant for three blocks in a row. Word blocks consisted of 36 items each, 12 positive, 12 negative, and 12 neutral. Word order was randomized within each block separately. Words were presented in white font (Arial, 40 pts) on a black screen, each for 616 ms, followed by a white fixation cross prompting participants to respond. The fixation cross was presented for a randomly varying inter-stimulus interval (ISI) of 1.9–2.3 s.

After block 1 and 2, participants were given a short self-paced break to allow them to relax briefly. After the third block of each session, participants were asked to assess their emotional state, i.e., rating subjectively felt valence and arousal on a nine-point Self-Assessment Manikin scale (Bradley and Lang, 1994) as well as their level of current happiness or sadness, on a seven point Likert scale.

Then, participants were allowed a longer break and the experimental language and mood induction were switched. The above described procedure was repeated with mood inductions and words presented in the other language. The experiment was controlled via Presentation software¹.

At the end of the first experimental session, another appointment was made for a second, analogously structured, experimental session. Experimental conditions were counterbalanced with the restriction that participants always underwent two different mood blocks and two different languages per session.

Analyses of Behavioral Data

Behavioral data were analyzed according to their match with predefined word categories. Number of word assignments

per category (positive, neutral, and negative) as well as reaction times were analyzed within a response window of 1,500 ms following stimulus onset. Reaction times were corrected for outliers, excluding responses that exceeded ± 2 SD of the individual mean and recalculating the reaction time. Statistical analyses were performed using repeated measures ANOVA with the factors Mood (Happy, Sad), Language (L1: German, L2: English), and Word Content (positive, neutral, negative).

TABLE 3 | Means for the set of German word attributes are given with standard deviations in parentheses; valence and arousal values are derived from the Berlin Affective Word List revised (BAWL-R; Vö et al., 2009) BAWL-R values range from –3.5 to + 3.5; lexicographic values come from dlex (Heister et al., 2011); means sharing the same superscript do not differ statistically.

Adjectives	Positive	Neutral	Negative
Valence	2.0 (0.06) ^a	0.04 (0.07) ^b	–2.0 (0.04) ^c
Arousal	2.9 (0.07) ^a	2.3 (0.08) ^b	3.1 (0.11) ^a
Concreteness	3.2 (0.14) ^a	3.1 (0.18) ^a	3.4 (0.15) ^a
Word length	6.6 (0.21) ^a	6.9 (0.19) ^a	7.1 (0.16) ^a
Word frequency (dLex)	31.4 (13.74) ^a	21.8 (7.17) ^a	20.0 (9.19) ^a
Orthographic neighborhood	0.7 (0.22) ^a	0.8 (0.18) ^a	0.6 (0.20) ^a
Bigram frequency	167086.6 (20480.01) ^a	175062.9 (19243.04) ^a	181801.1 (22722.29) ^a

Means not sharing superscripts differ. Comparisons are based on Fisher's LSD test post-hoc comparisons.

TABLE 4 | Means for the word attributes for the set of English words.

Adjectives	Positive	Neutral	Negative
Valence	6.9 (0.7) ^a	5.4 (0.9) ^b	3.1 (0.8) ^b
Arousal	4.8 (0.9) ^a	3.6 (0.6) ^b	4.6 (1.0) ^a
Concreteness	2.3 (0.5) ^a	2.5 (0.8) ^a	2.4 (0.5) ^a
Word length	6.8 (1.7) ^a	6.8 (1.8) ^a	6.7 (1.9) ^a
Word frequency (zipf)	3.9 (0.6) ^a	3.4 (0.8) ^a	3.6 (0.9) ^a
Orthographic neighborhood	6.8 (1.7) ^a	3.7 (4.1) ^{a1}	6.4 (9.0) ^a

Standard deviations in parentheses; means sharing the same superscript do not differ statistically. Means not sharing superscripts differ. Evaluations of the English word set for valence, arousal come from Warriner et al. (2013); scales range 1–9; zipf frequencies from SUBTLEX_US (van Heuven et al., 2014); concreteness values from Brysbaert et al. (2014); scales range: 1–5 (abstract-concrete); Orthographic Neighborhood based on CLEARPOND (Marian et al., 2012).

¹ Unfortunately a number of neutral words was missing from the CLEARPOND set, hence the larger standard deviation relative to positive and negative word valence.

¹ www.neurobehavioralsystems.com

EEG Recording and Analyses

EEG was recorded from 32 BioSemi active electrodes² sampled at 1,024 Hz. Two separate electrodes were used as ground electrodes, a Common Mode Sense active electrode (CMS) and a Driven Right Leg passive electrode (DLR), which formed a feedback loop that enabled measuring the average potential close to the reference in the A/D-box³. Four additional electrodes (EOG) placed near the outer canthi and below the eyes measured horizontal and vertical eye movement.

Pre-processing and statistical analyses were performed using BESA⁴ and EMEGS (Payk et al., 2011). Offline, data was re-referenced to an average reference and a forward 0.16 Hz high-pass and a zero-phase 30 Hz low-pass filter were applied. Filtered data were segmented from 100 ms before word onset until 1,000 ms after stimulus presentation. The 100 ms before stimulus onset were used for baseline correction. Eye-movements were corrected using the automatic eye-artifact correction method implemented in BESA (Ille et al., 2002). ERP data were statistically analyzed with EMEGS (Payk et al., 2011).

ERPs were averaged according to predefined word categories matched for other lexical variables (see section “Materials and Methods”) and analyzed in 5 different time windows and components, namely the N1 (125–200 ms), EPN1 (200–300 ms), EPN2 (300–400 ms), N400 (350–450 ms), and LPP (500–700 ms). Time-windows largely correspond to those in previous studies (see e.g., Herbert et al., 2008; Scott et al., 2009). We divided EPN into two time windows in order to be able to assess any processing delay for emotional content in L2 as suggested by previous research (Conrad et al., 2011; Opitz and Degner, 2012). Analyses were performed at two symmetrical occipital (O1, PO3, P3, P7 and O2, PO4, P4, P8) and temporal (CP5, T7, FC5, F7 and CP6, T8, FC6, F8) electrode groups for N1, EPN1, and EPN2 components. For N400 a fronto-central group consisting of Cz, Fz, FC1, and FC2 and for LPP a centro-parietal group comprising P3, CP1, Pz, P4, and CP2 were employed. Number and location of the grouped electrodes largely corresponded to the one presented by Scott et al. (2009) who also used four electrodes per cluster. As in Dehaene (1995) and according to the observed scalp topographies, we analyzed early negativities (N1 and EPN) at temporal as well as occipital sites.

Statistical analyses were conducted in EMEGS and SPSS 25. Analyses of variance (ANOVAs) were performed with the repeated measurement factors Mood Induction (happy, sad), Language (L1: German, L2: English), Word Valence (positive, neutral, negative) for behavioral data and N400 and LPP components. For the N1 and EPN components, laterality of Channel Group (left, right) was added to assess expected hemispheric asymmetries in language and mood processing. Significant higher-level ANOVAs were broken down into follow-up ANOVAs and the shapes of any valence-dependent (positive, neutral, negative) differences were determined with pairs of polynomial trend tests, comparing linear and quadratic trends, significant linear trends indicating valence-dependent effects and

significant quadratic trends indicating arousal-driven effects (see also Lang et al., 1993; Schindler et al., 2019a). If the sphericity assumption was violated, degrees of freedom and *p*-values were corrected according to the Huynh-Feldt procedure. In line with the literature, we report uncorrected degrees of freedom and corrected *p*-values for better readability. Partial eta-squared (η_p^2) was estimated to describe effect sizes (Cohen, 1988).

RESULTS

Behavior

Manipulation Check

Participants rated their moods as significantly more positive [$F(1, 22) = 30.43, p < 0.001, \eta_p^2 = 0.58$] after the happy than after the sad mood induction. Mood valence did not differ between L1 and L2 [$F(1, 22) = 0.4, p = 0.53, \eta_p^2 = 0.02$] and the effect of mood induction did not interact with language [$F(1, 22) = 1.3, p = 0.27, \eta_p^2 = 0.06$]. By contrast, mood induction did not impact self-rated arousal [$F(1, 22) = 0.0, p = 1, \eta_p^2 = 0.0$] in either language [$F(1, 22) = 0.27, p = 0.61, \eta_p^2 = 0.01$] and the interaction was likewise insignificant [$F(1, 22) = 0.24, p = 0.63, \eta_p^2 = 0.01$]. Self-rated sadness was higher following sad than happy mood induction [$F(1, 21) = 27.34, p < 0.001, \eta_p^2 = 0.57$], with no difference between the languages [$F(1, 21) = 2.25, p = 0.15, \eta_p^2 = 0.1$] and no interaction [$F(1, 21) = 0.96, p = 0.34, \eta_p^2 = 0.04$]. One participant failed to complete the sadness rating.

Word Evaluations

An analysis of evaluations according to predefined valence categories revealed that in L1 considerably more words were evaluated as expected than in L2 [$F(1, 22) = 48.74, p < 0.001, \eta_p^2 = 0.69$]. Overall, more words were evaluated as either positive or negative than as neutral [$F(2, 44) = 15.87, p < 0.001, \eta_p^2 = 0.42$; quadratic: $F(1, 22) = 19.08, p < 0.001, \eta_p^2 = 0.46$, linear: $F(1, 22) = 0.84, p = 0.37, \eta_p^2 = 0.04$], but a highly significant interaction of valence and language [$F(2, 44) = 21.0, p < 0.001, \eta_p^2 = 0.49$] reflected that this was considerably more pronounced in L1 [quadratic: $F(1, 22) = 42.24, p < 0.001, \eta_p^2 = 0.65$; linear: $F(1, 22) = 1.36, p < 0.25, \eta_p^2 = 0.06$] than in L2 [quadratic: $F(1, 22) = 4.64, p < 0.05, \eta_p^2 = 0.17$, linear: $F(1, 22) = 6.21, p < 0.05, \eta_p^2 = 0.22$]. In particular, whereas in L1 considerably more words were assigned to both the positive [$t(22) = 5.3, p < 0.001$] and the negative [$t(22) = 7.32, p < 0.001$] than to the neutral category, in L2 assignment to positive differed from neutral [$t(22) = 2.58, p < 0.05$] whereas negative and neutral did not differ [$t(22) = 1.48, p = 0.15$]. **Figure 1** shows how word evaluations were distributed across the valence categories in the two languages.

Reaction Times

As shown in **Figure 2**, words were evaluated faster in L1 (German) than in L2 (English) [$F(1, 22) = 8.07, p < 0.01, \eta_p^2 = 0.27$] and emotional words were evaluated faster than neutral ones [$F(2, 44) = 41.1, p < 0.001, \eta_p^2 = 0.65$]. A trend-level interaction indicated that participants took a little longer when they evaluated negative L2 words than negative L1 words

²www.biosemi.com

³www.biosemi.com/faq/cms&drl.htm

⁴www.besa.de

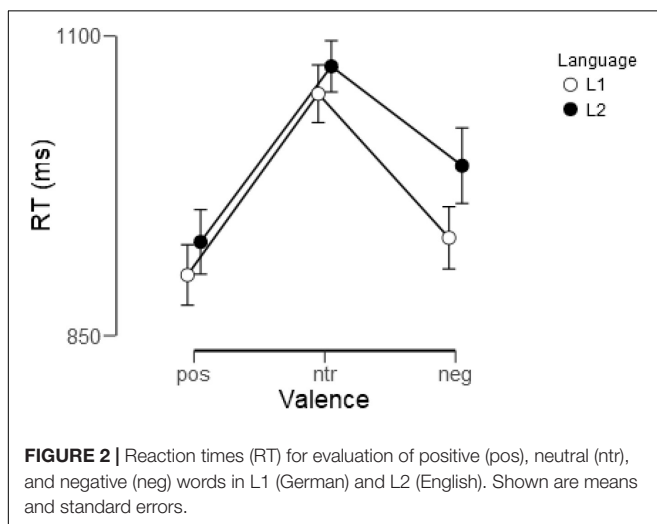
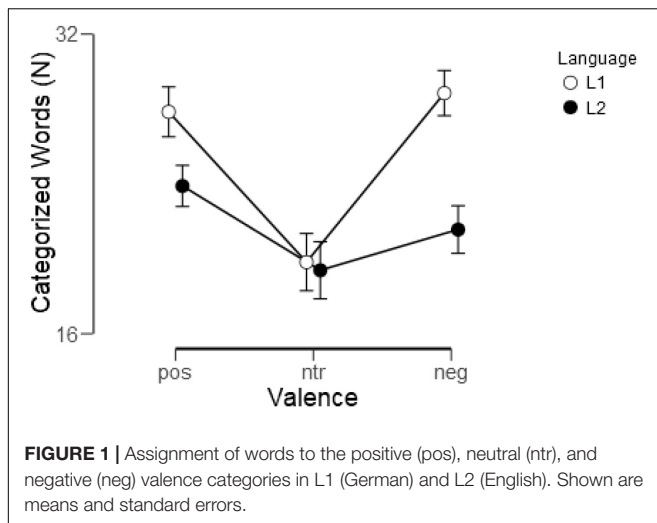


TABLE 5 | Reaction times for evaluative decisions.

Mood	Language	Valence	Mean	SD
Happy	L1	pos	884.89	123.52
		ntr	1046.67	126.22
		neg	913.79	129.39
	L2	pos	926.34	119.39
		ntr	1072.42	148.30
		neg	989.23	142.21
Sad	L1	pos	916.39	125.72
		ntr	1056.96	123.47
		neg	949.56	113.18
	L2	pos	930.12	126.15
		ntr	1076.91	111.74
		neg	994.41	152.74

[$F(2, 44) = 2.74, p < 0.1, \eta_p^2 = 0.11$], reaction times for the other two categories not differing. Reaction times for the individual experimental conditions are detailed in **Table 5**.

ERP data

Occipital N1

Left-lateralization of the occipital N1 in word processing was reflected in a main effect of channel group [$F(1, 22) = 5.9, p < 0.05, \eta_p^2 = 0.21$].

Moreover, as shown in **Figure 3**, valence differentiation was found following happy but not sad mood induction, as evident in an interaction of Mood with Valence [$F(2, 44) = 4.62; p = 0.02; \eta_p^2 = 0.17$]. In detail, in happy mood an effect of valence was found [$F(2, 44) = 9.65; p < 0.01, \eta_p^2 = 0.30$] in that N1 was largest for positive words, negative and neutral not differing [linear: $F(1, 22) = 12.87, p < 0.01, \eta_p^2 = 0.37$; quadratic: $F(1, 22) = 6.82, p < 0.05, \eta_p^2 = 0.24$]. By contrast, no valence effect was seen following sad mood induction [$F(2, 44) = 0.25, p = 0.8, \eta_p^2 = 0.01$]. **Figure 3** illustrates the interaction of mood and valence, showing the valence effect in happy but not in sad mood. **Figure 3** also suggests valence differentiation in the occipital N1 to be primarily driven by the right hemisphere, but the interaction was not significant [Valence \times Channel Group, $F(2, 44) = 2.92, p = 0.06; \eta_p^2 = 0.12$, see also **Figure 3**]. Finally, a three-way interaction of language with mood and channel group was found [$F(1, 22) = 5.54, p < 0.05, \eta_p^2 = 0.2$] in that mood affected N1 lateralization differently in the two languages. This effect was mainly driven by temporal rather than occipital activity (see detailed analysis below). No other main effects or interactions were significant.

Lateral N1

Over lateral parts of the N1, mood induction interacted with valence [$F(2, 44) = 6.96, p < 0.01, \eta_p^2 = 0.22$, see **Figure 4**]. Following happy mood induction, the valence effect [$F(2, 44) = 5.66, p < 0.01, \eta_p^2 = 0.21$] occurred because ERPs elicited by negative words were more negative-going than ERPs elicited by positive words, neutral words falling in between [linear: $F(1, 22) = 15.1, p < 0.01, \eta_p^2 = 0.4$; quadratic: $F(1, 22) = 0.15, p = 0.7, \eta_p^2 = 0.007$]. This was not the case following sad mood induction [$F(2, 44) = 1.64, p > 0.1, \eta_p^2 = 0.07$].

As a main finding in this time-window, mood induction impacted the lateralization of word processing differently in L1 and L2 [Mood \times Language \times Channel Group: $F(2, 44) = 11.045, p < 0.005, \eta_p^2 = 0.33$]. As shown in **Figure 5**, in L1 (German), mood induction had a highly significant effect on the lateralization of word processing [$F(1, 22) = 11.35, p < 0.005, \eta_p^2 = 0.34$]. N1 was more negative over the left than over the right channel group following happy mood induction [$F(1, 22) = 4.8, p < 0.05, \eta_p^2 = 0.18$] with no lateralization following sad mood induction [$F(1, 22) = 1.05, p < 0.32, \eta_p^2 = 0.01$]. An interaction of mood and channel group was also present in L2 [$F(1, 22) = 6.67, p < 0.05, \eta_p^2 = 0.23$] and its pattern seemed reversed (see **Figure 5**, bottom row). However, in L2 follow-up tests were not significant.

There was also a three-way interaction of language with valence and channel group [$F(2, 44) = 3.352, p = 0.044, \eta_p^2 = 0.13$], but follow up ANOVAs were all insignificant.

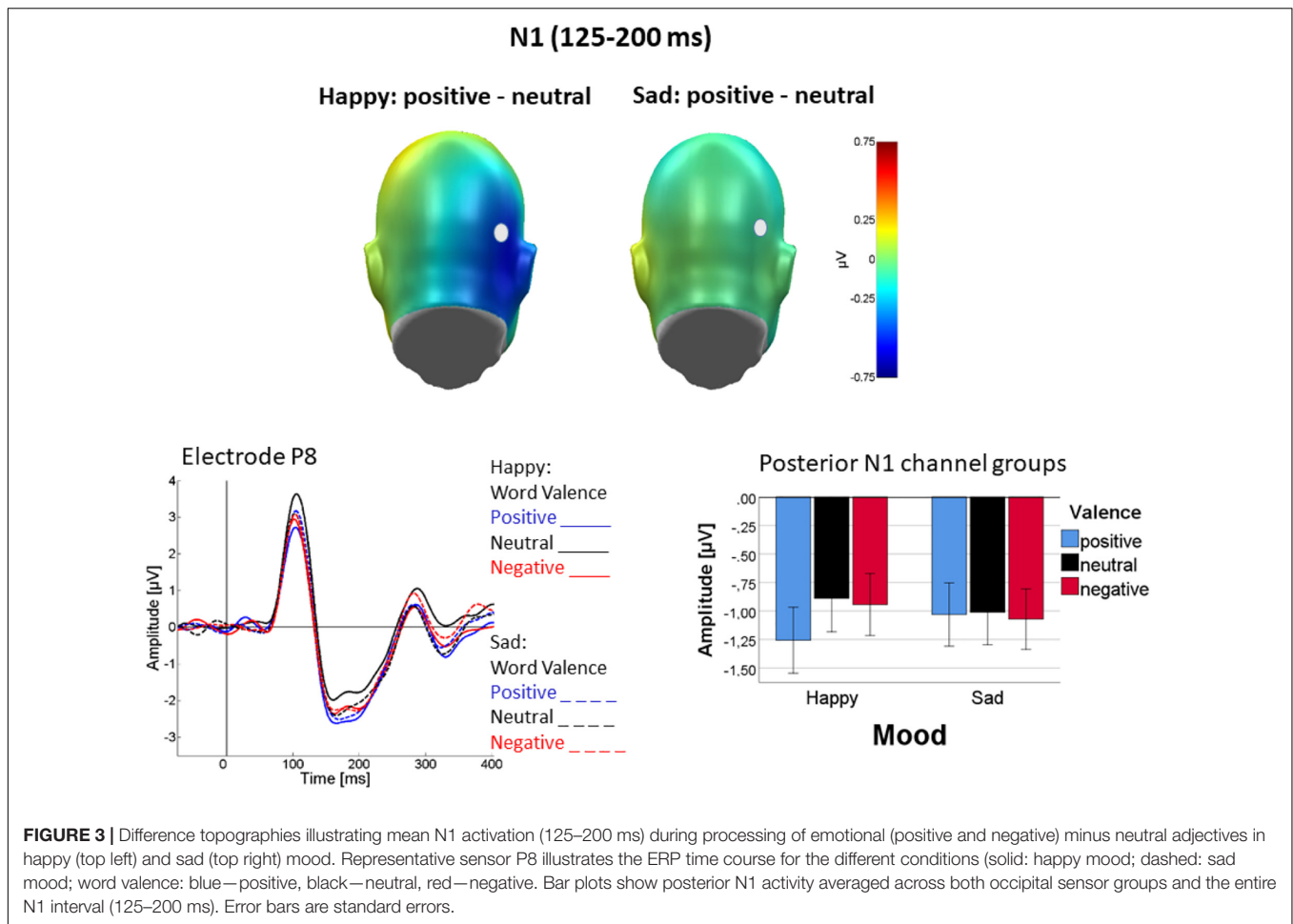


FIGURE 3 | Difference topographies illustrating mean N1 activation (125–200 ms) during processing of emotional (positive and negative) minus neutral adjectives in happy (top left) and sad (top right) mood. Representative sensor P8 illustrates the ERP time course for the different conditions (solid: happy mood; dashed: sad mood; word valence: blue—positive, black—neutral, red—negative). Bar plots show posterior N1 activity averaged across both occipital sensor groups and the entire N1 interval (125–200 ms). Error bars are standard errors.

Furthermore, a complex four-way interaction of mood with language, valence, and channel group [$F(2, 44) = 4.722, p = 0.014; \eta_p^2 = 0.17$] occurred. However, none of the follow-up tests was significant.

Occipital EPN1

In the early occipital part of the EPN, valence interacted with channel group [$F(2, 44) = 3.57, p < 0.05, \eta_p^2 = 0.14$] reflecting linear valence discrimination, with more negative-going ERPs for positive than negative words over the left occipital cortex [linear: $F(1, 22) = 5.67, p < 0.05, \eta_p^2 = 0.20$; quadratic: $F(1, 22) = 0.1, p = 0.76, \eta_p^2 = 0.76$], whereas over right occipital cortex valence discrimination was insignificant [$F(1, 44) = 1.11, p > 0.05, \eta_p^2 = 0.05$]. **Figure 6** illustrates this pattern.

Lateral EPN1

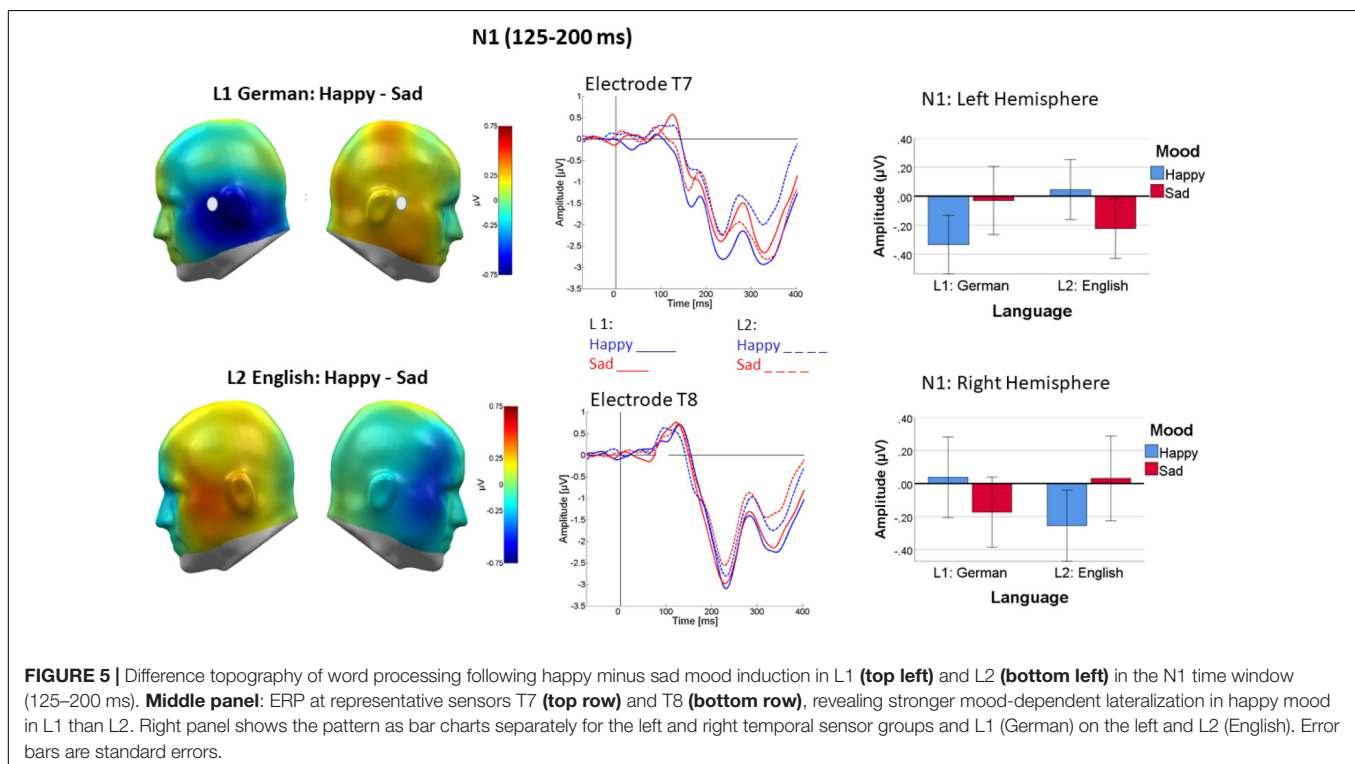
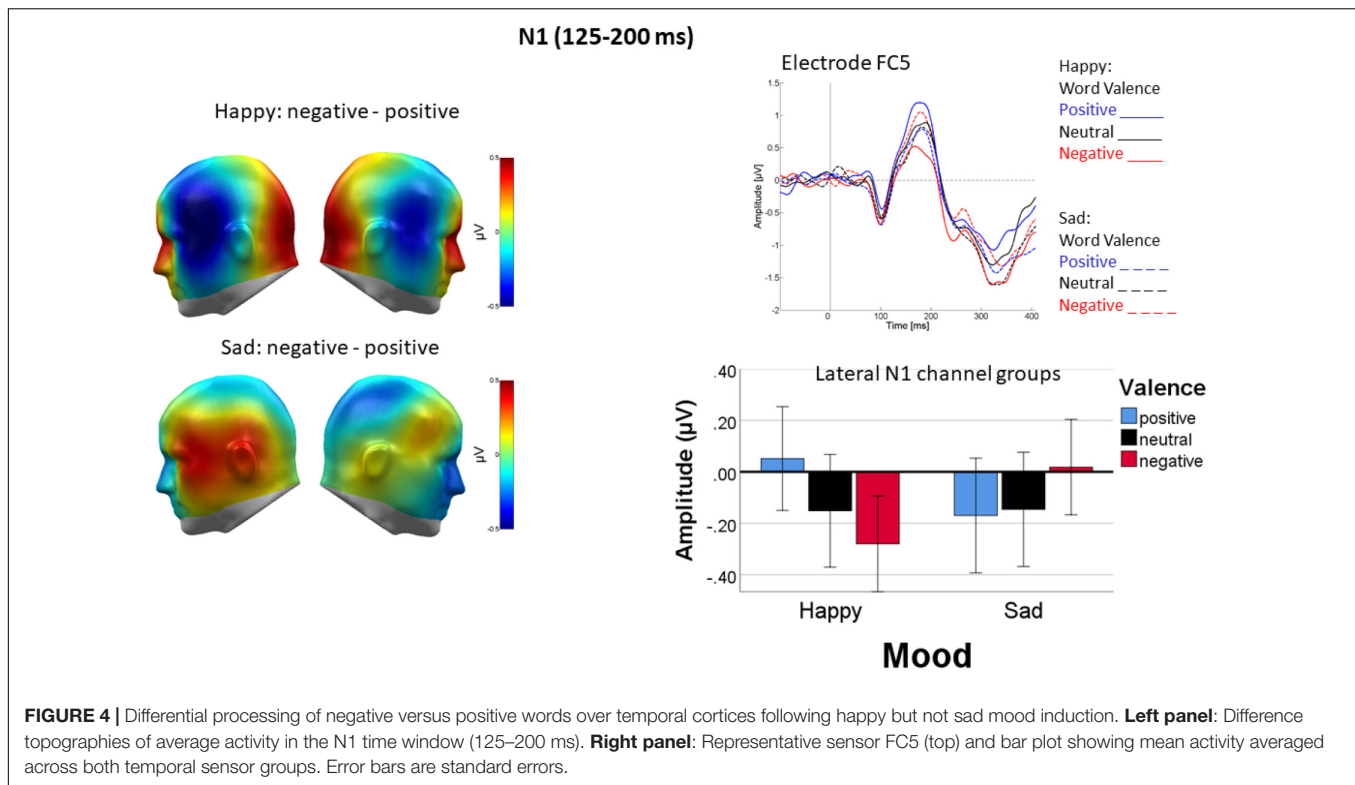
Over lateral sensors, an interaction of mood with valence occurred [$F(2, 44) = 5.33, p < 0.01, \eta_p^2 = 0.19$]. Its pattern was descriptively similar to the N1 (see **Figure 4**), but follow-up tests were not significant ($p_s > 0.07$). An interaction of valence and channel group was also found [$F(2, 44) = 4.096, p = 0.023, \eta_p^2 = 0.16$], but follow-up tests could not resolve it (all $p_s > 0.1$). No other effects approached significance ($p > 0.11$).

Occipital EPN2

For the later part of the EPN, the interaction of valence with channel group persisted [$F(2, 44) = 3.51, p < 0.05, \eta_p^2 = 0.14$]. The valence effect over the left channel group [$F(2, 44) = 4.54, p < 0.05, \eta_p^2 = 0.17$] was due to linearly more negative-going potentials for positive than negative words [linear: $F(1, 22) = 13.42, p < 0.36, \eta_p^2 = 0.38$; quadratic: $F(1, 22) = 0.084, p = 0.77, \eta_p^2 = 0.004$], whereas the descriptively inversely u-shaped valence differentiation over the right hemisphere was insignificant [$F(2, 44) = 1.03, p > 0.1, \eta_p^2 = 0.05$]. **Figure 6** illustrates EPN modulation by emotional words for both analyzed time-windows. No other effects were significant ($p > 0.07$).

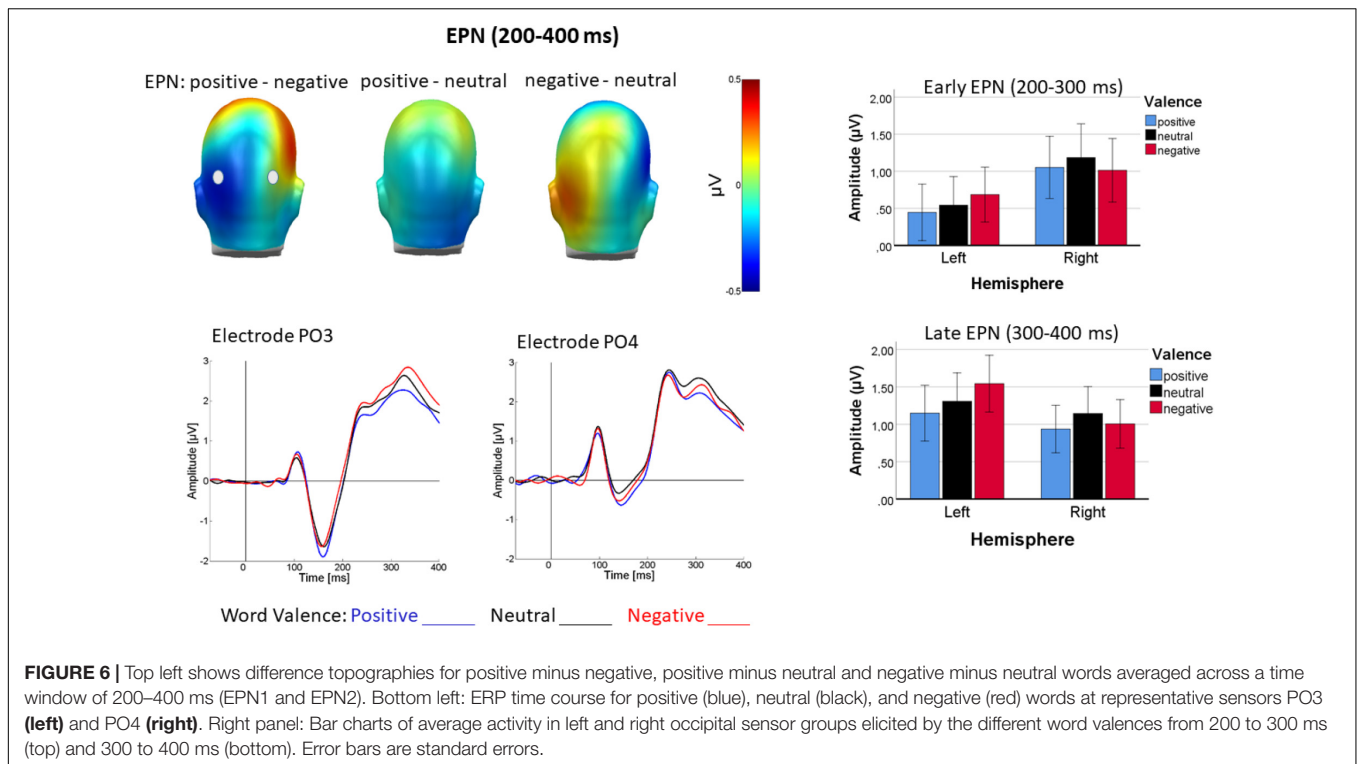
Lateral EPN2

Over lateral temporal sensors, an effect of language occurred [$F(1, 22) = 4.34, p < 0.05$] in that ERPs were more negative-going for L1 (German) than L2 (English). An interaction of mood and word valence [$F(2, 44) = 6.57, p < 0.01, \eta_p^2 = 0.23$] resembled the pattern found for the N1 and can be seen in the sensor tracings in **Figure 4**. It was due to a valence effect following happy mood induction [$F(2, 44) = 4.89, p < 0.01, \eta_p^2 = 0.18$] such that negative words were most negative going [linear: $F(2, 22) = 6.57, p < 0.05, \eta_p^2 = 0.23$; quadratic: $F(2, 44) = 3.26, p = 0.08, \eta_p^2 = 0.13$], whereas the valence effect in sad mood was not significant [$F(2,$



44) = 2.5, $p = 0.09$, $\eta_p^2 = 0.1$]. In particular, negativity elicited by negative words was more pronounced following happy than following sad mood induction [$t(22) = -2.79$, $p < 0.05$]. An

interaction of valence with channel group [$F(2, 44) = 3.361$, $p < 0.05$, $\eta_p^2 = 0.13$] was also present, but none of the follow-up tests was significant.



N400

On the N400, a main effect of language [$F(1, 22) = 5.061$; $p = 0.035$, $\eta_p^2 = 0.19$] emerged, reflecting a larger N400 in L2. No other main effects or interactions occurred ($ps > 0.2$).

LPP

On the LPP, a main effect of word valence [$F(2, 44) = 5.925$, $p = 0.005$, $\eta_p^2 = 0.21$], reflecting higher LPP amplitudes for both positive and negative rather than for neutral words [linear: $F(1, 22) = 0.63$, $p < 0.43$, $\eta_p^2 = 0.03$], quadratic: [$F(1, 22) = 10.61$, $p < 0.005$, $\eta_p^2 = 0.47$], and an interaction of mood and word valence [$F(2, 44) = 8.815$; $p = 0.001$, $\eta_p^2 = 0.29$] were found. **Figure 7** illustrates that the interaction reflected cortical accentuation of mood-incongruent content. The valence effect following positive mood induction [$F(2, 44) = 9.37$, $p < 0.001$, $\eta_p^2 = 0.30$] arose, because here negative words elicited highest amplitudes and amplitudes for positive words fell between negative and neutral [linear: $F(1, 22) = 5.46$, $p < 0.05$, $\eta_p^2 = 0.20$], quadratic: [$F(1, 22) = 12.35$, $p < 0.005$, $\eta_p^2 = 0.36$]. By contrast, for the valence effect following negative mood induction [$F(2, 44) = 5.65$, $p < 0.01$, $\eta_p^2 = 0.20$], positive words elicited highest amplitudes, with little difference between negative and neutral words [linear: $F(1, 22) = 7.23$, $p < 0.05$, $\eta_p^2 = 0.25$], quadratic: [$F(1, 22) = 4.24$, $p = 0.05$, $\eta_p^2 = 0.16$]. No other effects approached significance ($p > 0.13$).

DISCUSSION

In the present study, we investigated how the processing of emotional word content is modulated by moods. We specifically

compared how happy and sad moods affect the processing timeline of emotional adjectives when participants responded to them in L1 (German) and L2 (English), respectively. In two mood induction conditions, the same sets of positive, neutral and negative trait adjectives were presented either in participants' L1 or L2 while they evaluated the emotional content of the presented adjectives. Based on previous literature, we considered predictions from mood-congruence (Bower, 1981; Niedenthal et al., 1994), motivated attention (Kuperman, 2015), automatic vigilance (Pratto and John, 1991) and more emotion differentiation in good mood (Fiedler, 2001; Kiefer et al., 2007). Against these backgrounds, we aimed to establish empirically the timeline of mood and word valence interactions, examining specifically whether these effects would be observed already at the early (N1, EPN) ERPs and whether the brain potentials evoked in L1 and L2 would differ in amplitude and latency.

Behavioral data showed that participants differed in their responses to emotional word content in their L1 and L2. Their evaluations were faster in L1, their functionally dominant language, than in L2. They were also faster on both positive and negative emotional words than on neutral words, which is theoretically in line with the pattern expected by the motivated attention model (see Kuperman, 2015). However, reaction times for negative L2 words, while being considerably faster than for neutral ones, were slower than for positive L2 words which provides further evidence for attenuated processing of negative contents in L2 as suggested by some previous research (Wu and Thierry, 2012; Sheikh and Titone, 2016; Baumeister et al., 2017; Jończyk et al., 2019). Also, in L1, relative to L2, considerably more words were evaluated according to predefined

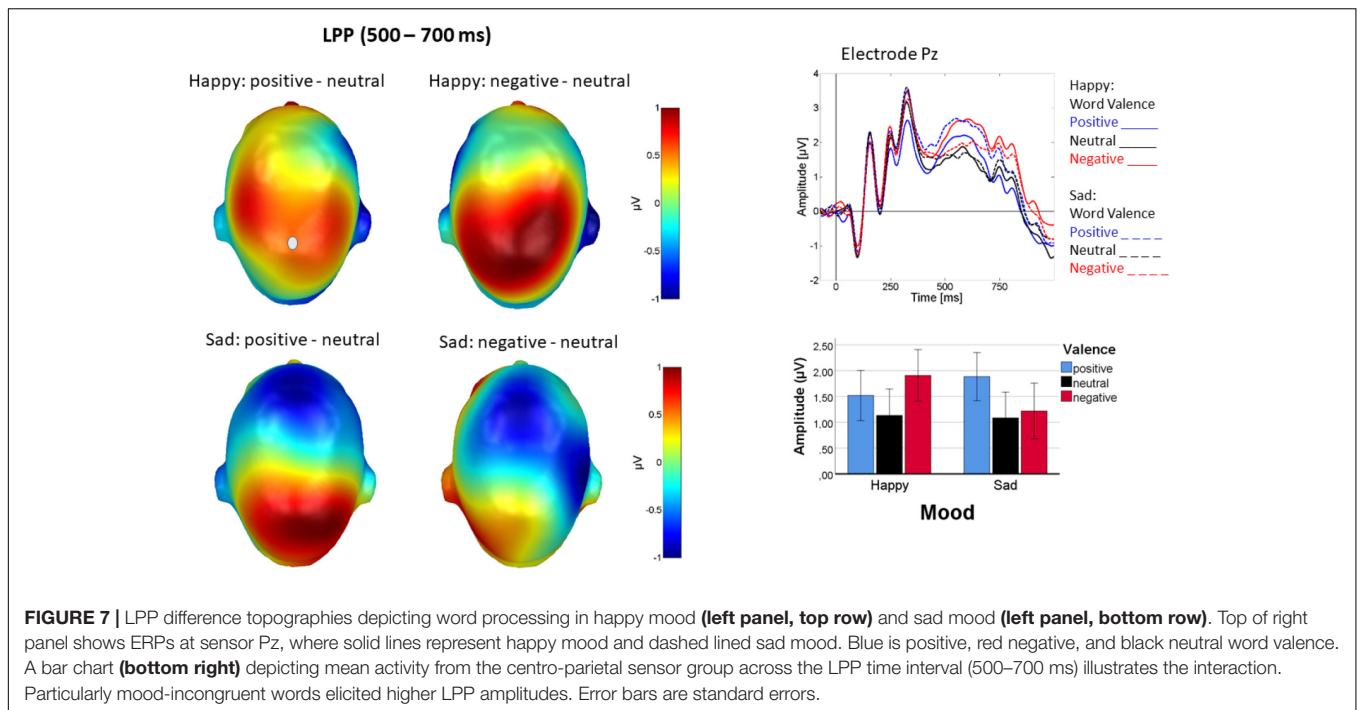


FIGURE 7 | LPP difference topographies depicting word processing in happy mood (left panel, top row) and sad mood (left panel, bottom row). Top of right panel shows ERPs at sensor Pz, where solid lines represent happy mood and dashed lined sad mood. Blue is positive, red negative, and black neutral word valence. A bar chart (bottom right) depicting mean activity from the centro-parietal sensor group across the LPP time interval (500–700 ms) illustrates the interaction. Particularly mood-incongruent words elicited higher LPP amplitudes. Error bars are standard errors.

word valence. This finding, indicating greater emotional distance in L2 relative to emotional words in L1, is consistent with a range of studies exploring a phenomenon referred to as a *foreign language effect*, which posits that when functioning in L2, people adopt a more utilitarian thinking style, which leads to different decision patterns as compared to when they operate in their L1 (Keysar et al., 2012; Costa and Sebastián-Gallés, 2014; Hayakawa et al., 2017).

Evaluation and reaction time data showed no effect of mood induction in either language, although in both languages self-rated mood questionnaires showed an expected difference on subjective valence and sadness ratings. One reason for this might be that the evaluative decision task, while making the affective dimension more salient than lexical decision does, also results in longer and more variable reaction times. Moreover, unlike anticipated, explicit emotion evaluation may override any more subtle implicit processes that mood might have on overt behavior. At any rate, present behavioral data provide no evidence in favor of mood-congruent processing suggested by older lexical decision studies (Niedenthal and Setterlund, 1994; Niedenthal et al., 1997) or any other of the above models, although a more recent study (Serenio et al., 2015) revealed mood effects on lexical decisions. However, mood clearly impacted the neurophysiology of word processing.

Neural Effects of Mood Induction

N1

Cortical differentiation of word valence for happy, but not for sad mood, already occurred on the N1. This was observed over occipital as well as temporal areas, although the pattern differed in that over occipital regions enhancement of N1 to positive

words was mood-congruent, whereas over temporal areas mood-incongruent negative words induced more negativity. More pronounced valence differentiation in happy than sad mood, as seen in several time windows, is in agreement with findings by Kiefer et al. (2007) who hypothesized that specifically in positive moods, contents would be encoded in an assimilative manner, favoring mood-congruent processing (Fiedler, 2001). Whereas we found stronger valence differentiation in word-evoked ERPs in happy mood across all early components, the pattern was not always mood-congruent, which was previously observed by Kiefer et al. (2007).

Strikingly, regardless of word valence, mood induction affected the lateralization of word-evoked N1 over temporal sites, the pattern differing between the two languages. In L1, the N1 component was strongly left-lateralized following happy mood induction, which was not the case following sad mood induction. This pattern was absent in L2. In line with previous research, both the early interaction of mood and valence and the interaction of mood and language status confirm the N1 time-window as an important, and possibly the first, window of integration of word meaning with its presentation context (Serenio et al., 2003; Schindler et al., 2019a). Our data extend this notion from meaning-biasing sentence context (Serenio et al., 2003) and putative social contexts (Schindler et al., 2019a) to mood as an emotional context of word processing. Crucially, results demonstrate that lateralization of word processing is malleable by mood-induction and that these effects further differ between L1 and L2. This novel finding was valence-general and resonates with the results of a recent fmri study that likewise indicated that mood-states affect language lateralization, with stronger left-lateralization in insular cortex in positive mood (Costanzo et al., 2015). Given the topography of the lateral N1,

we observed, the insula might well be one source of this effect. Costanzo et al. (2015) also showed that mood affected language lateralization differently in atypically lateralized people. Given the evidence that L2 often exhibits a different, more rightward lateralization than L1, particularly in related languages (D'Anselmo et al., 2013) as are German and English, the apparently inverted mood effect in L2 is in general agreement with Costanzo et al.'s findings. A differential pattern of right hemisphere (RH) activation in L2 would be consistent with the more widespread neural activation in the RH (e.g., Román et al., 2015; Połczyńska et al., 2017) found especially in the second language of less highly functional bilinguals (as is the case with our L1 dominant participants). Going beyond the language-general effects, additional higher-order interactions suggested that some of the early mood induction effects on neural correlates of word processing in L1 versus L2 may be valence specific, but since follow-up testing could not clearly identify their origin, further research with more participants will be needed to clarify this issue.

EPN

We divided the subsequent EPN in two time-windows, one from 200 to 300 ms and one from 300 to 400 ms to address the possibility of delayed valence processing in L2 (Conrad et al., 2011; Opitz and Degner, 2012). In line with ample previous research (for review see e.g., Citron, 2012), the time-window between 200 and 300 ms, presently scored as early EPN, was emotion sensitive, albeit not reflecting the more often observed u-shaped, arousal-driven pattern, but linear valence discrimination with more negativity elicited by positive than negative words, at least over left occipital areas. Over left hemisphere sites, mood also impacted word processing in a valence-specific manner. Cortical valence differentiation was primarily present after happy mood induction, again in line with the findings by Kiefer et al. (2007). In general, the observed early cortical processing of emotional words was valence-specific, differentiating between positive and negative, which would neither be expected by a motivated attention account (Kuperman, 2015), nor fully in line with automatic vigilance (Pratto and John, 1991), since the early visual attention-sensitive ERPs responded selectively to positive rather than negative words. Previous research on emotional word processing typically revealed arousal-driven ERP modulations (Fischler and Bradley, 2006), at least during free-viewing (e.g., Kissler et al., 2007) or lexical decision (Schacht and Sommer, 2009b). Present results suggest that explicit evaluation accentuates valence-specific perceptual processing, apparently particularly in positive mood as already suggested by Kiefer et al. (2007). Surprisingly, however, the pattern was reversed over temporal regions, and apparently generally in higher-level processing (see below). This theoretically unexpected finding was observed in several time-windows, lending it conceptual credibility. It may reflect the need for alerting by an unexpected input, similar to what is sometimes seen as processing interrupt in the startle literature (Herbert and Kissler, 2010; Blumenthal, 2015). Mood effects over perceptual brain areas, by contrast, exhibited a mood-congruence pattern with larger amplitudes for the mood-congruent words.

The latter portion of the EPN, between 300 and 400 ms, conceptually replicated what was seen in the early EPN as well N1. We found no evidence for delayed valence processing in L2 which would have been evident in an interaction of word valence with language in either of the EPN windows, which might be due to the relatedness of the two languages used. Instead, between 300 and 400 after word onset, over temporal areas, ERPs were generally more negative-going in L1 than in L2, probably reflecting a polarity reversal of the fronto-central N400.

N400

On the N400, a main effect of language was prominent. In line with the ERP literature pointing to N400 as an index of more wide-spread search in language networks (e.g., Kutas and Hillyard, 1980; Kutas and Federmeier, 2000, 2011), we found more negative N400 amplitudes in L2 relative to L1. Larger N400 in L2 than in L1 has been previously observed in word and sentence processing tasks (e.g., Ardal et al., 1990; Moreno and Kutas, 2005; Martin et al., 2013). For instance, in visual processing of words and sentences, bigger N400 amplitudes for L2 stimuli typically have been interpreted as indices of cognitive effort increase, i.e., more extensive lexical search for the L2 word meaning, or more difficulty in integrating L2 word meaning with the representation of the ongoing context (e.g., Moreno and Kutas, 2005; Thierry and Wu, 2007; Martin et al., 2013). Therefore, more negative amplitudes evoked in L2 relative to L1, as we observed here, should indicate of more extensive lexical search in L2 irrespective of mood. German-English bilinguals were employing more cognitive resources to perform the evaluation task in English (L2) than in German (L1). This finding contributes to the body of literature already showing that the N400 amplifications might be qualitatively different in the two languages of bilingual individuals, with factors such as language proficiency, or age of L2 acquisition most likely modulating N400 amplitude. However, unlike shown in previous sentence level (e.g., Federmeier et al., 2001; Pinheiro et al., 2013) or word level (Kiefer et al., 2007) research, no mood effects were present on the N400. This might be due to a combination of word level processing and the evaluative decision task that might have shifted neural mood and content effects in time, perhaps pushing them into earlier negativity or later positivity windows. No effects of word valence were found on this component either, which across emotional word processing studies is not unusual as only some studies report emotion effects on this component (e.g., Sass et al., 2010; Palazova et al., 2013; Zhang et al., 2014).

LPP

LPP amplitude responded to emotional content, being larger for both positive and negative than for neutral words. This accords with a large body of literature on emotional word processing (see e.g., Citron, 2012 for an overview), and is seen particularly during active tasks, requiring attentive processing of emotional content (Schindler and Kissler, 2016). The u-shaped, arousal-driven effect of emotional content is in line with the motivated attention account which is generally influential in

the emotional stimulus perception literature (Lang et al., 1997) and also described by Kuperman (2015) for word processing. Crucially, emotional LPP modulation further varied with mood induction in that LPP amplitude was particularly pronounced for the mood-incongruent word valence. This pattern is similar to what was observed for the lateral N1 and EPN effects and may be in line with the above mentioned alerting by interrupt account. Herring et al. (2011) investigating evaluative affective priming also found that the LPP, but not the N400, responded to the priming manipulation, with the response pattern indicating incongruity-sensitivity on the LPP. The present data extend this pattern from picture and word priming to the effect of experimentally induced moods across blocks of stimuli. Although we have not found language effects in the LPP, a recent EEG study (Kao and Zhang, 2020) points out differences in emotional speech processing between L1 and L2 exactly in the late ERP components—N400 and LPP in the auditory modality. This shows that that language effects for emotional meaning are also modality-related, and future studies need to account not only for mood but also for modality effects when examining how bilinguals process emotional language in their respective linguistic systems.

Limitations and Open Questions

The present study provides evidence for very early effects of mood on lateralization of language processing in L1, as well as of mood on emotion word processing in general. It also replicates several established effects, providing good conceptual credibility for the present findings. Since several observed effects were found in consecutive time-windows, there is also good internal consistency in the data. However, our aim of characterizing the full processing timeline across several time-windows necessitated numerous statistical tests. Therefore, the present findings should be replicated in the future and, if possible, larger groups should be studied. In fact, some early interactions also suggested that early mood effects on L1 versus L2 processing might be valence-specific as we had originally hypothesized. With more experimental power, it should be possible to further specify the nature of these effects. Using different, possibly less related languages may provide a further means of replication, but also help reveal specific effects. Finally, directly contrasting word and sentence-level effects in the same participants would allow us to test whether temporal shifts occur depending on processing load. Early mood effects might be specific to word-level processing and later ones (e.g., in the N400 window) might be found in sentence-level studies.

REFERENCES

- Ardal, S., Donald, M. W., Meuter, R., Muldrew, S. M., and Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain Lang.* 39, 187–205. doi: 10.1016/0093-934X(90)90011-5
- Baumeister, J. C., Foroni, F., Conrad, M., Rumiati, R. I., and Winkielman, P. (2017). Embodiment and emotional memory in first vs. second language. *Front. Psychol.* 8:394. doi: 10.3389/fpsyg.2017.00394
- Bayer, M., Sommer, W., and Schacht, A. (2012). P1 and beyond: functional separation of multiple emotion effects in word recognition. *Psychophysiology* 49, 959–969. doi: 10.1111/j.1469-8986.2012.01381.x
- Beck, A. T., Steer, R. A., and Hautzinger, M. (2001). *Beck-Depressions-Inventar (BDI): Testhandbuch (2., überarb. Aufl., 1. Nachdr.)*. Bern: Huber.
- Blumenthal, T. D. (2015). Presidential address 2014: the more-or-less interrupting effects of the startle response. *Psychophysiology* 52, 1417–1431. doi: 10.1111/psyp.12506

SUMMARY

Overall, we found that moods started modifying emotional word content processing very early, already at N1. This early influence was stronger for happy mood, bigger for L1, relative to L2, and clearly lateralized: left-lateralized for L1 and right-sided, in tendency, for L2, demonstrating language-specific mood effects in the bilingual brain that call for further characterization. Importantly, we found mood-congruent effects in perceptual processes and mood-incongruent ERP amplification during higher order evaluative processing, indicating that the effect of mood on the neurophysiology of language is stage-specific, rather than general. This needs to be taken into account by future models incorporating mood as a context factor in language processing.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JK and KB-D designed and setup the study and wrote the manuscript. JK analyzed the data. Both authors contributed to the article and approved the submitted version.

FUNDING

We acknowledge support for the publication costs by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.

ACKNOWLEDGMENTS

We thank Maria Vogt, Ria Vormbrock, and Philip Lembcke for help with data acquisition and preprocessing.

- Bower, G. H. (1981). Mood and memory. *Am. Psychol.* 36, 129–148. doi: 10.1037/0003-066X.36.2.129
- Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9
- Briesemeister, B. B., Kuchinke, L., and Jacobs, A. M. (2014). Emotion word recognition: discrete information effects first, continuous later? *Brain Res.* 1564, 62–71. doi: 10.1016/j.brainres.2014.03.045
- Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behav. Res. Methods* 46, 904–911. doi: 10.3758/s13428-013-0403-5
- Caldwell-Harris, C. (2015). Emotionality differences between a native and foreign language: implications for everyday life. *Cur. Dir. Psychol. Sci.* 24, 214–219. doi: 10.1177/0963721414566268
- Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain Lang.* 122, 211–226. doi: 10.1016/j.bandl.2011.12.007
- Citron, F. M., Weekes, B. S., and Ferstl, E. C. (2013). Effects of valence and arousal on written word recognition: time course and ERP correlates. *Neurosci. Lett.* 533, 90–95. doi: 10.1016/j.neulet.2012.10.054
- Clore, G. L., and Huntsinger, J. R. (2009). How emotions inform judgment and regulate thought. *Trends Cogn. Sci.* 11, 393–399. doi: 10.1016/j.tics.2007.08.005
- Cohen, J. (ed.). (1988). “The effect size index: d,” in *Statistical Power Analysis for the Behavioral Sciences*, Vol. 2, (New York, NY: Routledge).
- Conrad, M., Recio, G., and Jacobs, A. M. (2011). The time course of emotion effects in first and second language processing: a cross cultural ERP study with German-Spanish bilinguals. *Front. Psychol.* 2:351. doi: 10.3389/fpsyg.2011.00351
- Costa, A., and Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nat. Rev. Neurosci.* 15, 336–345. doi: 10.1038/nrn3709
- Costanzo, E. Y., Villarreal, M., Drucaroff, L. J., Ortiz-Villafañe, M., Castro, M. N., Goldschmidt, M., et al. (2015). Hemispheric specialization in affective responses, cerebral dominance for language, and handedness Lateralization of emotion, language, and dexterity. *Behav. Brain Res.* 288, 11–19. doi: 10.1016/j.bbr.2015.04.006
- D’Anselmo, A., Reiterer, S., Zuccarini, F., Tommasi, L., and Brancucci, A. (2013). Hemispheric asymmetries in bilinguals: tongue similarity affects lateralization of second language. *Neuropsychologia* 51, 1187–1194. doi: 10.1016/j.neuropsychologia.2013.03.016
- De Groot, A. M. (2011). *Language and Cognition in Bilinguals and Multilinguals: An Introduction*. Hove: Psychology Press.
- De Pascalis, V., Arwari, B., D’Antuono, L., and Cacace, I. (2009). Impulsivity and semantic/emotional processing: an examination of the N400 wave. *Clin. Neurophysiol.* 120, 85–92. doi: 10.1016/j.clinph.2008.10.008
- Dehaene, S. (1995). Evidence for category-specific word processing in the normal human brain. *NeuroReport* 6, 2153–2157. doi: 10.1097/00001756-199511000-00014
- Federmeier, K. D., Kirson, D. A., Moreno, E. M., and Kutas, M. (2001). Effects of transient, mild mood states on semantic memory organization and use: an event-related potential investigation in humans. *Neurosci. Lett.* 305, 149–152. doi: 10.1016/S0304-3940(01)01843-2
- Ferraro, F. R., King, B., Ronning, B., Pekarski, K., and Risan, J. (2003). Effects of induced emotional state on lexical processing in younger and older adults. *J. Psychol.* 137, 262–272. doi: 10.1080/00223980309600613
- Fiedler, K. (2001). “Affective states trigger processes of assimilation and accommodation,” in *Theories of Mood and Cognition: A User’s Guide*, eds L. L. Martin, and G. L. Clore, (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 85–98.
- Fischler, I., and Bradley, M. (2006). Event-related potential studies of language and emotion: words, phrases, and task effects. *Prog. Brain Res.* 156, 185–203. doi: 10.1016/S0079-6123(06)56009-1
- Gross, J. J., and Levenson, R. W. (1995). Emotion elicitation using films. *Cogni. Emot.* 9, 87–108. doi: 10.1080/02699939508408966
- Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., and Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage* 30, 1383–1400. doi: 10.1016/j.neuroimage.2005.11.048
- Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J., and Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychol. Sci.* 28, 1387–1397. doi: 10.1177/095679761772094
- Heister, J., Würzner, K. M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., et al. (2011). dlexDB—eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychol. Rundsch.* 62, 10–20. doi: 10.1026/0033-3042/a000029
- Herbert, C., Junghofer, M., and Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiol* 45, 487–498. doi: 10.1111/j.1469-8986.2007.00638.x
- Herbert, C., and Kissler, J. (2010). Motivational priming and processing interrupt: startle reflex modulation during shallow and deep processing of emotional words. *Int. J. Psychophysiol.* 76, 64–71. doi: 10.1016/j.ijpsycho.2010.02.004
- Herbert, C., Kissler, J., Junghofer, M., Peyk, P., and Rockstroh, B. (2006). Processing of emotional adjectives: evidence from startle EMG and ERPs. *Psychophysiol* 43, 197–206. doi: 10.1111/j.1469-8986.2006.00385.x
- Herring, D. R., Taylor, J. H., White, K. R., and Crites, S. L. Jr. (2011). Electrophysiological responses to evaluative priming: the LPP is sensitive to incongruity. *Emotion* 11, 794–806. doi: 10.1037/a0022804
- Herz, N., Baror, S., and Bar, M. (2020). Overarching states of mind. *Trends Cogn. Sci.* 24, 184–199. doi: 10.1016/j.tics.2019.12.015
- Higgins, E. T. (1998). The aboutness principle: a pervasive influence on human inference. *Soc. Cogn.* 16, 173–198. doi: 10.1521/soco.1998.16.1.173
- Hinojosa, J. A., Moreno, E. M., and Ferré, P. (2019). Affective neurolinguistics: towards a framework for reconciling language and emotion. *Lang. Cogn. Neurosci.* 35, 813–839. doi: 10.1080/23273798.2019.1620957
- Hofmann, M. J., Kuchinke, L., Tamm, S., Vo, M. L., and Jacobs, A. M. (2009). Affective processing within 1/10th of a second: high arousal is necessary for early facilitative processing of negative but not positive words. *CABN* 9, 389–397. doi: 10.3758/9.4.389
- Holt, D., Lynn, S., and Kuperberg, G. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *J. Cogn. Neurosci.* 21, 2245–2262. doi: 10.1162/jocn.2008.21151
- Ille, N., Berg, P., and Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *J. Clin. Neurophysiol.* 19, 113–124. doi: 10.1097/00004691-200203000-00002
- Jończyk, R., Boutonnet, B., Musiał, K., Hoemann, K., and Thierry, G. (2016). The bilingual brain turns a blind eye to negative statements in the second language. *Cogn. Affect. Behav. Neurosci.* 16, 527–540. doi: 10.3758/s13415-016-0411-x
- Jończyk, R., Korolczuk, I., Balatsou, E., and Thierry, G. (2019). Keep calm and carry on: electrophysiological evaluation of emotional anticipation in the second language. *Cogn. Affect. Behav. Neurosci.* 14, 885–898. doi: 10.1093/scan/nsz066
- Kao, C., and Zhang, Y. (2020). Differential neurobehavioral effects of cross-modal selective priming on phonetic and emotional prosodic information in late second language learners. *J. Speech Lang. Hear. Res.* 63, 2508–2521. doi: 10.1044/2020_jslhr-19-00329
- Keysar, B., Sayuri, L., Hayakawa, S., and An, S. (2012). The foreign-language effect: thinking in a foreign tongue reduces decision biases. *Psychol. Sci.* 23, 661–668. doi: 10.1177/0956797611432178
- Kharkhurin, A., and Altarriba, J. (2016). The effect of mood induction and language of testing on bilingual creativity. *Bilingualism* 19, 1079–1094. doi: 10.1017/S1366728915000528
- Kiefer, M., Schuch, S., Schenck, W., and Fiedler, K. (2007). Mood states modulate activity in semantic brain areas during emotional word encoding. *Cereb. Cortex* 17, 1516–1530. doi: 10.1093/cercor/bhl062
- Kissler, J., Assadollahi, R., and Herbert, C. (2006). Emotional and semantic networks in visual word processing: insights from ERP studies. *Prog. Brain Res.* 156, 147–183. doi: 10.1016/S0079-6123(06)56008-X
- Kissler, J., and Herbert, C. (2013). Emotion, etmnooi, or emittoon?—faster lexical access to emotional than to neutral words during reading. *Biol. Psychol.* 92, 464–479. doi: 10.1016/j.biopsycho.2012.09.004
- Kissler, J., Herbert, C., Peyk, P., and Junghofer, M. (2007). Buzzwords: early cortical responses to emotional words during reading. *Psychol. Sci.* 18, 475–480. doi: 10.1111/j.1467-9280.2007.01924.x
- Kissler, J., Herbert, C., Winkler, I., and Junghöfer, M. (2009). Emotion and attention in visual word processing – An ERP study. *Biol. Psychol.* 80, 75–83. doi: 10.1016/j.biopsycho.2008.03.004

- Klauer, K. C., and Musch, J. (2003). Affective priming: findings and theories. *Psychol. Eval.* 7, 1–50.
- Kuperman, V. (2015). Virtual experiments in megastudies: a case study of language and emotion. *Q. J. Exp. Psychol.* 68, 1693–1710. doi: 10.1080/17470218.2014.989865
- Kutas, M., and Federmeier, F. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470. doi: 10.1016/s1364-6613(00)01560-6
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Ann. Rev. Psychol.* 62, 621–641. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., and Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biol. Psychol.* 11, 99–116. doi: 10.1016/0301-0511(80)90046-0
- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). “Motivated attention: affect, activation, and action,” in *Attention and Orienting: Sensory and Motivational Processes*, eds P. J. Lang, R. F. Simons, and M. T. Balaban, (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 97–135.
- Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm, A. O. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30, 261–273. doi: 10.1111/j.1469-8986.1993.tb03352.x
- Lemhöfer, K., and Broersma, M. (2012). Introducing LexTALE: a quick, and valid lexical test for advanced learners of English. *Behav. Res. Methods* 44, 325–343. doi: 10.3758/s13428-011-0146-0
- Li, P., Zhang, F., Tsai, E., and Puls, B. (2014). Language history questionnaire (LHQ 2.0): a new dynamic web-based research tool. *Bilingualism* 17, 673–680. doi: 10.1017/s1366728913000606
- Marian, V., Bartolotti, J., Chabal, S., and Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One* 7:e43230. doi: 10.1371/journal.pone.0043230
- Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., and Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *J. Mem. Lang.* 69, 574–588. doi: 10.1016/j.jml.2013.08.001
- Matheson, H. E., and Barsalou, L. W. (2018). “Embodiment and grounding in cognitive neuroscience,” in *The Stevens’ Handbook of Experimental Psychology and Cognitive Neuroscience*, 4th Edn, eds J. Wixted, E. Phelps, L. Davachi, J. Serences, S. Ghetti, S. Thompson-Schill, et al. (Hoboken, NJ: Wiley), 1–32. doi: 10.1002/9781119170174.epcn310
- Moreno, E. M., and Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish-English bilinguals. *Cogn. Brain Res.* 22, 205–220. doi: 10.1016/j.cogbrainres.2004.08.010
- Moreno, E. M., and Vázquez, C. (2011). Will the glass be half full or half empty? Brain Potentials and emotional expectations. *Biol. Psychol.* 88, 131–140. doi: 10.1016/j.biopsycho.2011.07.003
- Niedenthal, P. M., Halberstadt, J. B., and Setterlund, M. B. (1997). Being happy and being “happy”: emotional state facilitates visual encoding. *Cogn. Emot.* 11, 403–432. doi: 10.1080/026999397379863
- Niedenthal, P. M., and Setterlund, M. B. (1994). Emotion congruence in perception. *Pers. Soc. Psychol. Bull.* 20, 401–441. doi: 10.1177/0146167294204007
- Niedenthal, P. M., Setterlund, M. B., and Jones, D. E. (1994). “Emotional organization of perceptual memory,” in *The Heart’s Eye. Emotional Influences in Perception and Attention*, eds P. Niedenthal, and S. Kitayama, (London: Academic Press), 87–113. doi: 10.1016/b978-0-12-410560-7.50011-9
- Olafson, K. M., and Ferraro, F. R. (2001). Effects of emotional state on lexical decision performance. *Brain Cogn.* 45, 15–20. doi: 10.1006/brcg.2000.1248
- Opitz, B., and Degner, J. (2012). Emotionality in a second language: it’s a matter of time. *Neuropsychologia* 50, 1961–1967. doi: 10.1016/j.neuropsychologia.2012.04.021
- Palazova, M., Mantwill, K., Sommer, W., and Schacht, A. (2011). Are effects of emotion in single words non-lexical? Evidence from event-related brain potentials. *Neuropsychologia* 49, 2766–2775. doi: 10.1016/j.neuropsychologia.2011.06.005
- Palazova, M., Sommer, W., and Schacht, A. (2013). Interplay of emotional valence and concreteness in word processing: an event-related potential study with verbs. *Brain Lang.* 125, 264–271. doi: 10.1016/j.bandl.2013.02.008
- Pavlenko, A. (2012). Affective processing in bilingual speakers: disembodied cognition? *Int. J. Psychol.* 47, 405–428. doi: 10.1080/00207594.2012.743665
- Peyk, P., De Cesare, A., and Junghofer, M. (2011). ElectroMagnetoEncephalography software: overview and integration with other EEG/MEG toolboxes. *Comput. Intell. Neurosci.* 2011:861705.
- Pinheiro, A. P., del Re, E., Nestor, P. G., McCarley, R. W., Gonçalves, O. F., and Niznikiewicz, M. (2013). Interactions between mood and the structure of semantic memory: event-related potentials evidence. *SCAN* 8, 579–594. doi: 10.1093/scan/nss035
- Półczyńska, M., Japardi, K., and Bookheimer, S. (2017). Lateralizing language function with pre-operative functional magnetic resonance imaging in early proficient bilingual patients. *Brain Lang.* 170, 1–11. doi: 10.1016/j.bandl.2017.03.002
- Ponari, M., Norbury, C. F., and Vigliocco, G. (2017). Acquisition of abstract concepts is influenced by emotional valence. *Dev. Sci.* 10, 10–11. doi: 10.1111/desc.12549
- Pratto, F., and John, O. P. (1991). Automatic vigilance: the attention-grabbing power of negative social information. *J. Pers. Soc. Psychol.* 61, 380–391. doi: 10.1037/0022-3514.61.3.380
- Román, P., González, J., Ventura-Campos, N., Rodríguez-Pujadas, A., Sanjuán, A., and Ávila, C. (2015). Neural differences between monolinguals and early bilinguals in their native language during comprehension. *Brain Lang.* 150, 80–89. doi: 10.1016/j.bandl.2015.07.011
- Sass, S. M., Heller, W., Stewart, J. L., Siltan, R. L., Edgar, J. C., Fisher, J. E., et al. (2010). Time course of attentional bias in anxiety: emotion and gender specificity. *Psychophysiol* 47, 247–259. doi: 10.1111/j.1469-8986.2009.00926.x
- Schacht, A., and Sommer, W. (2009a). Emotions in word and face processing: early and late cortical responses. *Brain Cogn.* 69, 538–550. doi: 10.1016/j.bandc.2008.11.005
- Schacht, A., and Sommer, W. (2009b). Time course and task dependence of emotion effects in word processing. *CABN* 9, 28–43. doi: 10.3758/cabn.9.1.28
- Schindler, S., and Kissler, J. (2016). People matter: perceived sender identity modulates cerebral processing of socio-emotional language feedback. *Neuroimage* 134, 160–169. doi: 10.1016/j.neuroimage.2016.03.052
- Schindler, S., Miller, G. A., and Kissler, J. (2019a). Attending to Eliza: rapid brain responses reflect competence attribution in virtual social feedback processing. *Soc. Cogn. Affect. Neurosci.* 14, 1073–1086. doi: 10.1093/scan/nsz075
- Schindler, S., Vormbrock, R., and Kissler, J. (2019b). Emotion in context: how sender predictability and identity affect processing of words as imminent personality feedback. *Front. Psychol.* 10:94. doi: 10.3389/fpsyg.2019.00094
- Scott, G. G., O’Donnell, P. J., Leuthold, H., and Sereno, S. C. (2009). Early emotion word processing: evidence from event-related potentials. *Biol. Psychol.* 80, 95–104. doi: 10.1016/j.biopsycho.2008.03.010
- Sereno, S. C., Brewer, C. C., and O’Donnell, P. J. (2003). Context effects in word recognition: evidence for early interactive processing. *Psychol. Sci.* 14, 328–333.
- Sereno, S. C., and Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cog. Sci.* 7, 489–493. doi: 10.1016/j.tics.2003.09.010
- Sereno, S. C., Rayner, K., and Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport* 9, 2195–2200. doi: 10.1097/00001756-199807130-00009
- Sereno, S. C., Scott, G. G., Yao, B., Thaden, E. J., and O’Donnell, P. J. (2015). Emotion word processing: does mood make a difference? *Front. Psychol.* 6:1191. doi: 10.3389/fpsyg.2015.01191
- Sheikh, N. A., and Titone, D. (2016). The embodiment of emotional words in a second language: an eye-movement study. *Cogn. Emot.* 30, 488–500. doi: 10.1080/02699931.2015.1018144
- Spielberger, C. D., Sydeman, S. J., Owen, A. E., and Marsh, B. J. (1999). “Measuring anxiety and anger with the state-trait anxiety inventory (stai) and the state-trait anger expression inventory (STAXI),” in *The use of Psychological Testing for Treatment Planning and Outcomes Assess*, ed. M. E. Maruish, (New Jersey: Lawrence Erlbaum Associates Publishers), 993–1021.

- Thierry, G., and Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *PNAS* 104, 12530–12535. doi: 10.1073/pnas.0609927104
- Van Berkum, J. J., Hagoort, P., and Brown, C. M. (1999). Semantic integration in sentences and discourse: evidence from the N400. *J. Cogn. Neurosci.* 11, 657–671. doi: 10.1162/089892999563724
- Van Berkum, J. J. A. (2018). “Language comprehension, emotion, and sociality: aren’t we missing something?” in *The Oxford Handbook of Psycholinguistics*, eds Rueschemeyer, and G. Gaskell, (Oxford: Oxford University Press), 644–669.
- Van Berkum, J. J. A. (2019). “Language comprehension and emotion: where are the interfaces and who cares?” in *Oxford Handbook of Psycholinguistics*, eds G. de Zubicaray, and N. O. Schiller, (Oxford: Oxford University Press), 736–766.
- van Heuven, W., Keuleers, E., and Brysbaert, M. (2014). SUBTLEX-UK: a new and improved word frequency database for British English. *Q. J. Exp. Psychol.* 67, 1176–1190. doi: 10.1080/17470218.2013.850521
- Võ, M. L., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., and Jacobs, A. M. (2009). The Berlin affective word list reloaded (BAWL-R). *Behav. Res. Methods* 41, 534–538. doi: 10.3758/brm.41.2.534
- Warriner, A., Kuperman, V., and Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res.* 45, 1191–1207. doi: 10.3758/s13428-012-0314-x
- Wu, Y. J., and Thierry, G. (2012). How reading in a second language protects your heart. *J. Neurosci.* 32, 6485–6489. doi: 10.1523/JNEUROSCI.6119-11.2012
- Yao, Z., Yu, D., Wang, L., Zhu, X., Guo, J., and Wang, Z. (2016). Effects of valence and arousal on emotional word processing are modulated by concreteness: behavioral and ERP evidence from a lexical decision task. *Int. J. Psychophys.* 110, 231–242. doi: 10.1016/j.ijpsycho.2016.07.499
- Zajonc, R. B. (1980). Feeling and thinking. Preferences need no inferences. *Am. Psychol.* 35, 151–175. doi: 10.1037/0003-066x.35.2.151
- Zhang, D., He, W., Wang, T., Luo, W., Zhu, X., Gu, R., et al. (2014). Three stages of emotional word processing: an ERP study with rapid serial visual presentation. *SCAN* 9, 1897–1903. doi: 10.1093/scan/ns/t188

Conflict of Interest: 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 © 2021 Kissler and Bromberek-Dyzman. 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 to Use Narrative Function Words for the Organization and Communication of Experience

Gregoire Pointeau^{1,2}, Solène Miriaz^{1,2,3}, Anne-Laure Mealier^{1,2} and Peter Ford Dominey^{1,2*}

¹ INSERM UMR 1093-CAPS, Université Bourgogne Franche-Comté, UFR des Sciences du Sport, Dijon, France, ² Robot Cognition Laboratory, Marey Institute, Dijon, France, ³ École Normale Supérieure de Rennes, Bruz, France

OPEN ACCESS

Edited by:

Ramesh Kumar Mishra,
University of Hyderabad, India

Reviewed by:

Eva Wennäs Brante,
Malmö University, Sweden
Alistair Knott,
University of Otago, New Zealand

*Correspondence:

Peter Ford Dominey
peter.dominey@inserm.fr

Specialty section:

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

Received: 11 August 2020

Accepted: 22 January 2021

Published: 03 March 2021

Citation:

Pointeau G, Miriaz S, Mealier A-L
and Dominey PF (2021) Learning
to Use Narrative Function Words
for the Organization
and Communication of Experience.
Front. Psychol. 12:591703.
doi: 10.3389/fpsyg.2021.591703

How do people learn to talk about the causal and temporal relations between events, and the motivation behind why people do what they do? The narrative practice hypothesis of Hutto and Gallagher holds that children are exposed to narratives that provide training for understanding and expressing reasons for why people behave as they do. In this context, we have recently developed a model of narrative processing where a structured model of the developing situation (the situation model) is built up from experienced events, and enriched by sentences in a narrative that describe event meanings. The main interest is to develop a proof of concept for how narrative can be used to structure, organize and describe experience. Narrative sentences describe events, and they also define temporal and causal relations between events. These relations are specified by a class of narrative function words, including “because, before, after, first, finally.” The current research develops a proof of concept that by observing how people describe social events, a developmental robotic system can begin to acquire early knowledge of how to explain the reasons for events. We collect data from naïve subjects who use narrative function words to describe simple scenes of human-robot interaction, and then employ algorithms for extracting the statistical structure of how narrative function words link events in the situation model. By using these statistical regularities, the robot can thus learn from human experience about how to properly employ in question-answering dialogues with the human, and in generating canonical narratives for new experiences. The behavior of the system is demonstrated over several behavioral interactions, and associated narrative interaction sessions, while a more formal extended evaluation and user study will be the subject of future research. Clearly this is far removed from the power of the full blown narrative practice capability, but it provides a first step in the development of an experimental infrastructure for the study of socially situated narrative practice in human-robot interaction.

Keywords: narrative, situation model, discourse marker, reservoir computing, narrative practice

INTRODUCTION

Meaning is grounded in social and cultural conventions expressed in the forms of words (Waxman and Markow, 1995), grammatical constructions (Goldberg, 2003; Tomasello, 2003), and narrative patterns (Bruner, 1991; Hutto, 2007) that are elaborated through shared experience. Theories of narrative practice hold that through normal exposure to narratives about human social interaction,

the child will come to learn how to interpret, react to and respond to social contexts as provided by a theory of mind or folk psychology (Hutto, 2007; Gallagher and Hutto, 2008; Nelson, 2009). This narrative practice theory holds that in human interaction, people regularly generate folk psychological narratives that explain why a person acted on a particular occasion, and that through exposure to these narratives children acquire the skills to understand and themselves produce such narratives (Hutto, 2007). This provides an answer to questions of socially situated language learning - To what extent do representations gleaned from the social and cultural context influence language processing and learning? What mechanisms contribute to socially-situated language processing and learning? The current research provides a theory of how exposure to situations and language describing those situations can be used to establish norms about how language should be used to describe and answer questions about these situations. One method to demonstrate the feasibility of such a theory is in the context of social interaction between humans and robots. The objective of the research described in this paper is to spell out a framework for implementing the theory, and to establish its feasibility in a proof of concept demonstration, leaving more formal and extended user studies for the future.

Spoken language has historically played an important role in interactive robot-human communication (Crangle and Suppes, 1994; Lauria et al., 2002; Dominey et al., 2009; Kollar et al., 2010; Matuszek et al., 2013). In the most direct usage, language allows the robot to describe events that have just occurred (Dominey and Boucher, 2005), and allows the human to command actions that the robot should perform (Dominey et al., 2007a,b, 2009). Extending the usage of language in time, we have used spoken language to allow the human to explain a coordinated, cooperative shared plan to the robot, and then to help explain and show the robot how to perform the different actions in the shared plan (Petit et al., 2013; Sorce et al., 2015) as illustrated in **Figure 1**.

Ideally, however, language allows a much more extended access to events and relations between events and the mental states of the agents involved, as those events occur in extended time. This more extended use of language brings us to something approaching narrative. In her characterization how the child begins to go beyond purely canonical representations of its life events, Nelson states that "Narrative is the vehicle of communicating representations of events between people by verbal means." [(Nelson, 2003), p. 32]. Nelson specifies that narrative processing requires a grammatical processing capability sufficient to handle the complexity of the sentences used in the narrative, a form of working memory that allows the construction of a representation of the unfolding story, and appropriate experiential memory for encoding and interpreting the situations that the story refers to Nelson (2009).

That is, language is about something, and this something is the shared experience of the participants. In this context, we have made a significant effort to develop an autobiographical memory (ABM) system that allows the iCub humanoid robot to store its experience with humans, and to organize this experience in pertinent manner, thus allowing the iCub to learn and perform shared plans for joint action (Petit et al., 2013; Pointeau et al.,

2014; Moulin-Frier et al., 2017), as illustrated in **Figure 1**. This ABM system thus contributes in part to Nelson's requirement for experiential memory. The requirement for grammar processing can be met with our work in dynamic construction grammar (DCG) (Hinault and Dominey, 2013; Hinault et al., 2014, 2015; Dominey et al., 2017). These ABM and DCG capabilities have been integrated in a model of narrative processing (Mealier et al., 2017), illustrated in **Figure 2**.

Further responding to Nelson's requirements, we have developed a system where a situation model (Zwaan and Radvansky, 1998; Zwaan and Madden, 2004) is assembled from events coded in the ABM, and is then enriched by linking event representations with causal and temporal relations that are coded by narrative function words. This extends our work on dynamic construction grammar (DCG) using recurrent reservoir networks for sentence processing (Hinault and Dominey, 2013; Hinault et al., 2014, 2015). These models are called Dynamic Construction Grammar because of the internal dynamics of the recurrent reservoir network that produces on-line dynamic responses to model inputs, as required for simulating ERP responses (Hinault and Dominey, 2013). These reservoir computing models learn the relation between the structure of sentences, and meaning, as the mapping of semantic words in the sentences (nouns and verbs) onto their semantic roles of predicate, agent, object and recipient (PAOR). This corresponds to the elements in the Narrative Cx Model in **Figure 2**.

This model of narrative processing will form the core infrastructure for our study of narrative practice. In the following we outline the extension from grammar to narrative, the elaboration of the situation model, and the use of narrative function words to express relations between event components within the situation model. Then we demonstrate the proof of concept for the learning of how to use narrative function words in responding to questions and in the generation of canonical narrative patterns.

From Grammatical Construction to Narrative Construction

The extension from the original grammatical construction models is based on the introduction of narrative function words in the sentences, and corresponding narrative relations in the meaning. Whereas grammatical function words (e.g., *to*, *by*, *was*) specify relations between open class words and their semantic roles within a sentence - e.g., who did what to whom - narrative function words specify relations between events in multiple sentences, and their constituent elements at the level of the situation model - e.g., why someone did something to someone.

The original DCG models allowed the learning of the mapping between event meaning and sentences. We then introduce the notion of narrative relations into the meaning. So the sentence "I gave you the toy because you wanted it" corresponds to the meaning with two events gave (I, you, toy) and want (you, toy), linked by the causal relation *because*. This new component of the meaning is labeled "Narrative Relations" in **Figure 2**. Thus, the recurrent neural network and readout learns to extract the *predicate* (*agent*, *object*, *recipient*) (PAOR) representations of events, and the narrative relations. This

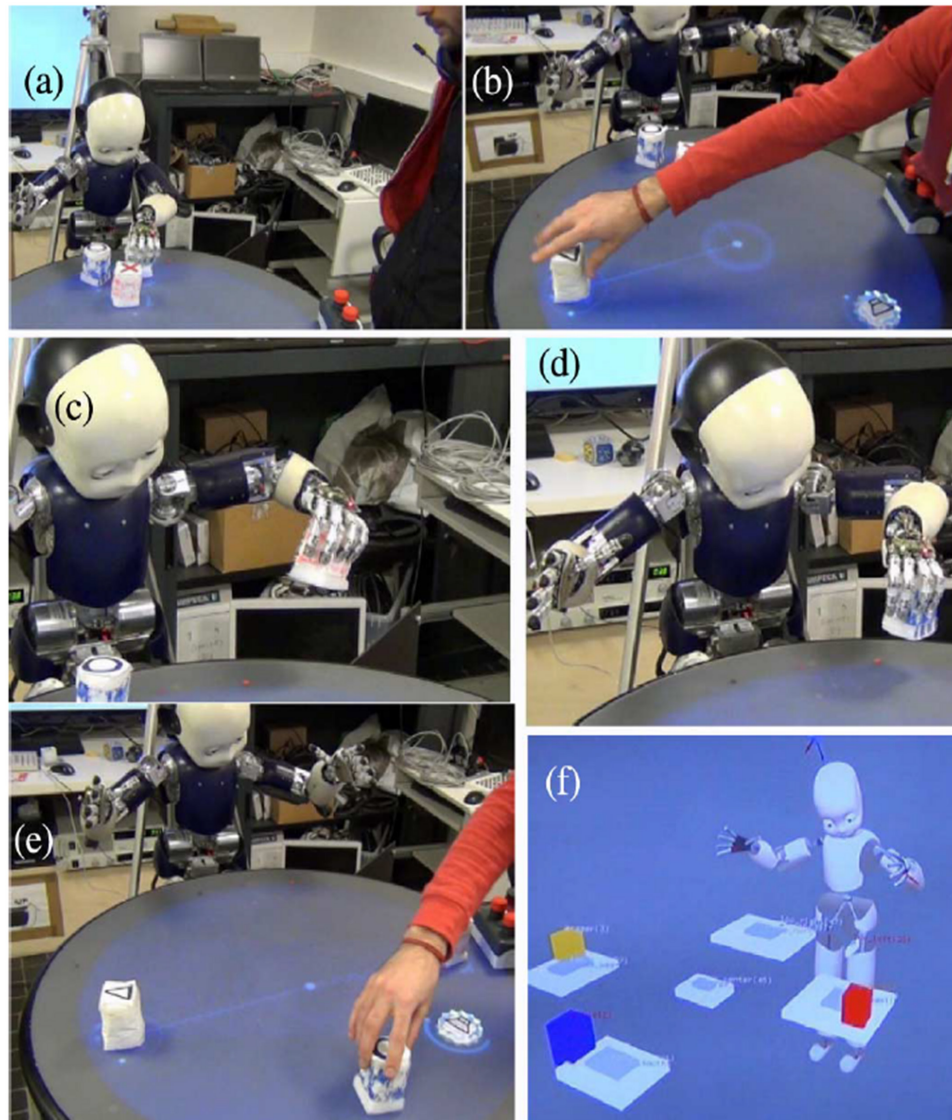
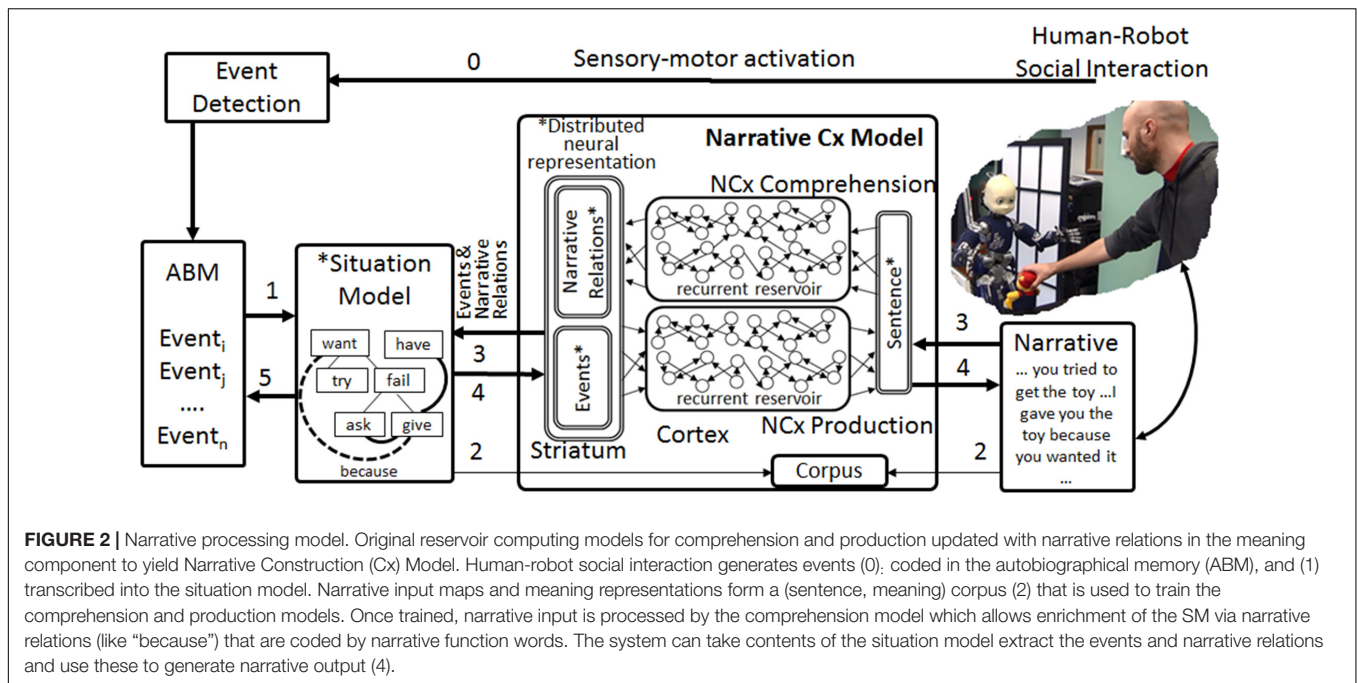


FIGURE 1 | Human-robot social interaction - joint execution of a shared plan that is learned from experience and coded in the autobiographical memory (ABM). Different steps of the iCub during the execution of a shared plan for the music game are illustrated. **(a)** Initial configuration of 3 elements. Robot places object 1 north. **(b)** Human takes this object and places it west. **(c)** Robot places object 2 North, for the human, who then puts it East (not shown). **(d)** Robot places final object north. **(e)** Human takes object and places it South. **(f)** Final internal representation of objects on ReacTable to produce the song as the joint goal of the shared plan. From Pointeau et al. (2014a).

is the content that can now be constructed into a coherent representation of the narrative, the situation model, based on the narrative construction.

The narrative construction is compositional, built up from multiple sentences that are linked by relations along these dimensions. The nature of such relations and their representation has been identified in various discourse models, such as Centering Theory (Grosz and Sidner, 1986; Grosz et al., 1995), rhetorical structure theory (Mann and Thompson, 1988), SDRT (Lascarides and Asher, 1993), or coherence and structure of discourse (Hobbs, 1985). Taking the analogy from grammatical constructions, these relations are coded by the order of the

sentences and by narrative function words (e.g., but, since, then, so, now, because, etc.). The crucial notion is that narrative structure provides a higher level of organization upon the events that it describes. New links—causal, intentional, temporal, etc., and aspects of meaning about people and events that may breach the canonical structure—are superimposed on the events by the narrative discourse, and this structuring results in the creation of meaning referred to by Bruner (1990, 1991, 2009). It is likely that there is a constructive interaction between pre-linguistic representations of such links, and language that labels and highlights these links as the child becomes increasingly proficient (Bruner, 2009).



We have developed methods for representing and expressing meaning about physical events in grammatical constructions (Dominey and Boucher, 2005; Hinaut et al., 2014). The constructions are learned in a manner similar to how humans communicate such meaning in sentences. Paired <sentence, meaning> corpora are created, and used to train the comprehension and production models. This form-meaning learning can be extended to narrative constructions, which allow humans to communicate meaning about a group of events that occurred in a coherent behavioral context, and importantly to express relations between events that may not be visible. Where the grammatical construction uses word order and grammatical functions words to map open class elements onto their thematic roles, the narrative construction uses sentence order and narrative function words to map multiple sentences onto events and relations between them. The form pole of the narrative construction is thus composed of a sequence of sentences that are linked via narrative function words—much like the grammatical function words (closed class words) that provide the grammatical structure at the sentence level (Mealier et al., 2017). Narrative function words have been characterized as discourse connectives which provide discourse structure (Grosz and Sidner, 1986; Knott, 1996; Knott and Sanders, 1998; Fraser, 1999; Webber et al., 2001), much like grammatical function words (closed class words) provide grammatical structure at the sentence level. Norrick (2001) shows how discourse markers “well” and “but” can take on special narrative functions distinct from their lexical meanings and usual discourse marker functions, supporting the psychological validity of the notion of narrative function word.

Narrative constructions are thus learned as conventions (Hutto, 2007), in the same way that grammatical constructions are learned as conventions. As with the grammatical construction model, the system must be furnished with matched

sentence-meaning pairs. The novelty is that these sentences will include narrative function words, whose role will also be reflected in the meaning representation. That is, they will be intrinsically present in the sequential structure of sentences and in the meaning representations in training corpora, and learned by the system. Crucially, however, as mentioned above, there may be components of the narrative structure that are not visible in the physical events, e.g., causal and logical relations. These relations will be introduced by the narrator in the narrative examples. This is part of how narrative is used to make meaning (Bruner, 1990, 1991), including the construction of the situation model.

The Situation Model

A narrative construction maps multiple sentences onto a situation model, specified as a network of these PAORs (predicate, agent, object, recipient frames), linked by relations along the five dimensions of Zwaan and Radvansky (1998): time, space, causation, motivation, and protagonist. Inspired by psycholinguistics (Zwaan and Radvansky, 1998; Zwaan and Madden, 2004), our situation model codes events, organized around an event structure with Initial state, Goal, Action, Result and Final state – IGARF. These events are linked with narrative relations (causal, temporal, intentional) from successive sentences in the narrative. Recalling from above, this involves an extension of the notion of grammatical construction to narrative construction which in turn involves the introduction of the notion of narrative function words. In analogy to the way in which grammatical function words operate on relations between open class words in a sentence, narrative function words operate on relations between events in a situation model (Dominey et al., 2017; Mealier et al., 2017). Narrative function words including “because, since, then, so, before, after” allow the construction of relations between events in

order to construct and enrich a situation model representation of meaning. A detailed situation model in IGARF format is illustrated in **Figure 3**.

The situation model addresses a major issue we had to resolve which concerned how the DCG model could accommodate multiple sentences that are linked by their narrative structure and contribute to the construction of a coherent meaning representation. The solution was to extend the meaning pole of

the DCG model. As illustrated in **Figure 2**, the DCG models have the meaning pole that continues to contain a representation of the events described in the sentence. In addition to coding the predicate-argument representation of the events, the meaning component is supplemented with an optional representation of the narrative context as coded by a narrative function word. This is indicated as narrative Relations in **Figure 2**. For example, in the sentence “I gave you the toy because you wanted it,”

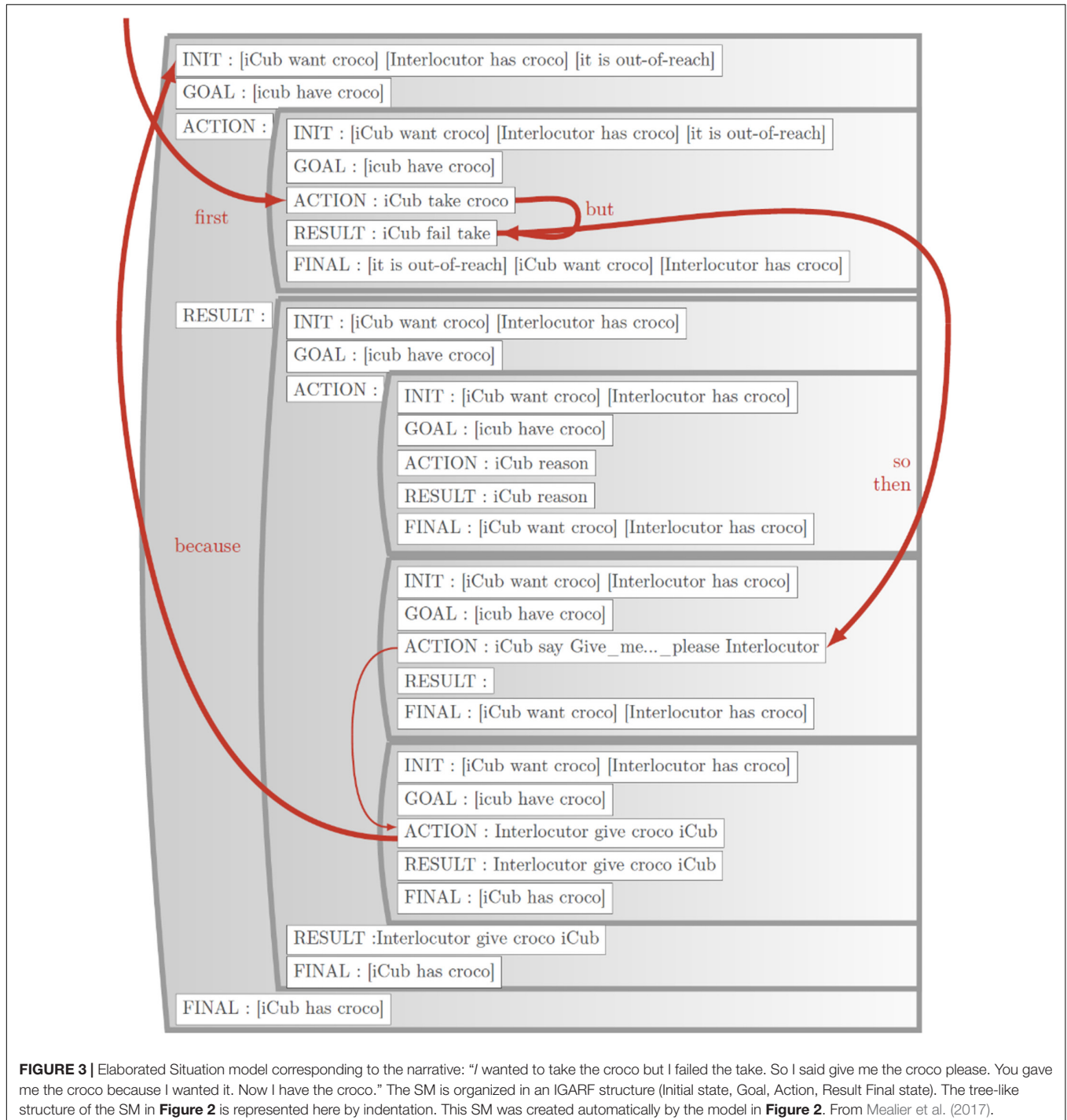


FIGURE 3 | Elaborated Situation model corresponding to the narrative: “I wanted to take the croco but I failed the take. So I said give me the croco please. You gave me the croco because I wanted it. Now I have the croco.” The SM is organized in an IGARF structure (Initial state, Goal, Action, Result Final state). The tree-like structure of the SM in **Figure 2** is represented here by indentation. This SM was created automatically by the model in **Figure 2**. From Mealier et al. (2017).

the meaning component is the standard predicate-agent-object-recipient (PAOR) of the two events gave (I, you, toy); want (you, toy), and the narrative relations component indicates the narrative function word that is now linked to these events. This link is then added to the situation model, as illustrated by the dotted line in **Figure 2**, and in the more detailed SM representation in **Figure 3**.

An important aspect of the situation model is that it provides a form of convergence zone between non-linguistic event representations - which can be internal representations of one's own actions - and linguistic representations of those same actions. **Figure 3** provides details of a situation model where the initial content was generated from a human-robot interaction, and the SM was then completed by human narration.

Integrated Function

In usage, the human and robot [here the iCub (Metta et al., 2008)] interact based on a complication and resolution scenario, where the human helps the robot to achieve its goal. For example, the robot wants a toy. It tries to grasp the toy, and fails. It then reasons on other actions that could be used to achieve the goal, and asks the human for help. The human gives the toy to the robot. These events, as generated by the robot, are coded in its Autobiographical Memory, and automatically converted into the SM representation. This yields an initial SM. The human then narrates what happened, which enriches the representation that has been initiated in the SM. Each sentence in the narrative is matched to the event that it describes. The resulting sentence-meaning pairs are assembled into a corpus that is then used to train the narrative construction (NCx) comprehension and production models. After the comprehension and production models are trained on the resulting corpus, the trained comprehension model can be

used to extract the meaning from the narrative. This extracts the events, which are assembled into a situation model (or used to enrich the existing SM), and narrative relations, that are used to create links between events (illustrated as the dotted link “because” between the give and the want actions in **Figure 2**, and the narrative links illustrated in **Figure 3**). Narrative relations are identified as those semantic elements that do not have a direct reference in the meaning component (e.g., there is no representation of “because” in any of the events).

Said in a different way, when learning from a narration of an experienced event, events in the sentences are matched with referenced events in the situation model. Those elements that don't match must then be narrative function words (NFWs). These will be used to create links, labeled with the NFW, between events mentioned in the same sentence. For “I gave you the toy because you wanted it,” *gave* and *wanted* in the sentence match with the *gave* and *want* events coded in the SM. *Because* cannot be found in the SM, and so must be considered an NFW. These events and the narrative relation make up the meaning that is paired with the sentence to constitute (with other pairs) the training corpus used to train the reservoir construction models. When this sentence is then presented to the trained model, the model generates the meaning as the events, and the narrative relation, *because*, which is used to create a labeled link between the want and gave events, as illustrated in the SM in **Figure 3**.

Learning to Produce Narrative Using Narrative Function Words From Narrative Practice

Given this infrastructure we can see how the SM can be generated from narrative. In order to generate narrative from the SM, we should just go in the opposite direction: the contents of the

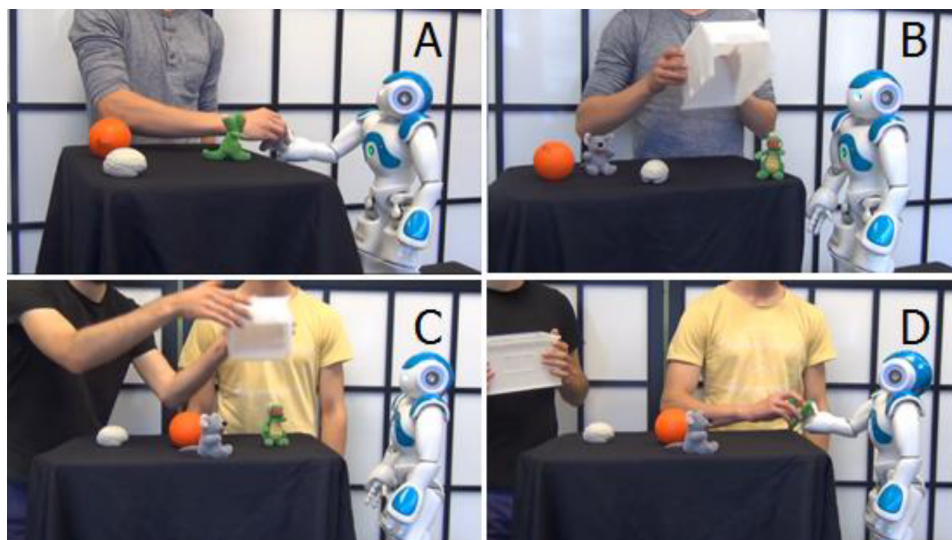


FIGURE 4 | Still images extracted from three of the six videos that were presented to subjects. **(A)** Partner hands an object to the robot. **(B)** Partner removes a box that blocked the robots access to the toy. **(C)** Partner 1 removes the box that obstructs access to the toy, and **(D)** partner 2 then hands the toy to the robot.

Robot wants the croco	Robot fails to take the croco	Robot asks for the croco to Human	Robot has the croco
Robot tries to take the croco	Robot reasons	Human gives the croco to Robot	

Discourse Function Word 2:

AFTER

Select second part

Robot wants the croco	Robot fails to take the croco	Robot asks for the croco to Human	Robot has the croco
Robot tries to take the croco	Robot reasons	Human gives the croco to Robot	

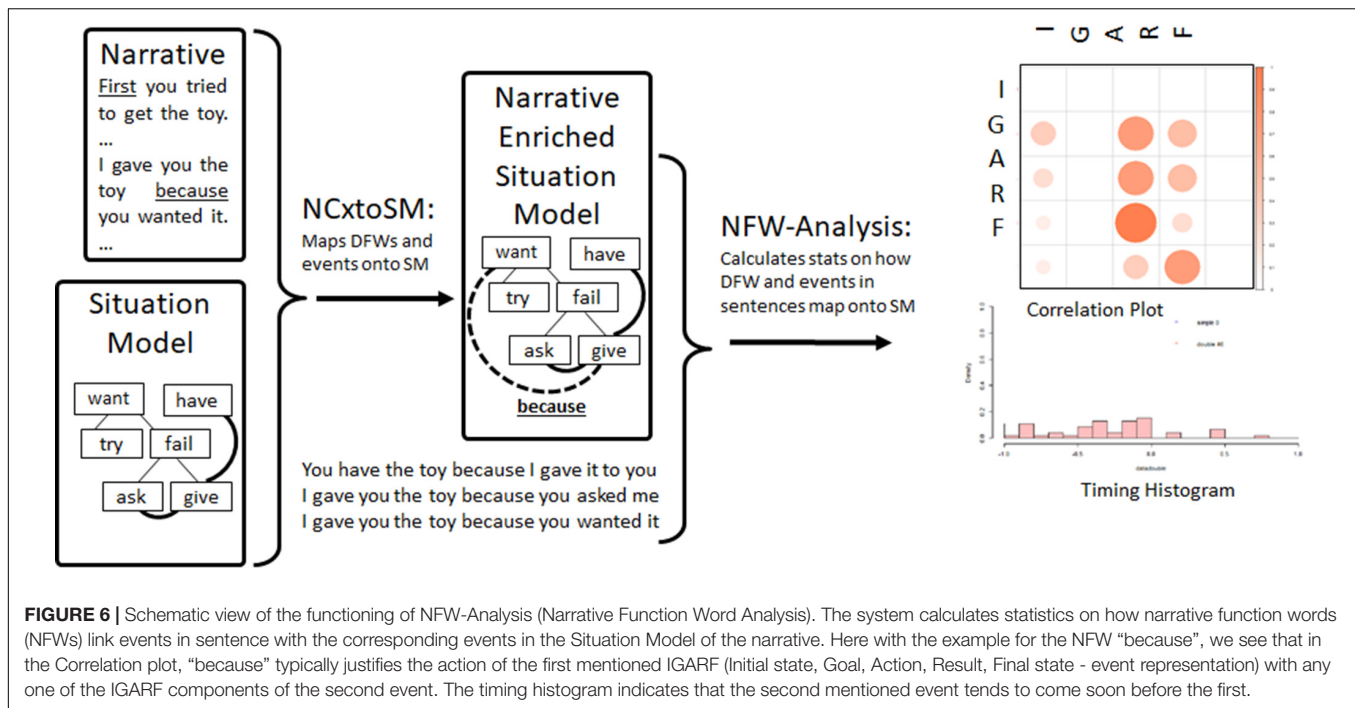
FIGURE 5 | Response panel from user interface for specifying use of NFWs.

situation model are used to generate meanings, with the two components – events and narrative relations, and this feeds into the narrative production model. The problem is that for a given situation model there a multiple different forms of sentences that can potentially be generated. Even more difficult, for new situations observed by the robot, the events will be encoded, but not the narrative relations, since they cannot be seen. Like the child, our system has to learn how events in the situation model are linked by causal and temporal relations, which can then be expressed in narrative. This is the problem we address here. This problem is of interest to researchers in developmental psychology, and developmental robotics.

Developmental studies of the acquisition of narrative function words indicate that there is a progression of complexity that typically starts with the use of “and” as an additive marker, then followed by markers for temporal, casual epistemic,

object specification, adversative, notice and other complement relations (Bloom et al., 1980). This emergence of discourse connectives is influenced by multiple factors including the conceptual complexity of the relations to be expressed, syntactic complexity of the forms used to express the relations, and the frequency of use in parental input (Evers-Vermeul and Sanders, 2009). Indeed, we should recall the importance of the parental/caregiver influence in the social context of interaction (Dominey and Dodane, 2004).

In this interactive context the child will learn how to express temporal and causal relations in the domain of human motivations for behavior. This problem has been approached by Hutto (2007) and Gallagher and Hutto (2008) in the context of the Narrative Practice Hypothesis. They argue that children engage in story-telling practices with others, and that through this narrative practice they are exposed to –and learn from– examples



of how narrative patterns are used to express reasons why people behave as they do. Like the child, our system will learn how to appropriately use narrative function words, based on experience. This experience takes the form of data characterizing how people talk about actions, and what kind of narrative function words they use to establish causal and temporal links between successive actions in a coherent scenario. These data can then be used to teach the system. To respond to this need, we gathered data from naïve human subjects who observe a human-robot interaction, and are then prompted to describe what they have seen. Their use of language then provides data for the learning system. This is described in section II.

Once we have data on how people use narrative function words to link events in narrative, we must render this data usable for the system. For this, we benefit from previous experience with a learning system that collects statistics on how pronouns are used, and generalizes so that the system learns to correctly use pronouns (Poiteau et al., 2014a). Here, we extended this system so that it accumulates statistics on how narrative function words like “because, first, so, then” specifically link different elements in a situation model, in order to talk about action in a meaningful way. This is described in section III.

Once the system has been trained on data from naïve human subjects, the system can then use this knowledge to discuss what happened with the human in a pertinent manner. These results are presented in Sections IV and V.

Ethics Statement

Written informed consent was obtained from the [individual(s) AND/OR minor(s)] legal guardian/next of kin] for the publication of any potentially identifiable images or data included in this article.”

TABLE 1 | Pseudo-code explaining how narrative function words are learned.

```
NarrativeFunctionWordLearning(Situation_Model, Narrative)
For each SENTENCE in Narrative
{
  MEANING = NarrativeReservoirComprehension(SENTENCE)
  // extract meaning with narrative reservoir model
  Locate MEANING in Situation_Model
  // MEANING may have 1 or 2 events
  Establish NFW link with the EVENT(s) in the Situation_Model
  //e.g. “because” links Event 1 (Result of IGARF N) with
  Event 2 (Action of IGARF M)
  Update Correlation_Plot statistics linking IGARF
  //e.g. “because” links Event 1 (Result of IGARF N) with
  Event 2 (Action of IGARF M)
  Update Timing Histogram statistics
  // e.g. in this example Event 1 is before Event 2
}
```

COLLECTING DATA ON HOW PEOPLE USE NARRATIVE FUNCTION WORDS

Certain dimensions of language structure can be learned through the extraction of statistical structure during exposure to language stimuli (Pelucchi et al., 2009). However, there are dimensions of language learning which require more direct social situation as in learning how to related other’s behavior to unseen goals and motivations. This is where Hutto’s narrative practice is pertinent, as it provides a framework to explain how children learn narrative patterns that explain behavior. Here we set out to initiate a simple modeling of these phenomena of socially situated learning.

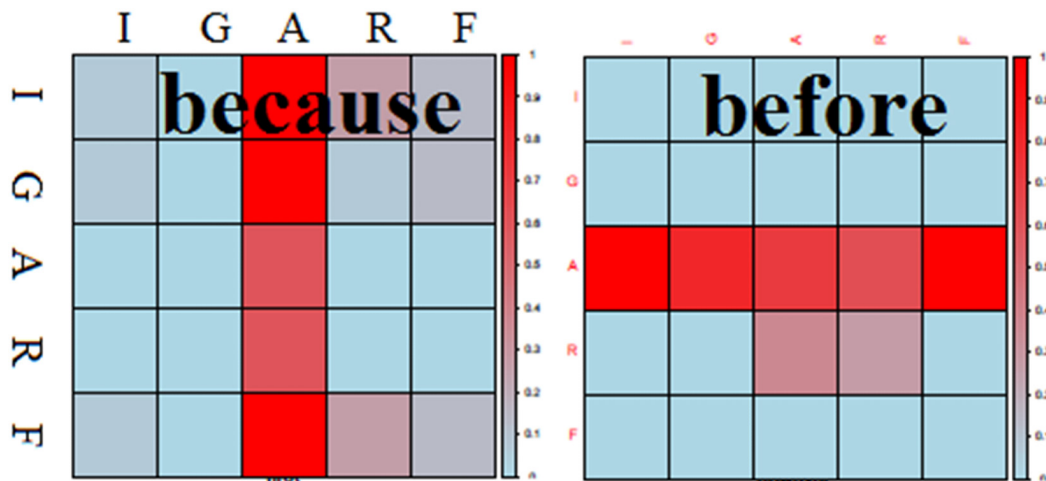


FIGURE 7 | Illustration of correlation plots for IGARF elements of events referred to by the NFW for the first event (horizontal axis) and second event (vertical axis) in data from nine naïve subjects. Correlation of the NFW linking the two event elements coded blue (0) to red (100%). Interesting behavior is observed for *because* and *before*. With *because* we see that the action of the IGARF for the first event can be causally linked to all components of a second event, as in the sentence “iCub take the croco because iCub want the croco.” In contrast, the use of “before” can link any element of the first event with the action of the second event.

TABLE 2 | Pseudo-code explaining how narrative function words are used in responding.

```

Use Narrative Function Word (pseudo-code)
UseNarrativeFunctionWord (Situation_Model, NFW, input_event(optional))
If input_event <> null
    Find input_event in Situation_Model
For each EVENT in Situation_Model
    // calculate statistics on most probably use of events
    // in SM with this NFW
    {
        Correlation = score(EVENT(i), input_event, Correlation_Plot(NFW))
        Timing = score(EVENT(i), input_event, Timing_Histogram(NFW))
        Update score_vector(i)(EVENT, Correlation_score, Timing_score)
    }
Response = select_best_event(score_vector)
Sentence = NarrativeGenerationModel(Response)
Say(Sentence)

```

In order to determine how people use narrative function words, we invited naïve subjects to watch a series of interactions that involved different levels of “complication” and “resolution,” involving a robot attempting to reach for a toy crocodile and a human helping the robot to achieve the goal. Still images from some of the videos are presented in **Figure 4**. These scenarios were designed to allow naïve subjects to be able to use narrative function words in order to describe when and why the partners and the robot performed as they did. By naïve, we mean that the subjects did not know anything about the algorithms being used to process their responses.

A set of six interactions that involved the human-robot interaction were filmed and put on YouTube, and via Qualtrics we allowed people to access these videos and then describe what they saw. Qualtrics is a tool that allows the creation of experimental protocols that can then be used in web applications such as Amazon Mechanical Turk.

- In Scenario 1, the robot tries to grasp the toy croco, but it is out of reach, so he asks for help, but gets the mouse instead of the croco, and so asks again and this time gets it. - In Scenario 2, the robot tries to grasp the croco that is covered by the box, asks Larry to remove the box, and Robert to give him the croco. - In Scenario 3, the robot tries to grasp the croco, but it is covered by a box, so he asks the human to remove the box, and then grasps the croco. - In scenario 4, the robot successfully grasps the croco. - In Scenario 5, the robot tries to grasp the croco that is out of reach, and then asks for the croco. - In Scenario 6, the robot tries to grasp the croco but it is covered by the box, so he asks Larry to remove the box, then he tries and fails to grasp the croco, and then asks Robert to give him the croco.

In a first data collection we asked people to describe what happened in the videos, using narrative functions words. In these unrestricted cases, the language produced was like this:

- (1) The robot seems to fail to pick up the crocodile because he is too far away
- (2) So he asks the person because that is easier than moving
- (3) The robots takes the time to thank the person, because that is what one should do
- (4) The robot finally got the toy
- (5) The robot asked the human to give the toy, and finally he could play with it
- (6) Although the robot first couldn't grasp the toy, it finally got it after asking for help
- (7) The robot needs help picking the toy because it cannot reach it
- (8) It cannot reach it because the toy is far away the user helps the robot because it cannot do it by itself

The NFW system requires that the meaning expressed in the sentences can be associated with meaning in the Situation Model. In some of these example sentences the mapping can be made

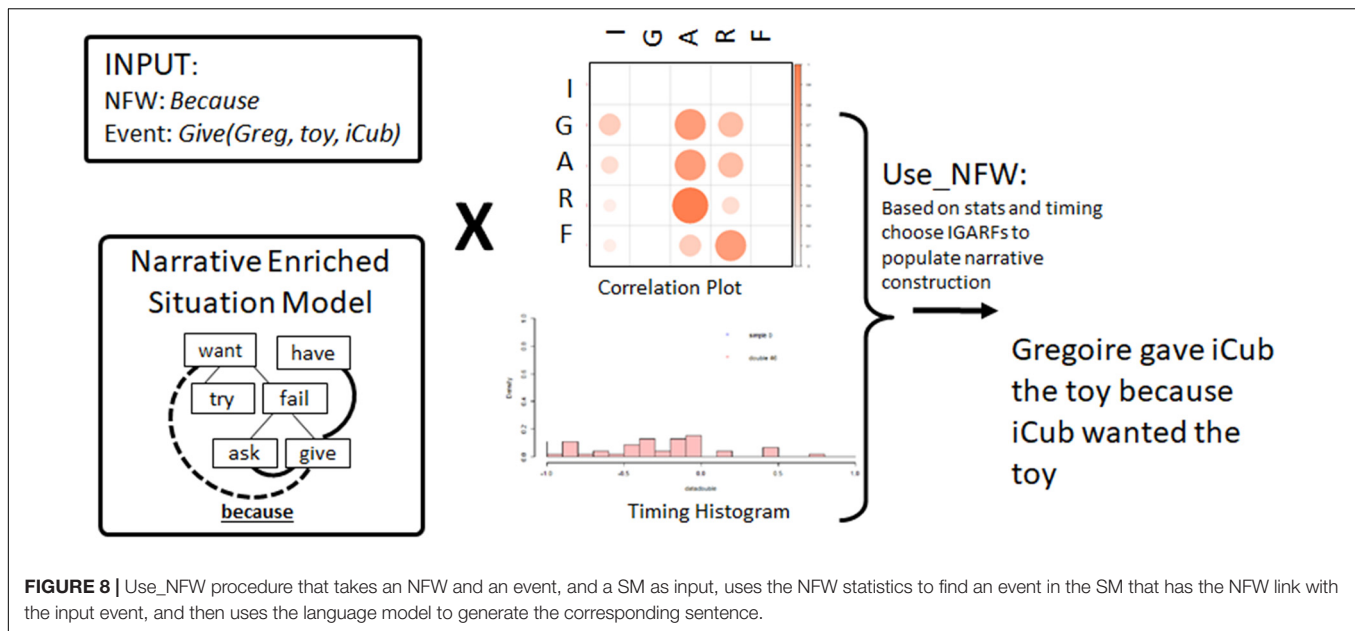


TABLE 3 | In order to generate event and NFW inputs for narrative to be generated by the system, we developed a simple mapping between questions that the user can pose, and the corresponding query that will be made to the situation model.

Question	Characteristics of returned events
What happened <first, then, finally>	Return events that respect the statistical (correlation and temporal) characteristics of the identified NFW
What happened <because, after, before> EVENT 1	Return events that respect the statistical (correlation and temporal) characteristics of the identified NFW and the cited event
Why did EVENT	Return events that respect the statistical (correlation and temporal) characteristics of the NFW “because” and the cited event
What else	CONTINUE with the same search
Why is that	Return events with a “because” link to the previously returned event
Do you remember when <first, then, finally> EVENT or EVENT <first, then, finally>	Search for a SM that contains the specified EVENT

(e.g., 1, 4, 6), while in others, the sentence refers to meaning components that are not in the SM (e.g., 2, 3). The sentences could be pre-processed, but what we were most interested in was how people used the NFWs to coordinate the main events in the scenarios.

In a second data collection using Qualtrics and Amazon Mechanical Turk, we tested nine subjects in a more structured way where they were given a narrative function word and could select a first event, and a second event, in order to make a sentence that described one of the scenarios. These sentences constructed in this more constrained situation allowed a more direct mapping onto the events in the SM for the discovery of how different NFWs are used to link these events. Here, we used a set of 12 NFW: “and, after, because, before, but, first, finally, however, so, then, therefore, while.” The data collection experiment that was performed by our subjects can be seen on this link:

https://survey.eu.qualtrics.com/jfe5/form/SV_6SF6NuZZdmTCr7D.

Here is an example of naïve subjects use of the narrative function words after, because and before.

The robot ask for the croco to the human after the robot fails to take the croco.

Human gives the croco to the robot because the robot asks the human for the robot.

The robot ask for the croco to the human before the human give the croco to the robot.

A screenshot of the interface for the choice of how to use a NFW is illustrated in **Figure 5**. This data collection campaign generated 432 distinct uses of the NFWs that could then be used for training the system.

LEARNING TO USE NARRATIVE FUNCTION WORDS

The situation model represents events and mental (goal) states, and different types of relations between them, expressed with narrative function words. In order to properly generate sentences that express these relations, we took a socially situated usage based approach (Tomasello, 2003) where the knowledge of how to use these narrative function words like “because” comes through narrative practice (Gallagher and Hutto, 2008; Hutto, 2009). We consider this as approximating learning contexts where people provide narrative about what happened, in the same way that caretakers would talk about events with a developing child, who learns by example.

Narrative function words express relations between events, and human knowledge about how these relations are expressed is encoded in the data we obtained from human subjects. Through a process of pattern matching statistical learning, the system

extracts regularities about how NFWs are used, and then re-employs these statistical patterns when generating narrative. In our algorithm, as illustrated in **Figure 6**, semantic words in sentences are matched with semantic words representing events in the situation model, in order to identify the referred events. A NFW in a sentence with two events will correspond to a link between these two events in the SM, and this link will be accounted for in the accumulating statistics, represented as the correlation plot in **Figure 6**. We learn based on two statistical analyses: Correlations between the Initial state, Goal, Action, Result and Final state (IGARF) elements referred to by the NFW, and relative Timing of the elements referred to by the NFW (e.g., does one of the events occur before or after the other).

Table 1 presents the pseudo-code algorithm for extracting this statistical structure. Sentences are analyzed to determine how the narrative function word, and its relation between the event(s) in the sentence, are to be represented in the situation model. For example, we showed that the system can observe that in sentences of the type “event-a because event-b,” there is a relation between the event component (initial state, goal, action, result, final state – IGARF) of the first and second event, and also a relation between the relative timing of event-a and event-b. Statistics on these relations can be accumulated, extending our work in this area for learning how to use pronouns (Pointeau et al., 2014b).

Figure 7 illustrates the results of this correlation analysis for two NFWs, *because* and *before*. We see the typical pattern of usage of these two words, reflected in these statistics. In our data, *because* is most often used to explain why an action occurred, corresponding to the vertical bar along the Action element of event 1. The reason can be related to an initial state, another action, result or final state (e.g., “Anne-Laure removes the box *because* the box cover the mug”). This corresponds to the distribution of probability along the different IGARF elements in the vertical band. For *before*, a different profile is observed. *Before* is used to explain what happened prior to a particular action (e.g., “the box cover the mug *before* Anne-Laure removes the box.”). What happened before can be any of the IGARF elements, and what happened after is an action. This corresponds to the horizontal band aligned along the Action dimension for *before*.

EXPLOITING NARRATIVE FUNCTION WORDS

The point of learning how NFWs are used is to then be able to use them in this conventional way – to talk about action the way that one has learned that others talk about action. So in answer to a question like “what happened first?” one can respond in the same way that one has heard others talking and using the word “first”. Likewise when asked why did an event occur, one would respond by using the word “because” in the same way that one has seen others do it.

Use_NFW is a procedure (described in **Table 2** and **Figure 8**) for extracting event representations in the form of IGARFs from the SM based on narrative links encoded in the statistics in the correlation plot and timing histogram learned from experience. The Correlation Plot encodes the source and target events typically referred to by a narrative function word (e.g., “because” often refers to the Action of the source and the multiple possible IGARF elements of the target). The Timing Histogram describes the temporal precedence for source and target.

The function Use_NFW exploits the statistics learned in the correlation plot and the timing histograms in order to determine what elements of the situation model should be extracted and used to generate a meaning representation for the narrative construction production reservoir model. Use_NFW takes as input [NFW, optional event [PAOR], optional order]. The optional event is because some NFWs do not take an event (like *first*, or *finally*), whereas others do (like *before* event, or *after* event). The optional order corresponds to the place of the given [PAOR] in the desired sentence: “What happened after I gave you the toy?” (Result: “After you gave me the toy, I have the toy.”) versus “I gave you the toy after what?” (Result: “You gave me the toy after I asked for it”). The function Use_NFW will thus return a set of events with a corresponding score (based on the distribution of the histogram and correlation plot), that will be sent to Narrative Construction Model to be generated as a sentence.

Narrative function words can also be used to search for a particular situation model. That is, a relation expressed by an

TABLE 4 | Example dialogue excerpt from an interaction with the iCub.

H: What happened first?
R: First I want the brain
H: What else?
R: First Greg have the brain
H: What else
R: First I take the brain
H: What happened then?
R: Then I said: “Give me the brain please” to Greg
H: What else?
R: Then Greg give the brain to me
H: Why did Greg give the brain to you?
R: Greg give the brain to me because I say: “Give me the brain please” to Greg
H: What happened after Greg gave the brain to you?
R: After Greg give the brain to me, I have the brain

TABLE 5 | This dialogue example illustrates how the human can start at the end of the interaction and ask questions proceeding from end to beginning.

H: What happened finally?
R: Finally I have the brain
H: Why do you have the brain?
R: I have the brain because Greg give the brain to me
H: Why did Greg give the brain to you?
R: Greg give the brain to me because I say ‘give me the brain please’ to Greg
H: Why did this happen?
R: I said “give me the brain please” to Greg because I failed the take
H: What else?
R: I said “give me the brain please” to Greg because I reasoned
H: Why did your reason?
R: I reasoned because I failed the take
H: Why did you take the brain?
R: I take the brain because I want the brain

TABLE 6 | Narrative generated from a novel situation, by applying a stored set of NFWs (marked in bold) from a set of questions posed by the human to yield a narrative template.

First Anne-Laure have the mouse.
Then I said: Give me the croco please, to Anne-Laure.
Then Anne-Laure give the mouse to me.
Finally I have the croco.
I have the croco **because** I have the mouse.*
I have the croco **because** I said: Give me the croco, to partner

NFW and one or more IGARFs can be used as a pattern that will be searched for in a set of SMs. This allows a form of interrogation of the system as: “Do you remember when . . .?”

HUMAN-ROBOT INTERACTION

Once the system has been trained as described above, it is almost ready to use the acquired knowledge in order to communicate about actions. The final element is another form of socially situated knowledge about conventions for how to answer questions. We believe that these conventions (specified in Table 3) can also be learned by narrative practice, but in the current demonstration they are pre-specified as described below.

Here we illustrate the ability of the system to use narrative function words in order to respond to questions in an interactive dialogue with the human. In the examples we present, an interaction first takes place, where the iCub wants to grasp a toy brain (see Figure 9). It attempts the grasp and fails. It then uses reasoning to determine if there is another method to get the brain, and determines that it can ask Greg. It does so, and Greg gives the iCub the toy brain.

The remarkable point here is that there has been no specific training for this scenario. That is, Greg and the iCub experienced

a shared cooperative activity where Greg helped the robot, but there was no specific language training about this interaction. Rather, based on several previous interactions, and narrative provided by people, the system was able to learn (as just described) how people use narrative function words like *first*, *then*, *because*, and *finally* – with respect to the ordering and relation between events in a situation model – when talking about events. So, the system learns that when using the word *first* one talks about things that occurred near the beginning of the scenario. When using the word *because* and an event, one talks about events that preceded the event in question.

In order to allow dialogic interaction a simple turn taking setup and a set of questions were developed. The questions and the selection of events that are used to generate the responses with the narrative reservoirs are depicted in Table 3. For an example of the functioning of the system identified in Table 3, consider the sentence “Why did Greg give the brain to you?” This is interpreted as “Why did EVENT” in Table 3. The interpreter thus calls the function Use_NFW to identify events that have a “because” link with the event Give (Greg, brain, iCub) in order to generate a sentence “Greg gave the brain to me because EVENT.” The system will look for events that precede the target event, and that correspond to an Initial state, Action, or Result.

The two dialogues illustrated in Tables 4, 5 were generated in on-line question answering with the system (see video)¹. We can observe that by the open-ended possibility of questions, the human is allowed to explore the past experience from different approaches. In the example illustrated in Table 4, the human starts at the beginning, by asking “What happened first?” and then moves forward from there. In the second example, Table 5, the human starts at the result by asking “What happened finally?” and works back through the causal chain.

¹<https://youtu.be/Lhs2aQ7zLK4>

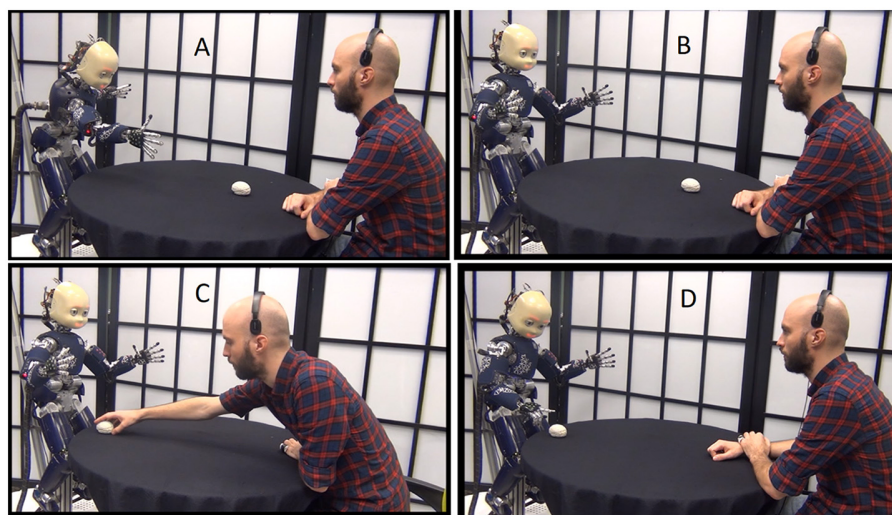


FIGURE 9 | New scenario - trying to take the brain. (A) iCub wants the brain and tries to take it. (B) After failing, and reasoning, iCub asks Gregoire to give him the brain. (C) Gregoire gives the brain to iCub. (D) iCub acknowledges that it has the brain.

In both of these examples, the system did not receive specific training in the scenario that was described in the dialogue. Rather, it was trained on data generated by naïve subjects in different scenarios. This training allowed the extraction of statistical rules about how NFWs are used, which could then be applied to this new situation.

GENERALIZATION OF A CANONICAL STORY TEMPLATE TO NEW SITUATIONS

This capacity to extract the rules of use of narrative function words, based on experience encoded in narrative that the robot has heard, allows the system an unprecedented level of generalization that we did not see with the grammatical constructions. In the narrative constructions, narrative function words like “because” specify relations between event components within a situation model. The system learns how *because*, and other NFWs, are used. An example is illustrated in **Table 6**, where a new situation is encountered and used to construct a situation model, then the system applies a learned sequence of NFWs to this situation model to generate a new simple narrative.

In contrast, in the grammatical constructions, grammatical function words like “by” specify relations between open class elements in the sentence, but rather than learning these relations as statistical rules associated with each grammatical function word (GFW), we learn an entire mapping from the whole sentence, with the global pattern of GFWs, onto the predicate-argument representation of the meaning. Thus, it is impossible for the system to learn the functions of individual GFWs. Interestingly, however, when the DCG model is exposed to sufficiently large corpora, it is able to generalize to new grammatical constructions that it was not trained on Hinaut and Dominey (2013). Thus, functionally, it learns how to interpret grammatical function words in a general manner. Still, the generalization on NFWs in the narrative model is much more powerful, and based on the compositionality of representations in the SM (Dominey, 2003).

However, the generalization is not entirely immune to error. We see with the sentence marked with “*” that the iCub says “I have the croco because I have the mouse.” This is not entirely wrong – prior to getting the croco the iCub did have the mouse. Part of the definition of the *because* relation is that the causal event does precede the caused event, and so from the temporal perspective this error can be understood. Interestingly, this kind of anomalous use of “because” is observed in situations where children (3 year, 6 month–9 year) narrate their personal experience. In the kinds of errors that these children can make, one event follows another, but the first event does not cause or enable the second, or vice versa: e.g., “I fell and just hurted my neck. Because I had to go to the doctor’s to get the shot for my mumps” (McCabe and Peterson, 1985). While it is highly probable that these observations in children reflect cognitive processes we do not model here, still, for children and our system, there is an observation of using *because* to preserve temporal order where causation is not directly present.

DISCUSSION

In its ecological form language is highly socially situated and is indeed a vehicle for social situation. Human interactions and events form the social matrix that is observed and must be explained and justified to others. Yet at the same time, the manner in which this communication is to be achieved is itself a social norm that is socially situated in its acquisition. Interestingly, such social conventions apply at the lexical, grammatical and narrative levels, in the service of meaning.

Actions typically do not take place in isolation, and when we talk about actions and events, we don’t simply state dully the action that took place, but instead we talk about the event in an interesting, pertinent way in the ongoing dialogical or narrative context in a way that has *meaning* (Bruner, 1990). We stress the notion of meaning, because meaning is derived not just from the action itself, but from the situated social and narrative context in which it is embedded: why did you do it, who did it, and when! This meaning is characterized by how an action is integrated into an intentional network of interrelated actions. These relations are described by a category of words that we refer to as narrative function words.

Crucially, the manner in which these words are used to express the organization of events and the relations that interconnect them are not arbitrary, nor are they innate. Rather, they are cultural artifacts that are transferred to the young new member of the culture (Tomasello, 1999), through narrative practice (Hutto, 2007; Gallagher and Hutto, 2008; Nelson, 2009; Hutto and Kirchhoff, 2015).

We emphasized how narrative practice could allow children to construct relations between events. A complementary account is that these relations (e.g., causal relations) are already perceived by the infant, and narrative practice allows the infant to learn how to appropriately label and refer to such relations. Our model is consistent with both accounts, and in human development it is likely that there is an interaction between them. Indeed, Bruner suggests that children may initially be limited in this causal paradigmatic thinking, and that adult discourse has a role in guiding children toward the right causal analyses, so that ultimately the child can perform these causal analyses autonomously (Bruner, 2009). Similarly, Lagerwerf (1998) considers that the use of the causal connective “because” presupposes an understanding of the relation between the causing and caused. Further supporting this position (Knott, 1996), considers that coherence relations like causality describe cognitive constructs that we use to represent the world, independent from linguistic processing. Thus, narrative would serve to label and make explicit cognitive constructs like causality that have already been perceived by the child. In Section II we sought to characterize how people use NFWs in the context of the scenarios that we study. This involved data collection in which naïve subjects narrated human-robot interactions, using narrative function words. We found that in fully unconstrained conditions, the sentences generated by subjects did not sufficiently map onto the events that were represented in the situation models, and so we modified our data collection so that subjects were more constrained to generate

sentences that refer to events in the situations models and the relations between them, thus providing more concrete demonstrations of how to use the narrative function words.

We then presented algorithms for learning how NFWs are used to express narrative relations about actions in section III, and for using this knowledge to allow the system to then use this knowledge in section IV, with demonstration in section V. At this point we made a remarkable observation about generalization and narrative function words: by learning how specific NFWs are used, the system is able to generalize to new situation models. That is, the learned narrative ability can be applied to contexts different from those used for learning.

In section V we thus demonstrate this ability for the system to communicate about action. The iCub is able to answer questions about diverse scenarios (illustrated here with one), based on the NFW processing.

The system is able to talk about actions in a rather advanced way. Rather than talking about actions in an isolated manner, with each action being independent, instead the system is able to situate actions in time, and with respect to other actions. Likewise, the human can approach the scenario in question from different perspectives – starting at the beginning and working forward, at the end and working backward, etc. This extends our work on perspective and construing an event in different ways (Mealier et al., 2016). This work on learning to use narrative functions words represents an advanced level of human-robot communication about actions. Future work will examine structural relations between situation models, and use of narrative function words in the context of specific (vs. statistical) use in these situation models.

One of the major limitations of the current research is illustrated in **Table 3**, which provides the form of responses to be used for different types of questions. The limitation is that this type of correspondence is exactly the kind of knowledge that can be acquired through narrative practice. That is, the child or learning system can observe what are the types of responses that are the social conventions for different types of questions. In the current research, this is a limitation in terms of what was actually done, but not in terms of what is theoretically possible.

This raises the point of the final remark in terms of limitations. This research demonstrates a form of feasibility or proof of concept for the ability to learn to use narrative function words

in order to organize and communicate experience. As stated at the outset, the behavior of the system is demonstrated as a set of illustrative behavioral interactions, and this proof of concept lays the foundation for a more extensive user study that will be the subject of future research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

All authors conceived the system, performed the experiments, and wrote the manuscript. GP, A-LM, and SM developed the software infrastructure.

FUNDING

This research was supported by the European Union, FP7 ICT Contract 612139 What You Say Is What You Did, WYSIWYD, and the French Région Bourgogne Franche Comté Project ANER RobotSelf.

ACKNOWLEDGMENTS

We thank Théophile Peyraud for help in collecting data on use of narrative function words.

REFERENCES

- Bloom, L., Lahey, M., Hood, L., Lifter, K., and Fiess, K. (1980). Complex sentences: acquisition of syntactic connectives and the semantic relations they encode. *J. Child Lang.* 7, 235–261. doi: 10.1017/s0305000900002610
- Bruner, J. (1991). The narrative construction of reality. *Crit. Inquiry* 18, 1–21.
- Bruner, J. S. (1990). *Acts of Meaning*. Cambridge: Harvard University Press.
- Bruner, J. S. (2009). *Actual Minds, Possible Worlds*. Cambridge: Harvard University Press.
- Crangle, C., and Suppes, P. (1994). *Language and Learning for Robots*, Vol. 41. Stanford, CA: CSLI.
- Dominey, P., and Boucher, J. (2005). Learning to talk about events from narrated video in a construction grammar framework. *Artif. Intel.* 167, 31–61. doi: 10.1016/j.artint.2005.06.007
- Dominey, P., Mallet, A., and Yoshida, E. (2007a). Progress in programming the hrp-2 humanoid using spoken language. *Paper Presented at the IEEE International Conference on Robotics and Automation*, Piscataway, NJ.
- Dominey, P., Mallet, A., and Yoshida, E. (2007b). Real-time cooperative behavior acquisition by a humanoid apprentice. *Paper Presented at the International Conference on Humanoid Robotics*, Pittsburg, CA.
- Dominey, P., Mallet, A., and Yoshida, E. (2009). Real-Time spoken-language programming for cooperative interaction with a humanoid apprentice. *Intl. J. Humanoids Rob.* 6, 147–171. doi: 10.1142/s0219843609001711
- Dominey, P. F. (2003). A conceptocentric shift in the characterization of language: comment on Jackendoff. *Behav. Brain Sci.* 26, 674–674. doi: 10.1017/s0140525x03290154
- Dominey, P. F., and Dodane, C. (2004). Indeterminacy in language acquisition: the role of child directed speech and joint attention. *J. Neurolinguistics* 17, 121–145. doi: 10.1016/s0911-6044(03)00056-3
- Dominey, P. F., Mealier, A.-L., Pointeau, G., Miriaz, S., and Finlayson, M. (2017). Dynamic construction grammar and steps towards the narrative construction of meaning. *Paper Presented at the AAAI Spring Symposium, Computational Construction Grammar and Natural Language Understanding*, Palo Alto, CA.

- Evers-Vermeul, J., and Sanders, T. (2009). The emergence of Dutch connectives; how cumulative cognitive complexity explains the order of acquisition. *J. Child Lang.* 36, 829–854. doi: 10.1017/s0305000908009227
- Fraser, B. (1999). What are discourse markers? *J. Pragmatics* 31, 931–952. doi: 10.1016/s0378-2166(98)00101-5
- Gallagher, S., and Hutto, D. (2008). “Understanding others through primary interaction and narrative practice,” in *The Shared Mind: Perspectives on Intersubjectivity*, eds E. Itkonen, J. Zlatev, C. Sinha, and T. P. Racine (Amsterdam: John Benjamin Publishing company), 17–38. doi: 10.1075/celcr.12.04gal
- Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends Cogn. Sci.* 7, 219–224. doi: 10.1016/s1364-6613(03)00080-9
- Grosz, B. J., and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Comput. Linguist.* 12, 175–204.
- Grosz, B. J., Weinstein, S., and Joshi, A. K. (1995). Centering: a framework for modeling the local coherence of discourse. *Comput. Linguist.* 21, 203–225.
- Hinaut, X., and Dominey, P. F. (2013). Real-time parallel processing of grammatical structure in the fronto-striatal system: a recurrent network simulation study using reservoir computing. *PLoS One* 8:e52946. doi: 10.1371/journal.pone.0052946
- Hinaut, X., Lance, F., Droin, C., Petit, M., Pointeau, G., and Dominey, P. F. (2015). Corticostriatal response selection in sentence production: insights from neural network simulation with reservoir computing. *Brain Lang.* 150, 54–68. doi: 10.1016/j.bandl.2015.08.002
- Hinaut, X., Petit, M., Pointeau, G., and Dominey, P. F. (2014). Exploring the acquisition and production of grammatical constructions through human-robot interaction with echo state networks. *Front. Neurobot.* 8:16. doi: 10.3389/fnbot.2014.00016
- Hobbs, J. R. (1985). *On the Coherence and Structure of Discourse*. Stanford, CA: CSLI.
- Hutto, D. (2009). Folk psychology as narrative practice. *J. Conscious. Stud.* 16, 9–39.
- Hutto, D. D. (2007). The narrative practice hypothesis: origins and applications of folk psychology. *R. Institute Philosophy Suppl.* 60, 43–68. doi: 10.1017/cbo9780511627903.004
- Hutto, D. D., and Kirchhoff, M. D. (2015). Looking beyond the brain: social neuroscience meets narrative practice. *Cogn. Syst. Res.* 34, 5–17. doi: 10.1016/j.cogsys.2015.07.001
- Knott, A. (1996). *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. Ph.D. Dissertation, University of Edinburgh, Edinburgh. Available online at: <http://hdl.handle.net/1842/583> (accessed 2016).
- Knott, A., and Sanders, T. (1998). The classification of coherence relations and their linguistic markers: an exploration of two languages. *J. Pragmatics* 30, 135–175. doi: 10.1016/s0378-2166(98)00023-x
- Kollar, T., Tellex, S., Roy, D., and Roy, N. (2010). Toward Understanding Natural Language Directions. *Paper presented at the Proceeding of the 5th ACM/IEEE International Conference on Human-robot Interaction – HRI '10*, Osaka.
- Lagerwerf, L. (1998). *Causal Connectives Have Presuppositions*. Ph.D. thesis, Catholic University of Brabant, Hague.
- Lascarides, A., and Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics Philosophy* 16, 437–493. doi: 10.1007/bf00986208
- Lauria, S., Bugmann, G., Kyriacou, T., and Klein, E. (2002). Mobile robot programming using natural language. *Rob. Auton. Syst.* 38, 171–181. doi: 10.1016/s0921-8890(02)00166-5
- Mann, W. C., and Thompson, S. A. (1988). Rhetorical structure theory: toward a functional theory of text organization. *Text Interdisciplinary J. Study Discourse* 8, 243–281.
- Matuszek, C., Herbst, E., Zettlemoyer, L., and Fox, D. (2013). “Learning to parse natural language commands to a robot control system,” in *Experimental Robotics. Springer Tracts in Advanced Robotics*, Vol. 88, eds J. Desai, G. Dudek, O. Khatib, and V. Kumar (Heidelberg: Springer).
- McCabe, A., and Peterson, C. (1985). A naturalistic study of the production of causal connectives by children. *J. Child Lang.* 12, 145–159. doi: 10.1017/s0305000900006280
- Mealier, A.-L., Pointeau, G., Gärdenfors, P., and Dominey, P. F. (2016). Construals of meaning: the role of attention in robotic language production. *Interact. Stud.* 17, 41–69.
- Mealier, A.-L., Pointeau, G., Miriaz, S., Ogawa, K., Finlayson, M., and Dominey, P. F. (2017). Narrative constructions for the organization of self experience: proof of concept via embodied robotics. *Front. Psychol. Lang.* 8:1331. doi: 10.3389/fpsyg.2017.01331
- Metta, G., Sandini, G., Vernon, D., Natale, L., and Nori, F. (2008). The iCub humanoid robot: an open platform for research in embodied cognition. *Paper presented at the PerMIS: Performance Metrics for Intelligent Systems Workshop*, Washington DC.
- Moulin-Frier, C., Fischer, T., Petit, M., Pointeau, G., Puigbo, J.-Y., Pattacini, U., et al. (2017). DAC-h3: a proactive robot cognitive architecture to acquire and express knowledge about the world and the self. *IEEE Trans. Cogn. Dev. Syst.* 10, 1005–1022. doi: 10.1109/tcds.2017.2754143
- Nelson, K. (2003). “Narrative and the emergence of a consciousness of self,” in *Narrative and Consciousness*, eds G. D. Fireman, T. E. McVay, and O. J. Flanagan (Oxford: Oxford University Press), 17–36.
- Nelson, K. (2009). Narrative practices and folk psychology: a perspective from developmental psychology. *J. Consciousness Studies* 16, 69–93.
- Norrick, N. R. (2001). Discourse markers in oral narrative. *J. Pragmatics* 33, 849–878. doi: 10.1016/s0378-2166(01)80032-1
- Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009). Statistical learning in a natural language by 8-Month-Old Infants. *Child Dev.* 80, 674–685. doi: 10.1111/j.1467-8624.2009.01290.x
- Petit, M., Lalle, S., Boucher, J.-D., Pointeau, G., Cheminade, P., Ognibene, D., et al. (2013). The coordinating role of language in Real-Time Multi-Modal learning of cooperative tasks. *IEEE Trans. Auton. Ment. Dev.* 5, 3–17. doi: 10.1109/tamd.2012.2209880
- Pointeau, G., Petit, M., and Dominey, P. F. (2014). Successive developmental levels of autobiographical memory for learning through social interaction. *IEEE Trans. Auton. Ment. Dev.* 6, 200–212. doi: 10.1109/tamd.2014.2307342
- Pointeau, G., Petit, M., Gibert, G., and Dominey, P. F. (2014a). “Emergence of the use of pronouns and names in triadic human-robot spoken interaction,” in *Proceeding of the IEEE Int Conf Dev and Learning & Epigenetic Robotics (ICDL-Epirob)*, New York, NY: IEEE.
- Pointeau, G., Petit, M., Gibert, G., and Dominey, P. F. (2014b). Emergence of the use of pronouns and names in triadic human-robot spoken interaction. *Paper Presented at the 4th International Conference on Development and Learning and on Epigenetic Robotics*, France.
- Sorce, M., Pointeau, G., Petit, M., Mealier, A.-L., Gibert, G., and Dominey, P. F. (2015). Proof of concept for a user-centered system for sharing cooperative plan knowledge over extended periods and crew changes in space-flight operations. *Paper Presented at the Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on*, New York, NY
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.
- Tomasello, M. (2003). *Constructing a Language: A Usage Based Approach to Language Acquisition*. Boston: MIT Press.
- Waxman, S. R., and Markow, D. B. (1995). Words as invitations to form categories: evidence from 12- to 13-month-old infants. *Cogn. Psychol.* 29, 257–302. doi: 10.1006/cogp.1995.1016
- Webber, B., Knott, A., and Joshi, A. (2001). “Multiple discourse connectives in a lexicalized grammar for discourse,” in *Computing Meaning. Studies in Linguistics and Philosophy*, eds H. Bunt, R. Muskens, and E. Thijss (Dordrecht: Springer), 229–245. doi: 10.1007/978-94-010-0572-2_13
- Zwaan, R. A., and Madden, C. J. (2004). Updating situation models. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 283–288; discussion 289–291. doi: 10.1037/0278-7393.30.1.283
- Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185. doi: 10.1037/0033-2909.123.2.162

Conflict of Interest: 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 © 2021 Pointeau, Miriaz, Mealier and Dominey. 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.



Effects of Dispositional Affect on the N400: Language Processing and Socially Situated Context

Veena D. Dwivedi^{1*} and Janahan Selvanayagam²

¹Department of Psychology and Centre for Neuroscience, Brock University, St. Catharines, ON, Canada, ²Department of Physiology and Pharmacology, University of Western Ontario, London, ON, Canada

OPEN ACCESS

Edited by:

Pia Knoeferle,
Humboldt University of Berlin,
Germany

Reviewed by:

Viviana Masia,
Sapienza University of Rome, Italy
Marijn Struiksma,
Utrecht University, Netherlands

*Correspondence:

Veena D. Dwivedi
vdwivedi@brocku.ca

Specialty section:

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

Received: 28 May 2020

Accepted: 10 March 2021

Published: 31 March 2021

Citation:

Dwivedi VD and
Selvanayagam J (2021) Effects of
Dispositional Affect on the N400:
Language Processing and Socially
Situated Context.
Front. Psychol. 12:566894.
doi: 10.3389/fpsyg.2021.566894

We examined whether the N400 Event-Related Potential (ERP) component would be modulated by dispositional affect during sentence processing. In this study, 33 participants read sentences manipulated by direct object type (congruent vs. incongruent) and object determiner type (definite vs. demonstrative). We were particularly interested in sentences of the form: (i) *The connoisseur tasted the wine on the tour* vs. (ii) *The connoisseur tasted the #roof...* We expected that processing incongruent direct objects (*#roof*) vs. congruent objects (*wine*) would elicit N400 effects. Previous ERP language experiments have shown that participants in (induced) positive and negative moods were differentially sensitive to semantic anomaly, resulting in different N400 effects. Presently, we ask whether individual dispositional affect scores (as measured by the Positive and Negative Affect Schedule; PANAS) would modulate N400 effects as shown previously. Namely, previous results showed larger N400 effects associated with happy moods and attenuated amplitudes associated with sad moods. Results revealed significant N400 effects, driven by *the #roof* vs. *the wine*, where larger amplitude differences were found for individuals showing smaller negative affect (NA) scores, thus partially replicating previous findings. We discuss our results in terms of theories of local (lexical) inhibition, such that low NA promotes stronger lexico-semantic links in sentences. Finally, our results support accounts of language processing that include social and biological characteristics of individuals during real-time sentence comprehension.

Keywords: N400, dispositional affect, conceptual semantics, sentence comprehension, mood/emotion, global-local

INTRODUCTION

It is well established that individuals differ in their responses to identical stimuli. Whereas one individual perceives, and interprets a glass filled at the 50% mark as half-full, another perceives, and interprets it as half-empty. A common-sense explanation for these differences is that the personality and/or mood of the perceiver plays a role in the above-mentioned interpretive possibilities.

The idea that the emotional mood of the perceiver may influence interpretation is supported by research in cognitive psychology (see, inter alia, Ashby et al., 1999; Gasper and Clore, 2002; Fredrickson and Branigan, 2005; Dale and Arnell, 2010). For example, it has been claimed

that positive moods are associated with greater global processing, such that individuals are more sensitive to top-down contextual knowledge. In contrast, negative moods tend to be associated with cognitive styles characterized by item specific processing, where attention is more focused on details (Kimchi and Palmer, 1982; Schwarz, 2002; Rowe et al., 2007). With respect to mood effects on language processing, Chwilla et al. (2011) conducted a series of studies using Event-Related Potential (ERP) methods (Vissers et al., 2010, 2013; see also Federmeier et al., 2001; van Berkum et al., 2013). They used a paradigm where participants were induced into happy or sad emotional moods by watching either happy or sad movies (e.g., short clips of either *Happy Feet* or *Sophie's Choice*). Their studies showed that participants in happy moods exhibited modulation of the language-related ERP effects of interest, whereas sad moods attenuated ERP responses. Relevant to our current study, Chwilla et al. (2011) focused on the N400 component, a negative-going waveform that peaks approximately 400 ms after a word, which does not match previous sentential context in terms of lexico-semantic fit (Kutas and Hillyard, 1980, 1984). In that work, they examined sentences of the form “*The pillows are stuffed with **feathers**...*” vs. “*The pillows are stuffed with **books**...*” and showed that the latter sentence with the low-cloze fragment showed larger N400 effects for happy vs. sad participants. They interpreted their findings as consistent with literature on mood and cognition (see above, as well as Schwarz and Clore, 2007), where happy individuals were more sensitive to global features of stimuli. That is, global features of sentence interpretation are associated with stereotypical event knowledge, also called script or schemata (cf. Schank and Abelson, 1977; St. George et al., 1994; Chwilla and Kolk, 2005). Finally, they concluded that their ERP results supported the recent shift in cognitive neuroscience that views cognition as “hot” (i.e., not independent of mood) vs. “cold.”

In the present work, we build on the findings above, as well as recent work in our own lab (Selvanayagam et al., 2019) and ask whether the related construct of dispositional affect would modulate sentence interpretation as measured via the N400 component.

In contrast to mood, which can fluctuate according to situation, dispositional affect reflects the stability across time and situation of individuals to view their world with approach-oriented positive affect (PA), or avoidance-oriented negative affect (NA). That is, it is a personality trait, reflective of individual differences (Watson and Clark, 1984; Staw and Ross, 1985; Levin and Stokes, 1989).

As such, we can consider dispositional affect of an individual as a socially relevant and arguably biological characteristic of an individual during language comprehension. The Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) indexes trait and state features of affect and mood. Given that it has been shown that there are individual differences in temperaments that can be more or less susceptible to mood induction (Larsen and Ketelaar, 1989, 1991; Brief et al., 1995), we ask whether dispositional affect, as measured by the PANAS, modulated N400 responses to sentences. In a preliminary investigation in our lab (Selvanayagam et al., 2019), we showed that the P300 effect

was indeed influenced by affect. Results were consistent with the view that less positive individuals, measured via the PANAS, were less sensitive to global heuristic cues of meaning.

In the current experiment, we focus on the N400 ERP component. We examined neural responses to words that were either congruent or incongruent with sentential context. That is, we contrasted sentences of the form *The connoisseur tasted the **wine** during the tour* vs. *The connoisseur tasted the **#roof**¹ during the tour*. The direct object in the latter sentence clearly violates our real-world expectations of what possible objects of tasting might be. As such, the critical word **roof** should elicit an N400 effect vs. its control **wine**. Our predictions regarding dispositional affect, *prima facie*, would be to replicate the findings of the induced mood participants in Chwilla et al. (2011). That is, N400 effects would be larger for individuals exhibiting higher PA scores, given findings of larger N400 effects for happy participants. Furthermore, N400 effects would be smaller for individuals with higher NA scores, mirroring previous findings of attenuated N400 effects for sad participants. These findings are expected on the cognitive view that positive participants should be more sensitive to lexico-semantic cues regarding global sentence meaning (also called heuristics, Townsend and Bever, 2001; Ferreira, 2003; Dwivedi, 2013) vs. negative participants, who are not. When a word does not match its lexico-semantic context, the greater sensitivity of positive (vs. negative) participants to this mismatch should therefore elicit larger N400 effects.

As such, correlational analyses will be conducted between N400 amplitudes and positive and negative affect scores. We predict (i) a significant positive correlation between N400 amplitude and positive affect scores as well as (ii) a significant negative correlation between the size of the N400 effect and negative affect scores of our participants.

Next, we note here that a separate condition was included in this experiment to ask a question independent of dispositional affect and is discussed in detail elsewhere (Dwivedi and Selvanayagam, 2020a,b). Briefly, we wanted to know whether neural responses to lexico-semantic incongruency differed from those elicited via discourse semantic violations. We did so by examining “double violations” (Hagoort, 2003). As such, the other independent variable we examined was determiner type at object position, such as *The connoisseur tasted ***that** # **roof**...* (without previous context, use of *that* results in presuppositional violation). In terms of our present study, the double violation condition might show the largest N400 effect, and if so, it too would correlate with positive and negative affect scores as above. See **Table 1** for list of conditions.

¹Note that the ‘#’ symbol (borrowed from semantic theory, cf., Roberts, 1989) indicates that the word does not match its context.

²The ‘*’ symbol is borrowed from syntactic theory, indicating ungrammaticality, as in **He likes Sarah*, where the pronoun *he* lacks an antecedent. We note that using a pronoun or referential noun phrase (e.g., *that wine*) without an explicit antecedent consists of a violation of presupposition and could be thus be equally characterized as a semantic anomaly (as in #). However, for ease of exposition, we use a separate symbol, ‘*’ for this condition.

TABLE 1 | Critical conditions with example stimuli.

		Object type	
		Congruent	Incongruent
Object	Definite	<i>The connoisseur tasted the wine during the tour.</i>	<i>The connoisseur tasted the roof during the tour.</i>
	Demonstrative	<i>The connoisseur tasted that wine during the tour.</i>	<i>The connoisseur tasted that roof during the tour.</i>

METHODS

Participants

Thirty seven Brock University undergraduates were recruited and either paid for their participation or received partial course credit. All participants were native, monolingual speakers of English, had normal or corrected-to-normal vision and were right-handed, as assessed by the Handedness Inventory. No participants reported any neurological impairment, history of neurological trauma, or use of neuroleptics.

Four participants with comprehension question accuracy for filler items (discussed below) at less than 85% were excluded from analysis leaving 33 eligible participants (25 females; mean age = 19.6; ranging from 18 to 25).

This study received ethics approval from the Brock University Bioscience Research Ethics Board (BREB) prior to the commencement of the experiment (REB 13-282). Written, informed consent was received from all participants prior to their participation in the experiment.

Materials

160 critical items (adapted from Dwivedi and Gibson, 2017) were presented in four conditions (see **Table 1**) counterbalanced across four lists. All sentences in this experiment were simple active sentences, using SUBJECT VERB OBJECT word order, followed by a prepositional phrase. Sentences varied according to two factors: object type (congruent vs. incongruent object) and determiner type (definite vs. demonstrative). All subjects were animate (e.g., *connoisseur*, *kid*) and preceded by the definite determiner *the*. An active, past-tense verb followed the subject (e.g., *tasted*, *climbed*). All direct objects were inanimate (e.g., *wine*, *roof*) and were either congruent with sentence context in the control conditions (e.g., *connoisseur – tasted – wine*, *kid – climbed – tree*) or incongruent with sentential context (e.g., *connoisseur – tasted – roof*, *kid – climbed – jade*). Direct objects were not repeated and were matched for word length item by item (e.g., *wine* vs. *#roof*). Next, word frequency for direct objects in congruent vs. incongruent conditions was controlled for, where log word frequencies (SUBTLEX-US database; see Brysbaert and New, 2009) indicated no significant difference, $t(159) = 0.63$, $p = 0.533$. Also, the direct object was never the final word in the sentence. Sentences ended with prepositional phrases that (crucially) did not alter the interpretation of the direct object. Instead, these phrases served to modify the event by referring to time (e.g., *in the morning*),

manner (e.g., *with difficulty*), and instrument (e.g., *with a pen*). Comprehension questions did not follow presentation of critical trials (stimuli available upon request).

To reduce predictability, 170 filler sentences were included, of varying syntactic and semantic structure.³ These sentences were 6–10 words in length and a subset of these (125 items, 38% of all trials) were followed by superficial Yes/No or True/False comprehension questions.

Offline Plausibility Ratings

We evaluated the plausibility of our critical materials by conducting a norming study using Qualtrics software, Version (March 2020) of the Qualtrics Research Suite (Qualtrics, 2020). Critical and filler sentences were rated in this web-based study according to plausibility on a scale from 0 (very implausible) to 5 (neutral) to 10 (very plausible), in steps of 0.1. The 160 critical items were presented in eight pseudorandomized, counterbalanced lists such that half of the critical items were presented in each list and each participant only saw each item once. 80 filler items were presented in all lists, for a total 160 items in each list. 34 participants completed the study, of which 30 met the eligibility criteria described above (as outlined in Section “Participants”). 10 participants were excluded for having a mean plausibility rating lower than seven on filler items (all of which were perfectly plausible). Data from the remaining 20 participants (20 females; mean age = 18.65; ranging from 18 to 25) were used to calculate plausibility ratings. A few trials had to be excluded due to software error (<0.3% of trials). Mean plausibility ratings for the critical conditions were: congruent definite (*the wine*; $M = 8.43$, $SD = 0.85$), congruent demonstrative (*that wine*; $M = 7.99$, $SD = 0.99$), incongruent definite (*the roof*; $M = 2.25$, $SD = 2.10$); incongruent demonstrative (*that roof*; $M = 2.05$, $SD = 2.07$). An ANOVA was conducted on the mean plausibility ratings with the independent variables of object type (congruent vs. incongruent), determiner type (definite vs. demonstrative). Significant main effects of object type, $F(1, 19) = 220.7$, $MSE = 3.33$, $p < 0.001$, $\eta_p^2 = 0.920$, and determiner type, $F(1, 19) = 6.00$, $MSE = 0.34$, $p = 0.024$, $\eta_p^2 = 0.240$ were observed. However, no significant interaction of object and determiner type was observed, $F(1, 19) = 1.60$, $MSE = 0.17$, $p = 0.221$, $\eta_p^2 = 0.078$. Overall, these results confirm the intended readings regarding congruent vs. incongruent sentences.

Electrophysiological Measures

Electroencephalographic (EEG) recordings were made using a 64-channel Active Two BioSemi system (BioSemi, Amsterdam). Data were sampled at a rate of 512 Hz and digitized with

³20 sentences began with a prepositional phrase (e.g., *After thirty minutes...*). 60 sentences began with a quantifier: (e.g., *Many*, *Most*). 20 sentences had direct objects as irregular plural nouns, half of which were plural (e.g., *teeth*, *women*) and half were singular, (e.g., *fireman*, *mouse*). The remaining 70 sentences were not controlled for any specific sentence construction. These filler items were used in other experiments in our lab.

a 24-bit analog-to-digital converter. EEG data were preprocessed offline using EMSE v5.5.1 software (Cortech Solutions, 2013). Two infinite impulse response filters were applied at 12 db/octave: a bandpass filter from 0.1 to 100 Hz used to remove high and low frequency noise and a bandstop filter from 59 to 61 Hz used to remove 60 Hz noise. All electrodes were re-referenced to the averaged mastoids for analysis. Prior to segmentation, eye movements artifacts and blinks were filtered from the data using a spatial ocular artifact correction algorithm (Pflieger, 2001). Due to equipment malfunction, data from electrode Fp1 was lost in some participants. A spatial interpolation filter (Cortech Solutions, 2013) was applied for this electrode, for all participants. Manual artifact rejection was applied. Epochs were created from an interval 200 ms prior to stimulus onset to 1,200 ms after stimulus onset.

Procedure

Participants were tested individually in one session of approximately 3 h. In each session, participants completed a short questionnaire regarding reading habits, a handedness inventory (Briggs and Nebes, 1975), and the PANAS (Watson et al., 1988) before the application of the electrodes. PANAS consists of 20 items (10 positive items, e.g., interested, excited, and 10 negative items, e.g., distressed, upset) for which the participants provided a response on a five-point Likert scale indicating the extent to which they felt this way (1 = “Very slightly or not at all,” 5 = “Extremely”; see Watson et al., 1988 for further details). Following a practice session of eight trials, each participant completed the experimental trials in six blocks of 55 trials, with rest periods between each block. Each participant saw one of four pseudorandomized, counterbalanced lists consisting of 330 items. The pseudorandomized lists were created using the Mix utility (van Casteren and Davis, 2006) such that the first three items and last two items of each block were always filler sentences; no more than two critical items were presented sequentially and items from the same condition were never presented sequentially. Using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA), sentences were presented in the center of the computer monitor (screen size 50.8 cm) in light gray, 18-point Courier New font on a black background at a viewing distance of approximately 70 cm. See **Figure 1** for a sample trial procedure (Psychology Software Tools, Inc. [E-Prime 2.0], 2012).

Each trial sentence began with the participant being prompted to press a button on the response pad, then the word “Blink” was presented for 1,000 ms, followed by a fixation cross (+) for 500 ms. After a variable inter-trial interval lasting between 200 and 400 ms, sentences were presented word-by-word in serial visual presentation mode with a stimulus onset asynchrony (SOA) of 600 ms and an inter-stimulus interval (ISI) of 200 ms. 125 filler items were followed by comprehension questions after the last word of the sentence, to which participants were asked to press a “1” or “2” key corresponding to answers on the screen using the response pad. Response time and accuracy was recorded for each response. The next trial began following another inter-trial interval lasting between 500 and 1,000 ms.

RESULTS

Behavioral Analyses

Filler Comprehension Questions

Comprehension rates for questions at filler conditions were at 95.20% ($SD = 2.75\%$), indicating that participants were indeed paying attention to sentence materials.

Electrophysiological Analyses

Figure 2 shows a topographic map for 33 participants, in the typical N400 time range (300–500 ms) after critical word (*wine/#roof*) onset. It shows a large, broadly distributed N400 effect with slight right lateralization in the definite condition, and a smaller N400 effect, constrained to centroparietal sites, also with slight right lateralization, in the demonstrative condition. Given this broad (and typical, see Federmeier and Kutas, 1999) distribution, we focus analyses at midline sites (see grand average ERPs, **Figure 3**).

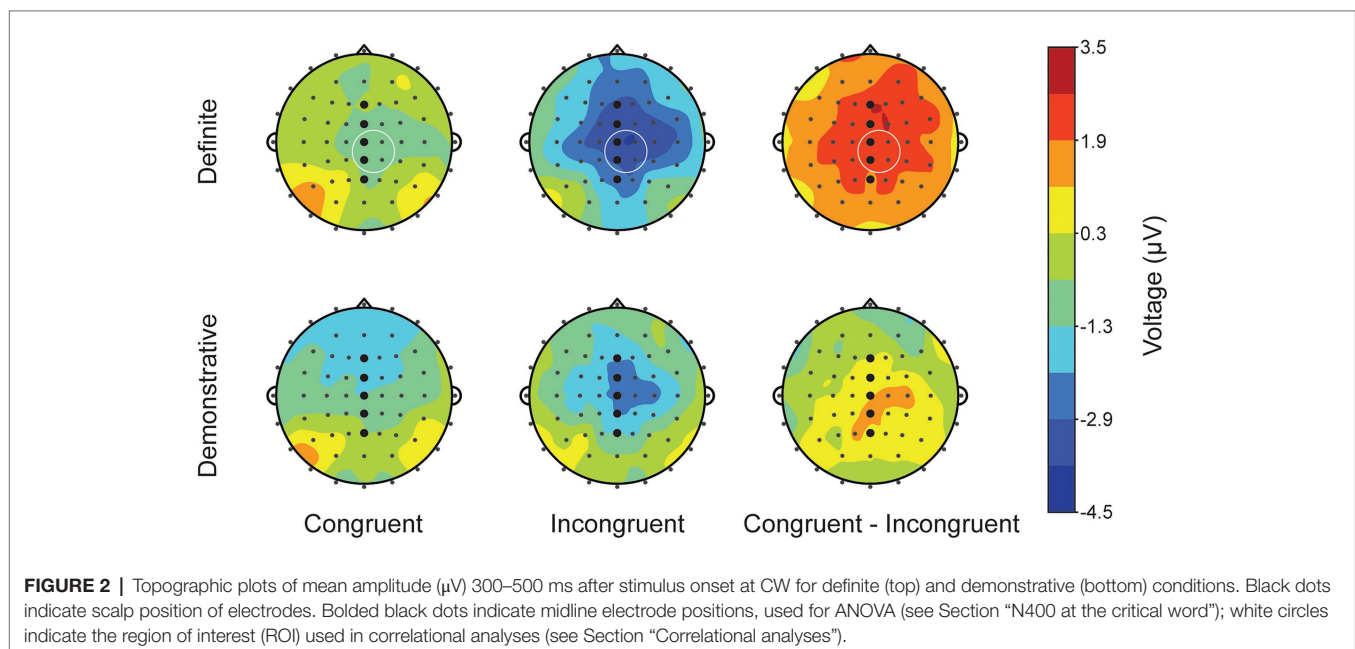
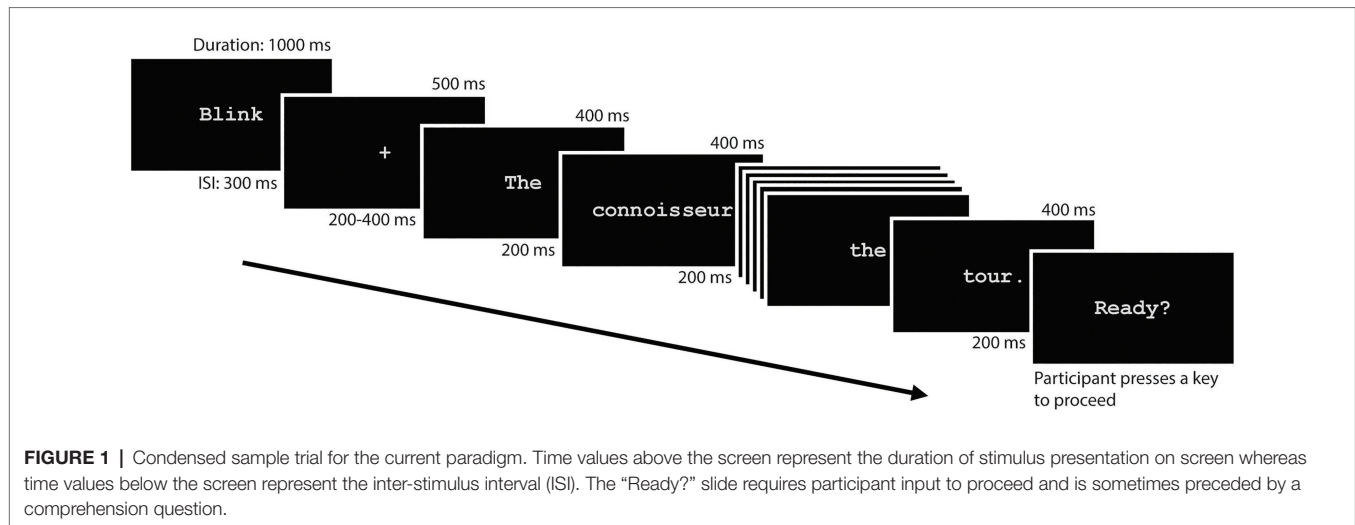
N400 at the Critical Word

A repeated measures ANOVA was conducted for midline electrode sites (Fz, FCz, Cz, CPz, and Pz) at the critical word (*wine* vs. *#roof*) on mean voltage in the traditional N400 time window (300–500 ms) using SPSS (IBM Corp, 2018) statistical software (see **Table 2**). We employed the Greenhouse-Geisser (Greenhouse and Geisser, 1959) non-sphericity correction for effects with more than one degree of freedom in the numerator. Following convention, unadjusted degrees of freedom are reported, along with the Greenhouse-Geisser epsilon value (ϵ) and adjusted p value. Mean square error values reported are those corresponding to the Greenhouse-Geisser correction. Partial eta squared (η_p^2) is reported as a measure of effect size and *post hoc* comparisons are Bonferroni corrected.

Here, we observed the incongruent condition was significantly more negative than its control in the definite condition at all midline electrode sites (*the wine/#roof*; $\Delta = 1.76$ – $2.65 \mu V$, all values of $p < 0.001$; see **Figure 3A**), whereas the incongruent condition was significantly more negative than its control in the demonstrative condition only at electrode site CPz (*that wine/#roof*; $\Delta = 1.42 \mu V$, $p = 0.006$; all other values of $p > 0.05$; see **Figure 3B**). Thus, a large, broadly distributed N400 effect was observed for a semantic anomaly preceded by the definite determiner *the*, whereas a heavily attenuated N400 effect restricted to one site was observed for a semantic anomaly preceded by the demonstrative determiner *that*.

Correlational Analyses

To investigate the relationship between N400 amplitude and dispositional affect, we computed difference scores by subtracting the mean amplitude at incongruent conditions from congruent conditions in the definite condition, which had elicited a robust N400 effect (e.g., *the wine* – *the #roof*). Based on visual inspection of the N400 effect (see **Figure 2**), we chose a region of interest (ROI) consisting of right lateralized centroparietal electrode sites (Cz, C2, CPz, and CP2) where the effect was maximal, a region consistent with and typical of N400 scalp topography



(Federmeier and Kutas, 1999; Kutas and Federmeier, 2011). We computed the difference in the mean amplitude at the average of these electrode sites at 300–500 ms following stimulus onset. These difference scores were correlated with PA ($M = 27.61$; $SD = 6.89$; ranging from 14 to 35) and Negative Affect (NA; $M = 15.73$; $SD = 4.71$; ranging from 10 to 28) scores.

We did not observe a significant correlation between PA scores and N400 amplitudes (see **Figure 4A**), $r(31) = -0.15$, $p = 0.419$.⁴ We observed a moderate negative correlation between NA scores and N400 amplitude (see **Figure 4B**), $r(31) = -0.36$, $p = 0.041$. That is, participants with larger NA scores had

smaller N400 amplitudes (conversely, participants with smaller NA scores had larger N400 amplitudes).

DISCUSSION

In the present study, we examined whether N400 responses to incongruent objects in sentences would be modulated by dispositional affect scores. Based on previous work on mood and language processing, we predicted (i) a significant positive correlation between N400 amplitude and positive affect scores, in addition to (ii) a negative correlation for this ERP component and negative affect scores. Our second prediction was borne out in this study. That is, N400 effects at incongruent *#the roof* vs. *the wine*, were modulated, such that more negative participants had smaller N400 amplitudes, whereas no correlation

⁴While the N400 effect was attenuated in the demonstrative condition, as it was still present at electrode CPz, we conducted the correlational analyses here using same ROI. No significant correlations were observed here for either PA, $r(31) = -0.02$, $p = 0.905$, or NA, $r(31) = -0.01$, $p = 0.967$.

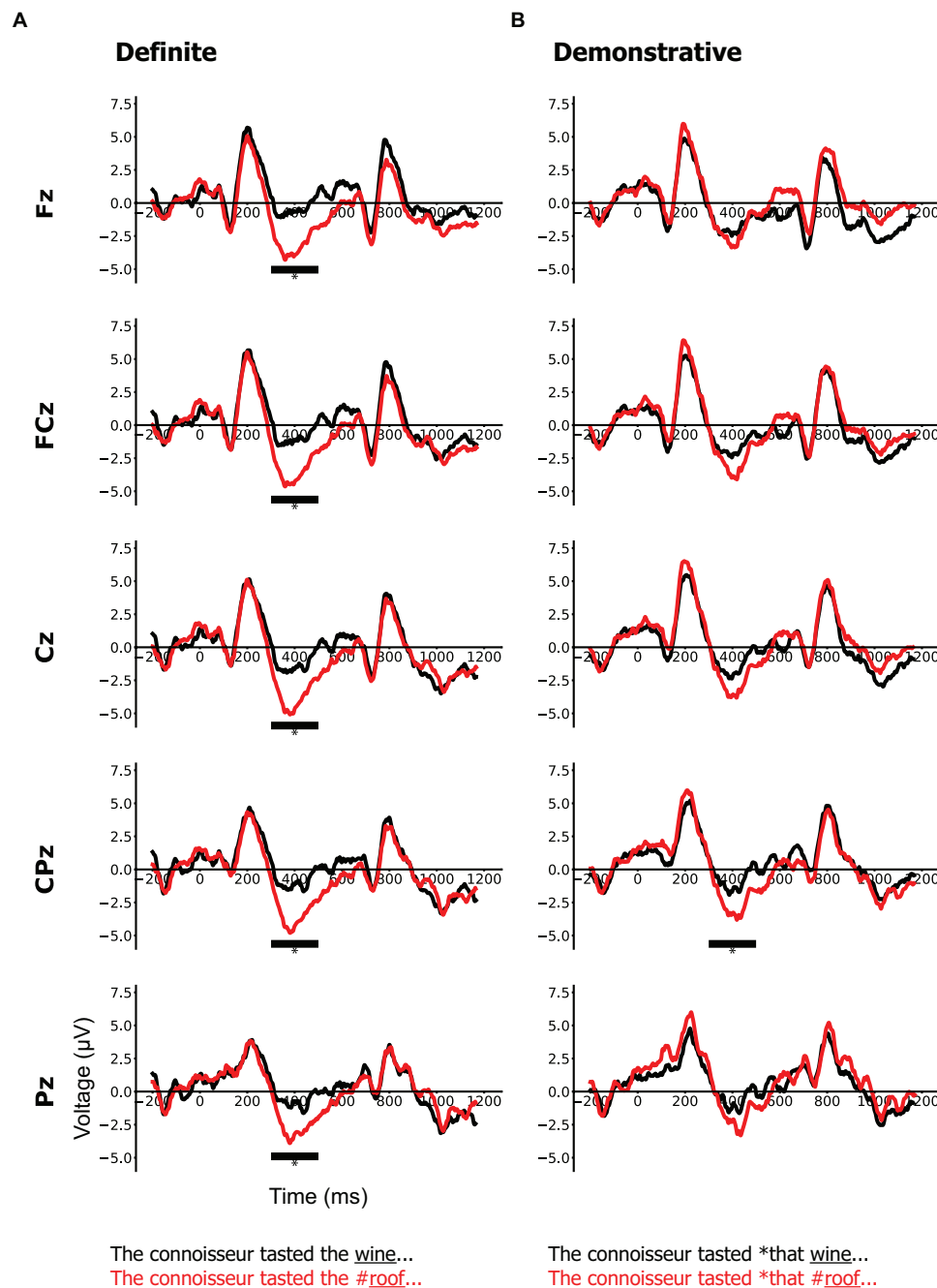


FIGURE 3 | Grand average event-related potential (ERP) waveforms for object type condition: congruent (black) and incongruent (red) separately for definite (A) and demonstrative (B) conditions at five midline electrodes (Fz, FCz, Cz, CPz, and Pz) time-locked to the onset of the critical word (*wine*/*#roof*). Asterisks * indicate significant amplitude differences (see Section “N400 at the critical word”).

was found with positive affect. In a separate question, we also examined how neural responses to sentences containing “double violations,” as in **that #roof* might differ from those with single violations. Results revealed an attenuated N400 response. This question and results are addressed in separate work (Dwivedi and Selvanayagam, 2020a,b) and we do not further discuss that question, to not detract from the issue at hand.

Below, we address the cognitive significance of our findings regarding dispositional affect and sentence perception. Specifically, we do so relying on previous work in emotion literature (Schwarz and Clore, 2007), which indicated that positive affect promotes global processing (focusing on the forest vs. the trees) vs. negative affect, which promoted more local processing (focusing on the trees vs. the forest).

Dispositional Affect and Conceptual Event Semantic Processing

To reiterate, we found no correlation with positive affect and N400 amplitude, in contrast to a significant negative correlation with negative affect. Previously, Chwilla et al. (2011) found that participants induced into happy moods produced large N400 effects, whereas those in sad moods produced attenuated effects. A corollary of the latter finding is that low negative affect individuals produce larger N400 effects – exactly what we found. We interpret the cognitive significance of our result to mean that low negative affect can facilitate contextual or heuristic processing, which we discuss in further detail below.

Real-time sentence comprehension necessarily involves global context/event knowledge for interpretation (Barton and Sanford, 1993; Zwaan and Radvansky, 1998; Ferretti et al., 2007; Dwivedi et al., 2018). We claim that low negative affect promotes local associations between words in sentences and in this way facilitates sensitivity to contextual knowledge (Chwilla et al., 1998). Here, we draw on work by Huntsinger et al. (2010, p. 725) who indicate that “a global focus should result either from having an accessible global orientation empowered by positive affect, or a local orientation inhibited by negative affect.” We draw on this logic and reason that if negative affect inhibits a local

orientation, then lower negative affect would lessen that inhibition. Less inhibition between words would allow for more spreading activation and build stronger lexical expectations between words during sentence comprehension (Kamide et al., 2003; Madden and Zwaan, 2003; McRae et al., 2005). A natural outcome of this increased spreading activation between words would be stronger lexical expectations during sentence comprehension (Collins and Loftus, 1975; Metusalem et al., 2012). In this way, larger N400 amplitudes are elicited for low negative affect individuals when lexical expectations are violated. In contrast, a high negative affect individual would strongly inhibit local lexical meaning, yielding weaker links between words in a sentence and therefore exhibit less sensitivity to incongruent meaning – thereby producing attenuated N400 effects.

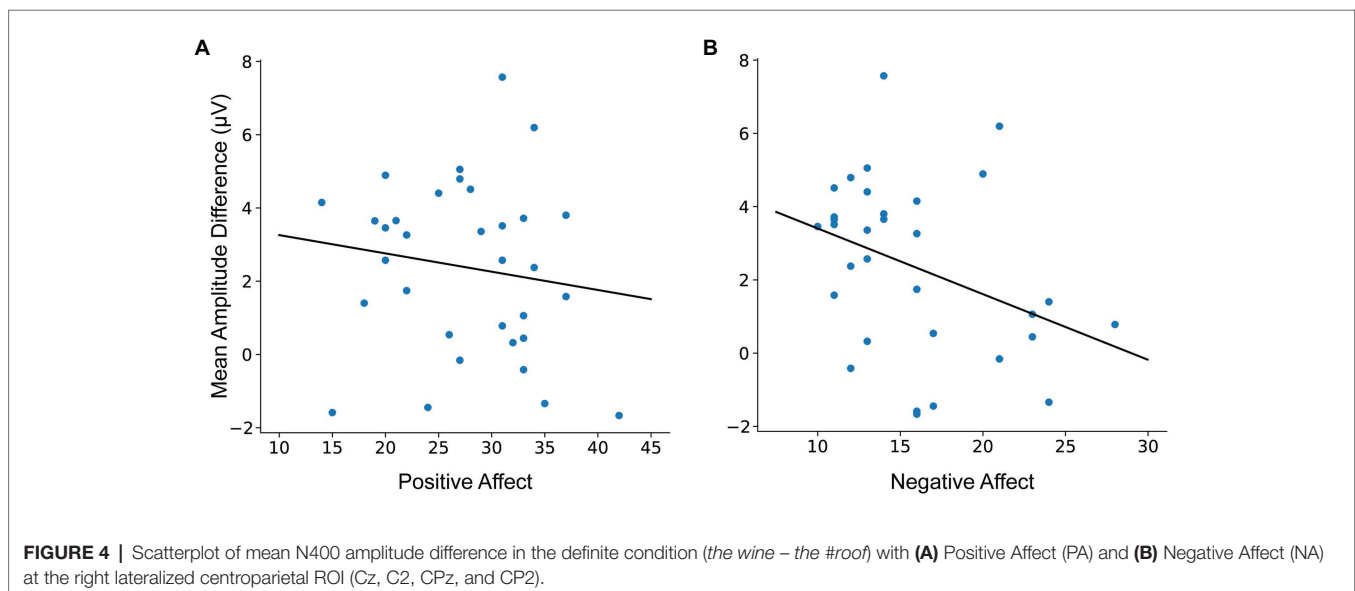
We speculate here that the account above could explain the lack of an effect found with positive affect in our experiment, vs. what was found previously in Chwilla et al. (2011). We claim that differences in experimental stimuli design between our work and Chwilla et al. (2011) resulted in different processing strategies, which would be differentially sensitive to affect. Recall that sentence stimuli of Chwilla et al. (2011) consisted of high-cloze vs. low-cloze word manipulations (e.g., *stuffing pillows with feathers* vs. *stuffing pillows with #books*). This manipulation meant that sentences with high-cloze words yielded meanings that were (by definition) consistent with event knowledge, that is, these were stereotypical events. In contrast, sentences inconsistent with real-world experience, (i.e., those with low-cloze words) yielded improbable or unlikely events. As such, conceptual event semantics was an important cue to access and attend to in order to compute meaning coherence.⁵ In contrast, meanings associated with our sentence stimuli with incongruent words consisted of impossible, not improbable events (to paraphrase above example, e.g., *stuffing pillows with #clouds*). As a result,

⁵Accessibility of this cue was further strengthened since half of all trials contained pseudo-homophones, yielding sentences with no propositional meaning.

TABLE 2 | *F*-values for midline analyses of Object and Determiner type at critical word for time window 300–500 ms.

Effect (df)	<i>F</i>	<i>p</i>	<i>MSE</i>	η_p^2
O (1, 32)	30.982	<0.001*	13.727	0.492
D (1, 32)	1.362	0.252	18.283	0.041
E (4, 128)	5.580	0.004*	5.655	0.148
O * D (1, 32)	5.047	0.032*	15.920	0.136
O * E (4, 128)	1.270	0.290	2.038	0.038
D * E (4, 128)	3.346	0.026*	1.843	0.095
O * D * E (4, 128)	3.221	0.047*	1.764	0.091

O, object type; D, determiner type; and E, electrode site. * < 0.05.



participants did not need to access top-down information regarding stereotypical events in the same way – word by word association and integration was enough to construct a coherent mental representation. As such, the design of our materials did not necessitate a processing strategy associated with accessing global event knowledge – a strategy linked with positive affect (Ashby et al., 1999; Isen and Reeve, 2005; Talarico et al., 2009; Gable and Harmon-Jones, 2010; Selvanayagam et al., 2019). In fact, preliminary evidence suggests that we are on the right track. Given that sentence plausibility is a good measure of conceptual event-based expectations (Matsuki et al., 2011), we note that the off-line plausibility ratings reported above show comparable correlations with negative ($r = -0.438$) and positive ($r = 0.069$) affect. This result supports our preliminary explanation.

CONCLUSION

Our work showed a significant N400 effect for sentences such as *The connoisseur tasted the wine/#roof on the tour*, which was modulated by dispositional negative affect. This is the first study to show that dispositional affect can modulate N400 magnitude, furthering our understanding of individual differences associated with this component (Tanner et al., 2013; Grey et al., 2017). Furthermore, the findings are consistent with studies in affect and other domains of cognition that suggest a link between global and local perception of stimuli and positive and negative affect. Finally, our work supports views espoused by Münster and Knoefler (2018), and others that relevant “contextual cues” regarding real-time language processing includes characteristics of the very human being perceiving sentences in context.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Ashby, F. G., Isen, A. M., and Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychol. Rev.* 106, 529–550. doi: 10.1037/0033-295X.106.3.529
- Barton, S. B., and Sanford, A. J. (1993). A case study of anomaly detection: shallow semantic processing and cohesion establishment. *Mem. Cogn.* 21, 477–487. doi: 10.3758/BF03197179
- Brief, A. P., Butcher, A. H., and Roberson, L. (1995). Cookies, disposition, and job attitudes: the effects of positive mood-inducing events and negative affectivity on job satisfaction in a field experiment. *Organ. Behav. Hum. Decis. Process.* 62, 55–62. doi: 10.1006/obhd.1995.1030
- Briggs, G. G., and Nebes, R. D. (1975). Patterns of hand preference in a student population. *Cortex* 11, 230–238. doi: 10.1016/S0010-9452(75)80005-0
- Brysbaert, M., and New, B. (2009). Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977
- Chwilla, D. J., Hagoort, P., and Brown, C. M. (1998). The mechanism underlying backward priming in a lexical decision task: spreading activation versus semantic matching. *Q. J. Exp. Psychol.* 51, 531–560. doi: 10.1080/17475773

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Brock University Bioscience Research Ethics Board (BREB) prior to the commencement of the experiment (REB 13-282). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

VDD conceived and designed the experiment. JS and VDD conducted the experiment, analyzed the data, and wrote the paper. Both the authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the Canada Foundation for Innovation (grant number 13831), and Social Sciences and Humanities Research Council grants (430-2016-00740 and 435-2019-1308) awarded to VDD. JS was supported by Match of Minds summer funds from Brock University. Open Access publication fees were supported by the Brock Library Open Access Publishing Fund.

ACKNOWLEDGMENTS

We would like to thank Alanna Kozak, Harmonie Chan, Brent Dryczewycz, Anushka Lalwani, and Azaan Adnan for help in running participants. In addition, we thank Louis Schmidt for comments and discussion, as well as audiences at Maastricht University, McGill University, and the Cognitive Neuroscience Society 2019, where versions of this work were presented. Finally, we also like to thank Aaron Ayub for additional help in manuscript preparation.

- Chwilla, D. J., and Kolk, H. H. J. (2005). Accessing world knowledge: evidence from N400 and reaction time priming. *Cogn. Brain Res.* 25, 589–606. doi: 10.1016/j.cogbrainres.2005.08.011
- Chwilla, D. J., Virgillito, D., and Vissers, C. T. W. M. (2011). The relationship of language and emotion: N400 support for an embodied view of language comprehension. *J. Cogn. Neurosci.* 23, 2400–2414. doi: 10.1162/jocn.2010.21578
- Collins, A., and Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 10.1037/0033-295X.82.6.407
- Cortech Solutions. (2013). EMSE suite (version 5.5.1). Cortech Solutions.
- Dale, G., and Arnell, K. M. (2010). Individual differences in dispositional focus of attention predict attentional blink magnitude. *Atten. Percept. Psychophysiol.* 72, 602–606. doi: 10.3758/APP.72.3.602
- Dwivedi, V. D. (2013). Interpreting quantifier scope ambiguity: evidence of heuristic first, algorithmic second processing. *PLoS One* 8:e81461. doi: 10.1371/journal.pone.0081461
- Dwivedi, V. D., and Gibson, R. M. (2017). An ERP investigation of quantifier scope ambiguous sentences: evidence for number in events. *J. Neurolinguistics* 42, 63–82. doi: 10.1016/j.jneuroling.2016.11.006
- Dwivedi, V. D., Goertz, K. E., and Selvanayagam, J. (2018). Heuristics in language comprehension. *J. Behav. Brain Sci.* 8, 430–446. doi: 10.4236/jbbs.2018.87027

- Dwivedi, V. D., and Selvanayagam, J. (2020a). An electrophysiological investigation of conceptual vs. compositional semantic anomaly. Ms. under review.
- Dwivedi, V. D., and Selvanayagam, J. (2020b). "Distinct neurocognitive effects of compositional vs. conceptual semantic violations." in *Poster presentation for the 12th annual meeting of the Society for Neurobiology of Language*. Philadelphia, USA [virtual].
- Federmeier, K. D., Kirson, D. A., Moreno, E. M., and Kutas, M. (2001). Effects of transient, mild mood states on semantic memory organization and use: an event-related potential investigation in humans. *Neurosci. Lett.* 305, 149–152. doi: 10.1016/S0304-3940(01)01843-2
- Federmeier, K. D., and Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495. doi: 10.1006/jmla.1999.2660
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cogn. Psychol.* 47, 164–203. doi: 10.1016/S0010-0285(03)00005-7
- Ferretti, T. R., Kutas, M., and McRae, K. (2007). Verb aspect and the activation of event knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 182–196. doi: 10.1037/0278-7393.33.1.182
- Fredrickson, B. L., and Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognit. Emot.* 19, 313–332. doi: 10.1080/02699930441000238
- Gable, P., and Harmon-Jones, E. (2010). The motivational dimensional model of affect: implications for breadth of attention, memory, and cognitive categorisation. *Cognit. Emot.* 24, 322–337. doi: 10.1080/02699930903378305
- Gasper, K., and Clore, G. L. (2002). Attending to the big picture: mood and global versus local processing of visual information. *Psychol. Sci.* 13, 34–40. doi: 10.1111/1467-9280.00406
- Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika* 24, 95–112. doi: 10.1007/BF02289823
- Grey, S., Tanner, D., and Van Hell, J. G. (2017). How right is left? Handedness modulates neural responses during morphosyntactic processing. *Brain Res.* 1669, 27–43. doi: 10.1016/j.brainres.2017.05.024
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *J. Cogn. Neurosci.* 15, 883–899. doi: 10.1162/089892903322370807
- Huntsinger, J. R., Clore, G. L., and Bar-Anan, Y. (2010). Mood and global-local focus: priming a local focus reverses the link between mood and global-local processing. *Emotion* 10, 722–726. doi: 10.1037/a0019356
- IBM Corp (2018). SPSS (Version 27.0) (No. 27). IBM Corp.
- Isen, A. M., and Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: facilitating enjoyment of play, responsible work behavior, and self-control. *Motiv. Emot.* 29, 297–325. doi: 10.1007/s11031-006-9019-8
- Kamide, Y., Altmann, G. T. M., and Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J. Mem. Lang.* 49, 133–156. doi: 10.1016/S0749-596X(03)00023-8
- Kimchi, R., and Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 521–535. doi: 10.1037//0096-1523.8.4.521
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., and Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657
- Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 137, 161–163.
- Larsen, R. J., and Ketelaar, T. (1989). Extraversion, neuroticism and susceptibility to positive and negative mood induction procedures. *Personal. Individ. Differ.* 10, 1221–1228. doi: 10.1016/0191-8869(89)90233-X
- Larsen, R. J., and Ketelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. *J. Pers. Soc. Psychol.* 61, 132–140. doi: 10.1037/0022-3514.61.1.132
- Levin, I., and Stokes, J. P. (1989). Dispositional approach to job satisfaction: role of negative affectivity. *J. Appl. Psychol.* 74, 752–758. doi: 10.1037/0021-9010.74.5.752
- Madden, C. J., and Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Mem. Cogn.* 31, 663–672. doi: 10.3758/BF03196106
- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., and McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 913–934. doi: 10.1037/a0022964
- McRae, K., Hare, M., Elman, J. L., and Ferretti, T. R. (2005). A basis for generating expectancies for verbs from nouns. *Mem. Cogn.* 33, 1174–1184. doi: 10.3758/BF03193221
- Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., and Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *J. Mem. Lang.* 66, 545–567. doi: 10.1016/j.jml.2012.01.001
- Münster, K., and Knoeferle, P. (2018). Extending situated language comprehension (accounts) with speaker and comprehender characteristics: toward socially situated interpretation. *Front. Psychol.* 8:2267. doi: 10.3389/fpsyg.2017.02267
- Pflieger, M. E. (2001). "Theory of a spatial filter for removing ocular artifacts with preservation of EEG." in *EMSE Workshop*. September 2001. 7–8.
- Psychology Software Tools, Inc. [E-Prime 2.0] (2012). Available at: <https://www.pstnet.com>
- Qualtrics (2020). Qualtrics (version March 2020). Qualtrics: Provo, Utah, USA.
- Roberts, C. (1989). Modal subordination and pronominal anaphora in discourse. *Linguist. Philos.* 12, 683–721. doi: 10.1007/BF00632602
- Rowe, G., Hirsh, J. B., and Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *Proc. Natl. Acad. Sci. U. S. A.* 104, 383–388. doi: 10.1073/pnas.0605198104
- Schank, R., and Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, New Jersey: Erlbaum.
- Schwarz, N. (2002). "Situated cognition and the wisdom of feelings: cognitive tuning" in *The wisdom in feelings: Psychological processes in emotional intelligence*. eds. L. Feldman Barrett and P. Salovey (New York: Guilford Press), 144–166.
- Schwarz, N., and Clore, G. L. (2007). "Feeling and phenomenal experiences" in *Social psychology: Handbook of basic principles*. 2nd Edn. eds. A. W. Kruglanski and E. T. Higgins (New York: The Guilford Press), 385–407.
- Selvanayagam, J., Witte, V., Schmidt, L. A., and Dwivedi, V. D. (2019). A preliminary investigation of dispositional affect, the P300, and sentence processing. *Brain Res.* 1721. doi: 10.1016/j.brainres.2019.146309
- Staw, B. M., and Ross, J. (1985). Stability in the midst of change: A dispositional approach to job attitudes. *J. Appl. Psychol.* 70, 469–480. doi: 10.1037/0021-9010.70.3.469
- St. George, M., Mannes, S., and Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. *J. Cogn. Neurosci.* 6, 70–83. doi: 10.1162/jocn.1994.6.1.70
- Talarico, J. M., Berntsen, D., and Rubin, D. C. (2009). Positive emotions enhance recall of peripheral details. *Cognit. Emot.* 23, 380–398. doi: 10.1080/02699930801993999
- Tanner, D., McLaughlin, J., Herschensohn, J., and Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Biling. Lang. Cogn.* 16, 367–382. doi: 10.1017/S1366728912000302
- Townsend, D. J., and Bever, T. G. (2001). *Sentence comprehension: The integration of rules and habits*. Cambridge, MA: MIT Press.
- van Berkum, J. J. A., De Goede, D., van Alphen, P. M., Mulder, E. R., and Kerstholt, J. H. (2013). How robust is the language architecture? The case of mood. *Front. Psychol.* 4:505. doi: 10.3389/fpsyg.2013.00505
- van Casteren, M., and Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behav. Res. Methods* 38, 584–589. doi: 10.3758/BF03193889
- Visser, C. T. W. M., Chwilla, U. G., Egger, J. I. M., and Chwilla, D. J. (2013). The interplay between mood and language comprehension: evidence from P600 to semantic reversal anomalies. *Neuropsychologia* 51, 1027–1039. doi: 10.1016/j.neuropsychologia.2013.02.007
- Visser, C. T. W. M., Virgillito, D., Fitzgerald, D. A., Speckens, A. E. M., Tendolkar, I., van Oostrom, I., et al. (2010). The influence of mood on the processing of syntactic anomalies: evidence from P600. *Neuropsychologia* 48, 3521–3531. doi: 10.1016/j.neuropsychologia.2010.08.001
- Watson, D., and Clark, L. A. (1984). Negative affectivity: the disposition to experience aversive emotional states. *Psychol. Bull.* 96, 465–490.
- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Zwaan, R. A., and Radvansky, G. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185. doi: 10.1037/0033-2909.123.2.162

Conflict of Interest: 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 © 2021 Dwivedi and Selvanayagam. 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.



Reading About Us and Them: Moral but no Minimal Group Effects on Language-Induced Emotion

Björn 't Hart¹, Marijn Struiksma¹, Anton van Boxtel² and Jos J. A. van Berkum^{1*}

¹Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands, ²Cognitive Neuropsychology, Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands

OPEN ACCESS

Edited by:

Pia Knoeferle,
Humboldt University of Berlin,
Germany

Reviewed by:

Johanna Maria Kissler,
Bielefeld University, Germany
David A. Havas,
University of Wisconsin–Whitewater,
United States

*Correspondence:

Jos J. A. van Berkum
j.vanberkum@uu.nl

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 31 July 2020

Accepted: 27 April 2021

Published: 20 May 2021

Citation:

't Hart B, Struiksma M, van Boxtel A
and van Berkum JJA (2021) Reading
About Us and Them: Moral but no
Minimal Group Effects on Language-
Induced Emotion.
Front. Commun. 6:590077.
doi: 10.3389/fcomm.2021.590077

Many of our everyday emotional responses are triggered by language, and a full understanding of how people use language therefore also requires an analysis of how words elicit emotion as they are heard or read. We report a facial electromyography experiment in which we recorded corrugator supercilii, or “frowning muscle”, activity to assess how readers processed emotion-describing language in moral and minimal in/outgroup contexts. Participants read sentence-initial phrases like “Mark is angry” or “Mark is happy” after descriptions that defined the character at hand as a good person, a bad person, a member of a minimal ingroup, or a member of a minimal outgroup (realizing the latter two by classifying participants as personality “type P” and having them read about characters of “type P” or “type O”). As in our earlier work, moral group status of the character clearly modulated how readers responded to descriptions of character emotions, with more frowning to “Mark is angry” than to “Mark is happy” when the character had previously been described as morally good, but *not* when the character had been described as morally bad. Minimal group status, however, did not matter to how the critical phrases were processed, with more frowning to “Mark is angry” than to “Mark is happy” across the board. Our morality-based findings are compatible with a model in which readers use their emotion systems to simultaneously simulate a character’s emotion and evaluate that emotion against their own social standards. The minimal-group result does not contradict this model, but also does not provide new evidence for it.

Keywords: psycholinguistics, communication, emotion, embodiment, morality, minimal groups, EMG, facial electromyography

INTRODUCTION

Part of the attraction of reading a story is that we can vicariously experience what it is like to be somebody else. For example, we can experience happiness when characters in a story find love, frustration when they quarrel, and sadness when they break up—all the while reclining in our armchairs or waiting for the train. Such vicarious experiences can take our mind off things or help us pass the time, provide entertainment, and help us learn about others, life, and possibly even ourselves.

Interestingly, “vicarious experience” may well have a literal meaning here. Partly motivated by the realization that the meaning of at least *some* concepts must be grounded in actual bodily experience (Barsalou, 2008), research on embodied language processing has indicated that reading or hearing a word can lead to a *simulation* of concrete experiences involving the concept, via the neural re-instantiation of perceptual, motor and other experience-induced states associated with what the

concept or phrasal combination of concepts is *about* (Barsalou, 2009; Vigliocco et al., 2009; Glenberg, 2011; Glenberg and Gallese, 2012; Havas and Matheson, 2013; Zwaan, 2014; Winkielman et al., 2015; Zwaan, 2016; Fingerhut and Prinz, 2018; Winkielman et al., 2018). For example, reading action words like “kick” or “pick” leads to activation of the motor cortex involved in *actually realizing* the described movements (Pülvermüller et al., 2005; Willems and Casasanto, 2011), and reading phrases such as “he saw an eagle in the sky” leads to a perceptual simulation of the described situation (Zwaan et al., 2002; Zwaan and Pecher, 2012). Such research suggests that when people process emotion words like “happy” or “angry”, they may actually reuse emotion-related neural systems (Anderson, 2010) to mentally simulate the emotional state described by the language at hand.

Compatible with this simulation idea, studies that use electromyography (EMG) to track subtle facial muscle activity have suggested that simply reading or hearing “angry” or “she is angry” leads to rapid contraction of the corrugator supercilli or ‘frowning muscle’, and, conversely, that reading or hearing “happy”, or “she is happy” leads to rapid contraction of the zygomaticus major, the cheek muscle involved in smiling (e.g., Foroni and Semin, 2009; Glenberg et al., 2009; Foroni and Semin, 2013; Künecke et al., 2015; Fino et al., 2016; see van Berkum et al., in press for review). The central idea here is not that people need to actually *move their face* to make sense of emotion words and phrases, but that the comprehension of an emotion word involves the spontaneous partial reinstatement of the described emotional state (including traces of the associated facial expression), as if one is having the emotion oneself. This reinstatement would occur as part of the retrieval of word meaning from memory (e.g., Foroni and Semin, 2009; Künecke et al., 2015), and/or as part of constructing a situation model in which some concrete character is having an emotion (e.g., Glenberg et al., 2009; Fino et al., 2016).

Evidence that readers use their emotion systems to *simulate* linguistic meaning poses an interesting puzzle, because during everyday language comprehension, people obviously also need their emotion systems for their *primary* function, which is to—consciously or unconsciously—*evaluate* how events in the world relate to their own concerns (e.g., Lazarus, 1991; Frijda, 2007; Tooby and Cosmides, 2008; Scarantino, 2014; van Berkum, 2018; Scherer and Moors, 2019; see van Berkum, in press, for review). Emotional evaluation is what makes us feel good over a verbal compliment, scared when receiving an unfavourable medical diagnosis, worried over what we read in the newspaper, or surprised by a fictional character’s actions in a novel. These everyday examples suggest that we continuously use our emotion systems to evaluate what we read or hear. So how does such language-driven emotional evaluation mesh with language-driven emotion simulation? When processing language, do we *simultaneously* use our emotion systems to simulate somebody else’s described emotion *and* to evaluate, i.e., have our own emotions about, what is described? If so, how? And if not, which of the two potential uses of our emotion systems receives priority?

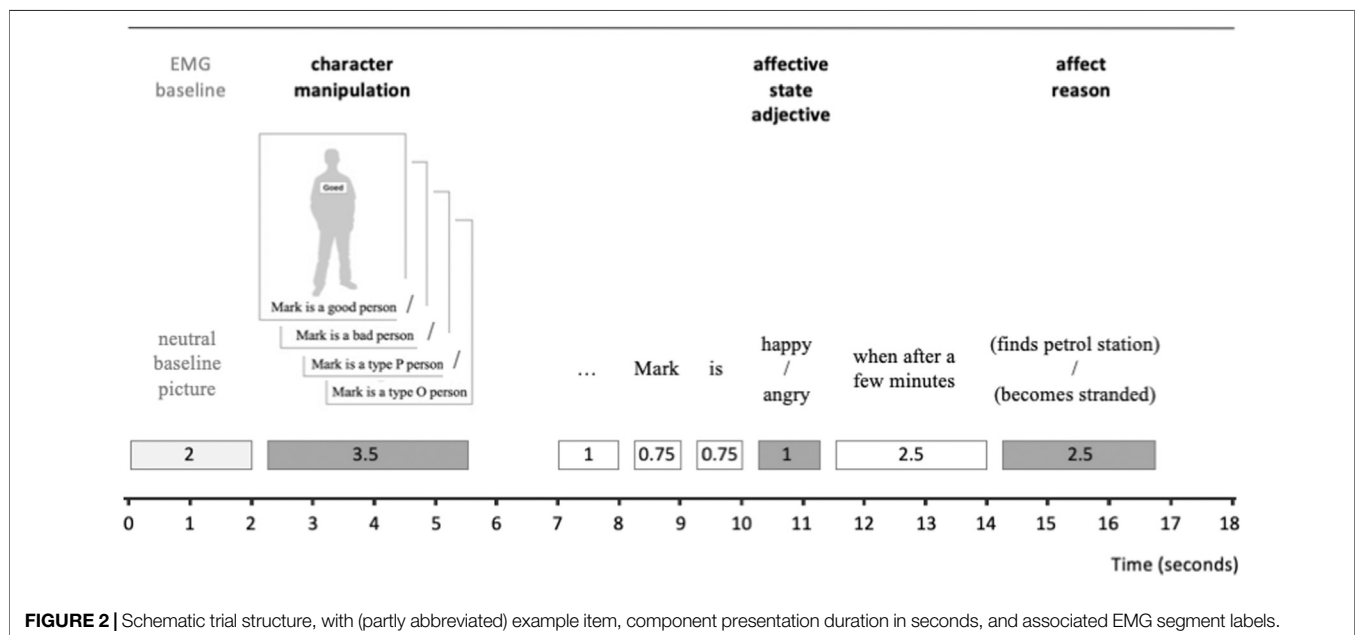
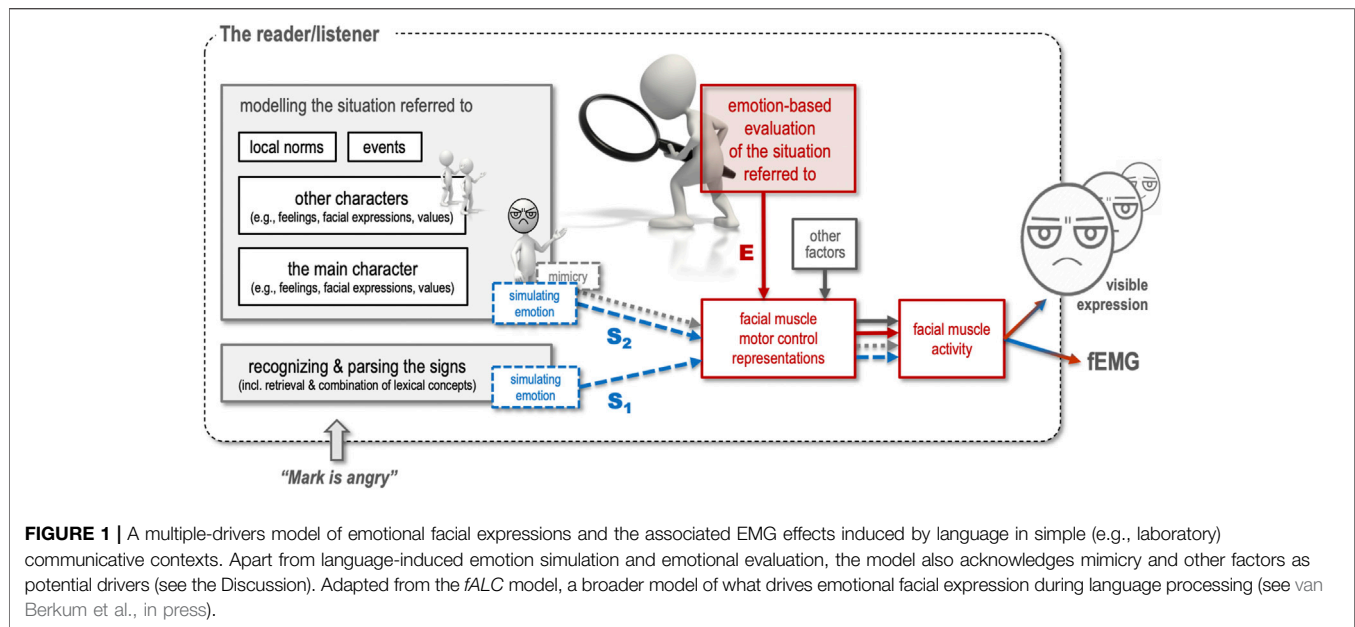
We explored this issue in two prior EMG-studies (’t Hart et al., 2018; ’t Hart et al., 2019), where we embedded phrases like “Mark

was angry” or “Mark was happy” in a narrative context that was designed to promote simulation as well as evaluation. Specifically, we compared the processing at negative or positive emotional state adjectives, e.g., “angry” vs. “happy”, in stories where the character experiencing those states had previously displayed morally good or morally bad behavior. We reasoned that any lexical and/or situation model simulation should in principle always generate more negative emotion at “Mark was angry” than at “Mark was happy”, independent of whether the character was morally good or bad. The reader’s moral evaluation of events, however, should depend on *who* the event is happening to, at least to some extent. When something bad happens to a morally *good* character, this should typically be seen as “unfair” or otherwise undesirable, and something good happening to him or her should be seen as desirable (as in a “feel-good” movie). Something bad happening to a morally bad character, however, should typically elicit a sense of fairness or “justice being served”, perhaps even a bit of Schadenfreude (e.g., Feather and Nairn, 2005; Singer et al., 2006; Leach and Spears, 2009; Cikara and Fiske, 2012), and something good happening to him or her should typically be seen as “unfair”.

Because our logic was cast in terms of the valence (positivity or negativity) of language-induced emotion, we looked for traces of reader emotion by recording EMG over the corrugator or “frowning” muscle, a sensitive and reliable indicator of valence (e.g., Larsen et al., 2003; Höfling et al., 2020; see van Bortel, in press; van Berkum et al., in press, for reviews). The EMG-results were very clear. In both studies, phrases like “Mark was angry” led to stronger corrugator activity than phrases like “Mark was happy” when the character had previously acted in a morally good way, but *not* when the character had previously acted in a morally bad way—in the latter case, corrugator activity to negative and positive emotion adjectives did not differ. Because simple models involving *only* simulation or evaluation cannot easily explain these results, we converged on a *multiple-drivers* model of corrugator activity during language comprehension (see **Figure 1**; adapted from van Berkum et al., in press), which, in our materials, would involve both simulation (at the lexical and/or situation model) *and* evaluation of what is being asserted.

Our multiple-drivers account proposed that in the case of a good character, negative emotion induced by simulation at “Mark is angry” adds up with the negative emotional evaluation associated with an undesirable outcome, and positive emotion induced by simulation at “Mark is happy” adds up with the positive emotional evaluation associated with a desirable outcome, leading to a much stronger corrugator activity at negative emotion words, as compared to positive ones. In the case of a bad character, however, negative emotion induced by simulation at “Mark is angry” is *counteracted* by the positive emotional evaluation of a “fair” outcome, and positive emotion induced by simulation at, e.g., “Mark is happy” is counteracted by the negative emotional evaluation of an “unfair” outcome, to such an extent that, with our materials, no net valence effect at negative vs. positive emotion words remains.

While adequate, this account of the ’t Hart et al. (2018), ’t Hart et al. (2019) results is only modestly parsimonious, as it explains a null result as the net effect of two counteracting forces. The



principal aim of the current experiment was to try to expose the counteracting drivers by “downtuning” the force of emotional evaluation. Morality is deeply intertwined with ingroup cohesion and intergroup competition (Haidt, 2012; Greene, 2014). As people tend to consider themselves as morally virtuous (Tappin and McKay, 2016), morally good people can be said to belong to a highly relevant ingroup, associated with strong positive feelings, and morally bad people can be said to belong to a highly relevant outgroup, associated with strong negative feelings. Taking this morality-based grouping as our starting point, we turned to a *minimal* group manipulation (e.g., Tajfel et al., 1971;

Diehl, 1990) to define in- and outgroups that are associated with *attenuated* emotional evaluations. In a minimal group paradigm, participants are divided into two or more groups on the basis of arbitrary characteristics, such as a coin flip, shirt color, or fake personality test score. Such classifications, although arbitrary, lead to subtle in- and outgroup biases with a preference for “us” and a dispreference for “them”, in face-to-face contact, but also when processing language (e.g., Morrison et al., 2012). We reasoned that with phrases like “Mark is angry”, the force of group-based emotional evaluation (e.g., a bit of Schadenfreude when something bad happens to an outgroup member) would be

weaker when the characters at hand belonged to *minimal* outgroup than when they belonged to a moral outgroup. With a smaller contribution of group-dependent evaluation, and on the assumption that at phrases like “Mark is angry”, language-driven lexical or situation-model simulation would remain the same, a simulation-based valence effect should begin to show up in corrugator EMG.

Before the critical task, participants filled out a fake personality questionnaire, which invariably scored them as a “type P” rather than a “type O” personality. Each participant subsequently viewed a series of composite stimuli as their EMG was being recorded. At each trial (see **Figure 2**), participants first saw a silhouette of a character together with a moral (“good” or “bad”) or a minimal (“type P” or “type O”) group classification, and then read a sentence in which the character was described having a positive or negative emotion because of some particular reason.

Each sentence contained three EMG-relevant segments. At the *character manipulation segment*, we predicted that designating a character as bad should elicit more corrugator activity than designating a character as good. Based on evidence for a mild negative bias toward minimal outgroups (e.g., Diehl, 1990), designating a character as type O to participants who themselves have been designated as type P might also elicit more corrugator activity, albeit to a lesser extent than with a *moral* outgroup designator.

At the *affective state adjective segment* (e.g., “angry” vs. “happy”), the critical segment for our study, predictions also depended on whether characters had been designated in terms of a moral or a minimal group dimension. For characters designated as morally good or bad, we expected to replicate the crucial corrugator EMG pattern observed in our two earlier EMG-studies: substantially more frowning at “(Mark was) angry” than at “(Mark was) happy”—i.e., a large adjective valence effect—for morally good characters because of simulation- and evaluation-driven activity adding up, but a zero, or close to zero, adjective valence effect for morally bad characters because of simulation- and evaluation-driven activity counteracting each other.

For characters designated as belonging to a *minimal ingroup* (type P, the same type as the participant), we again expected more frowning at “(Mark was) angry” than at “(Mark was) happy” because of simulation- and evaluation-driven activity adding up. However, because the fate of a member of an ingroup that the reader only weakly associates with should matter less than the fate of a member of an ingroup member that the reader strongly associates with, the size of the adjective valence effect with minimal ingroup characters should be smaller than with moral ingroup characters. Furthermore, for characters designated as belonging to a *minimal outgroup* (type O), we predicted that the negative emotion associated with simulating the meaning of “(Mark was) angry”, compared to “(Mark was) happy”, would not be fully counteracted by a weak outgroup-contingent positive evaluation of this particular outcome, leading to a small net adjective valence effect. Assuming *some* minimal ingroup favoritism, the net adjective valence effect should still be a bit larger with minimal ingroup characters (where any evaluation still aligns with simulation) than with minimal outgroup

characters (where it opposes simulation). But in *both* minimal group cases, adjective valence effects should lie between the adjective valence effects in the two moral group cases. With a smaller evaluation bias, EMG-responses should in the minimal-group part of the design be dominated more by language-driven simulation.

At the *affect reason segment*, the reason for the character’s emotion is revealed. Because the input provided here is distributed over a multi-word clause, with the reasons for positive and negative affect usually differing on more than one word, descriptions of affect reasons were much less well-controlled in terms of lexical variables and time-locking precision. We therefore made no detailed predictions for this segment. However, in line with the results of the one prior study where we had also temporally separated the affect reason from the affective state adjective (’t Hart et al., 2019), we expected a renewed phasic corrugator response to reasons for negative emotion, particularly for moral ingroup characters, but possibly also for other characters.

MATERIALS AND METHODS

Participants

We recruited 64 native speakers of Dutch (58 female and 6 male) aged between 18 and 27 ($M = 21.5$, $SD = 2.2$) from the Utrecht University Humanities faculty participant database, for an experiment on reading that focused on language, emotions and personality. None of the participants had been diagnosed with dyslexia, had taken Botox® injections in the face, or had participated in the earlier ’t Hart et al. (2018), ’t Hart et al. (2019) studies. Research procedures complied with Netherlands Code of Conduct for Academic Practice and with the Declaration of Helsinki. Participants gave written informed consent after reviewing a form that detailed the nature of the materials and the procedure, and emphasized their right to withdraw consent at any time without having to provide a reason and without losing financial compensation (€ 12,-). The study was approved by the Linguistics Chamber of the Faculty of Humanities Ethics Assessment Committee at Utrecht University.

Stimulus Materials and Design

We extracted the 64 critical story-final sentences from the larger Dutch stories of the ’t Hart et al. (2019) study, to now present them in a different context. All sentences described an affective episode according to the structure: <Character name> is/feels/becomes/notes <positive or negative affective state adjective> <neutral connector phrase> <reason for the affective state>. Positive and negative critical adjectives had comparable average length (positive: 8.0 letters, range 4–13; negative: 7.3 letters, range 4–14), and so did the reason fragments (positive: 40.3 characters, range 22–52; negative: 42.5 characters, range 21–58, including spaces). Each of the critical sentences was preceded by a neutral silhouette image of a male or female character together with a moral or a minimal group classification, signaled by a badge on their chest that said

either “good” (moral ingroup), “bad” (moral outgroup), “type P” (minimal ingroup) or “type O” (minimal outgroup), as well as by an accompanying qualification underneath the silhouette: “<Character name> is a really good person”, “...a really bad person”, “...a type P personality” or “...a type O personality” (see **Supplementary Section S1**).

Fully crossing character manipulation with critical sentence type yielded eight stimulus variants that realized our $2 \times 2 \times 2$ design: grouping dimension (moral vs. minimal) \times group (ingroup vs. outgroup) \times critical adjective valence (positive vs. negative). We constructed 8 pseudo-randomized 128-trials lists, such that (a) no specific stimulus variant was repeated in a list, (b) each list contained two pseudo-randomized blocks, with 64 moral-group items followed by 64 minimal-group items in four lists, and the reverse block order in the remaining lists, and (c) in each block, 32 items had a male character and 32 a female one. Each participant received one list only.

Procedure and Data Acquisition

After signing an informed consent form, participants first completed a (fake) digital personality test. The 22 items in this test, pseudo-randomly drawn from existing personality tests, queried aspects of personality unrelated to morality (e.g., “Sometimes I really lose myself in music”, “I have fairly fixed habits”, “I never worry”, and “I am very eager to learn”). Unbeknownst to the participants, the test automatically always classified them as “type P”. To make sure participants attended to their classification, they were asked to digitally enter their type themselves, and to wear a badge with a capital P for the remainder of the session.

In the subsequent EMG-task, participants read a series of descriptions of events involving different characters, each preceded by a character description. Apart from trying not to move and blink too much, no other task was imposed. Stimuli were presented with the structure and timing shown in **Figure 2** on a 15.6-inch laptop monitor (Lenovo E531 ThinkPad) positioned at about 60 cm distance, in white on a gray background, with a character silhouette image of approximately 10° vertical angle, a 26 points Times New Roman font for the sentence, and with the same neutral baseline picture of a forest scene presented at the beginning of each trial (providing a mental reset and a trial-specific EMG-baseline). Participants pressed the space bar to advance to the next trial, with their left hand so as to prevent cable movement artifacts. Each block was preceded by two practice trials, and the blocks were separated by a pause that contained a short and easy distractor task. Sentence presentation parameters were identical to that of 't Hart et al. (2019).

Facial EMG was recorded at 2048 Hz with a Nexus MKII biosignal system (Mind Media, Roermond-Herten), using reusable Ag/AgCl electrodes with a 2 mm contact surface, placed at standard recording sites over the right corrugator supercilii and zygomaticus major (Fridlund and Cacioppo, 1986; van Boxtel, in press). As in the 't Hart et al. (2018), 't Hart et al. (2019) studies, we recorded from the right side of the face only (on average, spontaneous facial expressions do not differ between the left and right side of the face; Ekman et al.,

1981). Also as in our earlier studies, we defined predictions for *corrugator* EMG only. Although the corrugator and the zygomaticus are often used together to assess emotional valence, only the former muscle tracks valence in a relatively monotonic way (zygomaticus activity can increase with both positive and very negative stimuli, relative to neutral stimuli; Kunkel, 2018, Chapter 3; Larsen et al., 2003; Lee and Potter, 2018; see van Berkum et al., in press, for review). To allow for comparison to other work, we document average zygomaticus results in the **Supplementary Section S5**, with the raw data available in our online repository (<https://doi.org/10.24416/UU01-YM9VPP>).

After the EMG-task, participants filled out the Adolescent Measure of Empathy and Sympathy (AMES, Vossen et al., 2015), the Moral Foundations Questionnaire (MFQ, Graham et al., 2011), and a structured exit survey. The AMES and MFQ data were of exploratory interest and are reported in the **Supplementary Section S3**. Finally, participants were debriefed and paid. The average total session lasted about 75 min, with about 45 min on the EMG-task.

Data Preparation and Analysis

The raw EMG-data were filtered with a band-pass of 20–500 Hz (48 dB/octave roll-off) and a notch filter at 50 Hz to remove common artifacts (see van Boxtel 2010), followed by signal rectification and segmentation per trial, all using BrainVision Analyzer 2 (BrainProducts, Gilching). A trigger placement error resulted in loss of data for two of originally 64 tested participants. For the remaining 62 participants, we used visual inspection to select maximally long epochs of “quiet signal” (free of extreme bursts) within the 2,000 ms baseline segment, with a minimum length of 500 ms for each muscle. If a continuous artefact-free baseline epoch of at least 500 ms was not found, the trial was excluded from the analysis (resulting in 3.45% lost trials).

After baseline epoch selection, the data were exported to MatLab for further segmentation time-locked to the onset of the character manipulation picture (segment length 3,500 ms), the affective state adjective (segment length 1,000 ms), and the affect reason (segment length 2,500 ms). Each of the resulting EMG segments was then partitioned into consecutive 100-ms bins, known to strike a good balance between sufficient temporal resolution and sufficient random error reduction (van Boxtel, 2010). To reduce random variance both within and between individuals (van Boxtel, 2010), average EMG activity was expressed as a percentage of the pre-stimulus baseline epoch activity level.

The three segments were analyzed separately using the linear mixed models procedure in SPSS (IBM, v25). In linear mixed models the item- and participant variance are estimated simultaneously, resulting in a cross-classified model (Quené and Van den Bergh 2004; Quené and Van den Bergh 2008). For each segment, we constructed models for the corrugator data by iteratively adding potentially relevant components and testing for significant model improvement at each addition (using the likelihood ratio (-2LL difference chi-square) test, $p < 0.05$). We only kept components whose addition

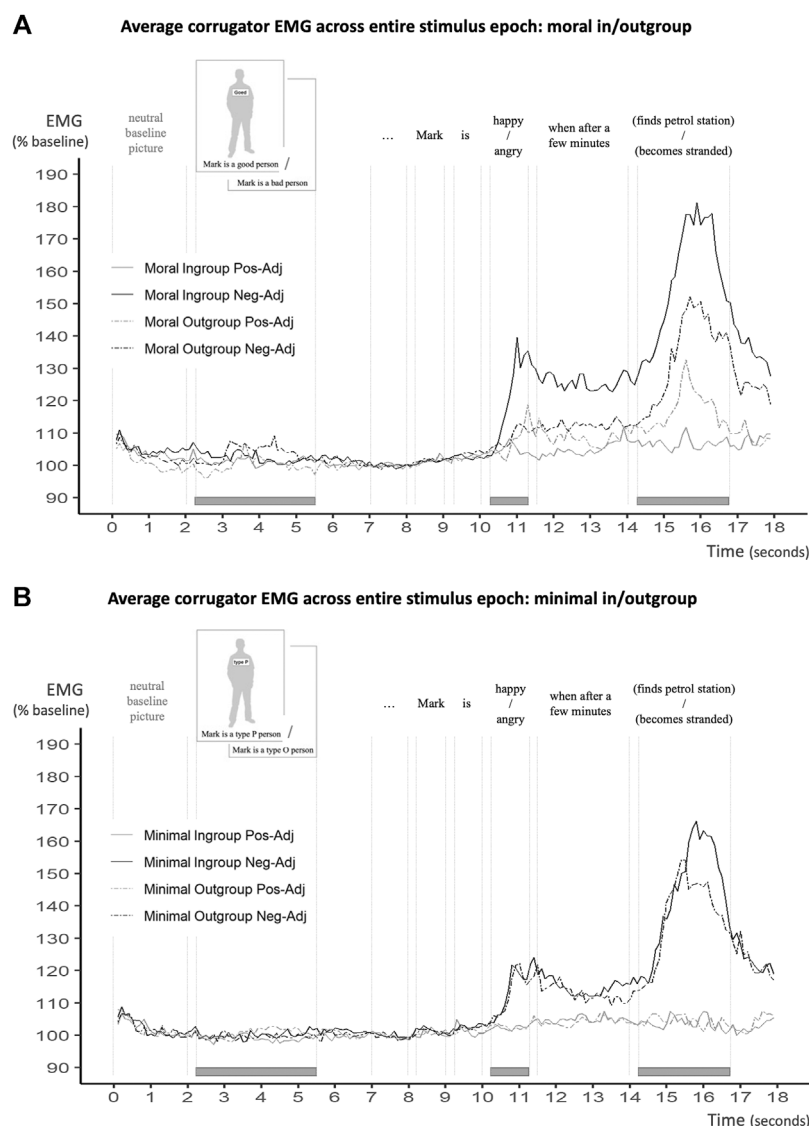


FIGURE 3 | Corrugator EMG across the entire stimulus epoch, for (A) moral and (B) minimal in- and outgroups, together with a (partly abbreviated) example item, the associated stimulus component presentation onsets and offsets (thin vertical lines) and the three critical EMG segments (gray bars). Note that on average, EMG activity in the first two seconds is above 100% because the values on which the baseline value is based may be a subset of all data points in this 2-s interval (see Methods for EMG baselining procedure).

explained a significant amount of variance or were necessary to test hypothesized interactions. Components that did not significantly improve the model were dropped in the next iteration (Winter, 2020).

Because we were not only interested in average corrugator activity in a segment but also in its development *over time*, we used a growth curve model approach (Peck and Devore, 2008; Mirman, 2015) with specific analysis designs that were optimized for assessing and comparing time trends across conditions. We first modeled participants and items as random factors. To assess the effect of our manipulations on the average activation across an entire segment, we subsequently added grouping dimension (moral vs.

minimal), group (ingroup vs. outgroup), and its interaction as a fixed factor in the model for the character manipulation segment, and grouping dimension (moral vs. minimal), group (ingroup vs. outgroup), affective state adjective/reason valence (positive vs. negative), and their 2-way and 3-way interactions as fixed factors in the model for the affective state adjective and affect reason segments. Afterward the most complex interaction was added to the random part of the model as a random slope. Next, linear, quadratic, and cubic trends were added as covariates in the fixed part of the model. Time trend (e.g., linear) components were added *per condition* to maintain flexibility in building the model, and to avoid forcing the model to fit, for example, a linear trend for all conditions

when only *one* condition contained a significant linear component. All trend components were centered to avoid correlation between trends (fixed effects final model intercepts therefore reflect the average corrugator activity across the entire segment, *not* the level at which corrugator activity intercepts the y -axis). By using trends up to the cubic component, we achieved some flexibility to fit responses without over-fitting or losing explanatory power (Mirman, 2015). Because we were particularly interested in temporal developments, the random part of the models always included random slopes for subjects for each time trend that initially improved the model (as well as standard random intercepts for subject and item).

To facilitate interpretation, in the final model the fixed factors grouping dimension, group and affective state adjective/reason valence were included as a single condition factor, which allowed for a no-intercept model where the estimates of the conditions reflect the segment average corrugator activation. This re-parametrization does not change the -2LL value, and as such still represents the optimal model. While trend components were fitted with a resolution of 100 ms, the associated parameter estimates (e.g., b for a linear slope) are reported on a 1-s basis. For the final model, custom two-tailed t -tests were used to assess theoretically relevant pairwise comparisons between condition averages. Theoretically relevant comparisons between two (e.g., linear) condition-specific trend components were done by explicitly comparing the difference between associated regression weights ($b_1 - b_2$) in a dedicated two-tailed t -test in case both components had been kept in the model, and by resorting to the simple fixed effects t -test for just one of them (e.g., b_1) when the other component had not been included in the final model (which effectively defined b_2 as 0). For each critical segment, **Supplementary Section S2** reports on model construction steps, followed by parameter tests and specific comparisons based on the final model (referring to our online repository for all original statistical analyses documents: <https://doi.org/10.24416/UU01-YM9VPP>).

RESULTS

Figure 3 shows average corrugator EMG responses across the entire stimulus epoch, together with an example item and the associated temporal structure. As can be seen, there is hardly any differential activity in the character manipulation segment, but substantial differential activity in the affective state adjective segment and the affect reason segment. In the following, we discuss these results per segment (see **Supplementary Section S2** for statistical details).

Character Manipulation

The character manipulation designated a character as morally good (moral ingroup), morally bad (moral outgroup), a type P personality (minimal ingroup) or a type O personality (minimal outgroup). **Figure 4** shows the corrugator EMG results for each

condition, time-locked to the onset of the character manipulation picture. For the average corrugator activity across the entire 3.5-s segment, the overall interaction test revealed a significant interaction between grouping dimension \times group (moral ingroup 103.9%, moral outgroup 112.0%, minimal ingroup 104.3%, minimal outgroup 104.4%, $F(4, 134.64) = 1542.44$, $p < 0.001$). We discuss all further effects for moral and minimal groups separately.

Moral In- and Outgroup

As expected, **Figure 4** reveals increased frowning to characters designated as bad (dashed black line), and no such increase for characters designated as good (solid black line). In line with this, average corrugator activity across the entire segment differed significantly at morally good vs. morally bad character descriptions (difference_{ingr-outgr} = -8.02 , $t(184.17) = -2.48$, $p = 0.01$, 95% CI $[-14.41, -1.64]$). As for the trend components, the model fitted a flat line for good character descriptions, indicating no change in corrugator activity at all. For bad character descriptions, two marginal effects hint at the phasic nature of the response, with linear and quadratic trend components further improving the statistical model (see **Supplementary Section S2a**), and the associated b -estimates almost significantly different from zero (positive linear trend $b = 2.19$, $t(62.02) = 1.97$, $p = 0.053$, 95% CI $[-0.03, 4.41]$; negative quadratic trend $b = -2.89$, $t(61.97) = -1.93$, $p = 0.059$, 95% CI $[-5.89, 0.11]$). In all, seeing a silhouette with “bad” accompanied by “X is a really bad person” fairly rapidly elicits a bit of frowning, starting at around 1,000–1,100 msec in the actual (non-modeled) data.

Minimal In- and Outgroup

We had considered that designating a character as a *minimal* outgroup member might increase corrugator activity too, although not to the extent observed for *moral* outgroup designators. However, in **Figure 4**, the corrugator response to characters labeled as type O (dashed gray line) and type P

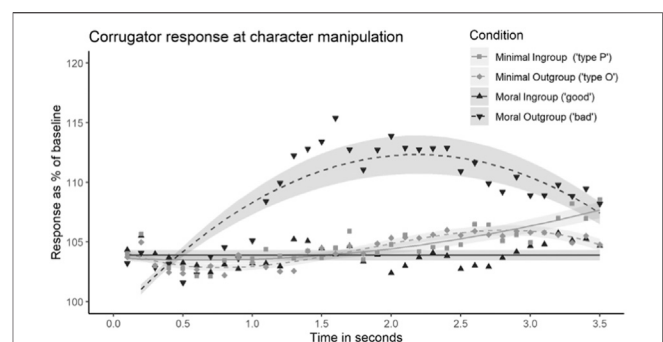


FIGURE 4 | Corrugator EMG response to the character manipulation. Dots show the observed data (averaged per 100 ms), and lines show the final growth curve model (incorporating all intercept and trend parameters that significantly improved the model). The gray bands represent 95% confidence intervals from the final growth curve model.

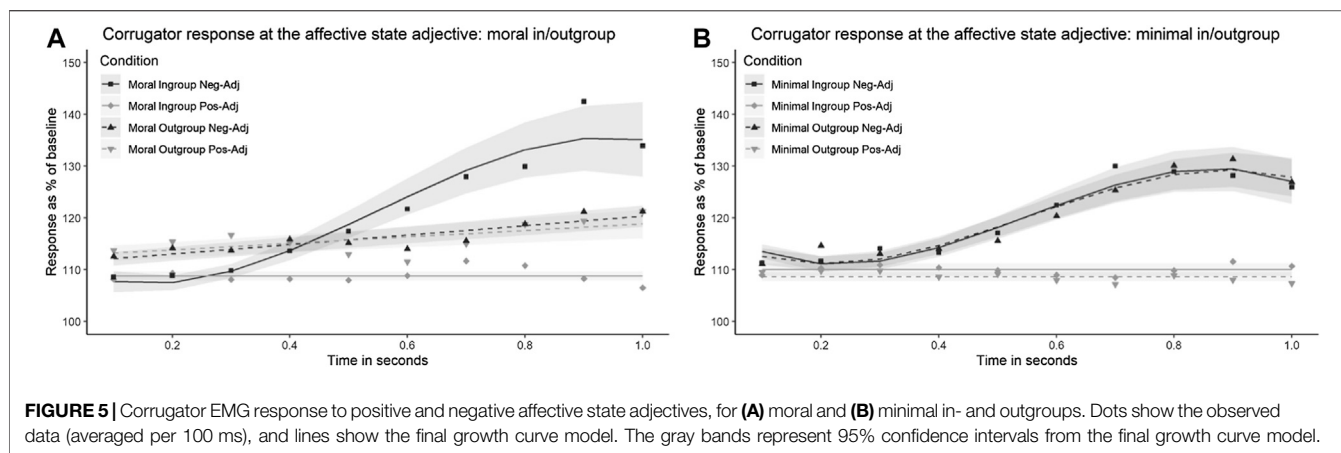


FIGURE 5 | Corrugator EMG response to positive and negative affective state adjectives, for (A) moral and (B) minimal in- and outgroups. Dots show the observed data (averaged per 100 ms), and lines show the final growth curve model. The gray bands represent 95% confidence intervals from the final growth curve model.

(solid gray line) are right on top of each other, and a pairwise comparison did not reveal a significant average corrugator activity difference between the two conditions (difference_{ingr-outgr} = -0.18, $t(184.15) = -0.05$, $p = 0.96$, 95% CI [-6.56, 6.21]). Also, trend analysis did not reveal any big differences there either. For minimal outgroup as well as minimal ingroup character descriptions, a positive linear trend component improved the model (see **Supplementary Section S2a**). For minimal outgroup characters, the positive linear trend significantly differed from zero ($b = 2.27$, $t(302.92) = 3.63$, $p < 0.001$, 95% CI [1.04, 3.50]), while for minimal ingroup characters the difference was only marginal ($b = 1.18$, $t(61.90) = 1.89$, $p = 0.06$, 95% CI [-0.07, 2.44]), but the pairwise comparison revealed no significant difference between these two trends (difference_{ingr-outgr} $b = 1.09$, $t(204.66) = 1.23$, $p = 0.22$, 95% CI [-0.66, 2.83]). Higher-order trend analysis revealed a significant but very small cubic trend in the minimal outgroup response (indicating a fall-rise-fall pattern, see **Supplementary Section S2b**), but as can be seen in **Figure 4**, the fitted curves for these two conditions are virtually on top of each other. In all, the corrugator EMG did not show a clear differential response to descriptions of minimal ingroup (type P) or outgroup (type O) characters. Participants did report feeling less similar (range: -3 *not similar at all*, 3 *very similar*) to minimal outgroup characters than to minimal ingroup characters ($M = -1.61$ vs. $M = 0.70$; $M_{diff} = -2.31$, $SD = 1.19$; two-tailed paired-samples t -test $t(61) = -9.06$, $p < 0.001$), but this did not translate to clearly differential EMG activity.

Affective State Adjective

At the affective state adjective (e.g., “happy/angry”), the most critical segment in our study, participants read about positive or negative emotion of the same character. This additional adjective valence factor expands the EMG-analysis to a 2 (grouping dimension: morality vs. minimal group) \times 2 (group: ingroup vs. outgroup) \times 2 (affective adjective valence: positive vs. negative) design. **Figure 5** displays the associated corrugator EMG-responses. Consistent with the first impression, analysis of the average EMG activity across the entire 1-s segment revealed a significant three-way-interaction of these factors ($F(8, 259.41) = 222.51$, $p < 0.001$).

As with the character manipulation, we discuss all further effects for moral and minimal groups separately.

Moral In- and Outgroup

For characters designated as morally good or bad, we expected to replicate the core result of our two earlier EMG-studies: substantially more frowning at “(Mark was) angry” than at “(Mark was) happy” for morally good characters because of simulation- and evaluation-driven corrugator activity adding up, but a zero, or close to zero, adjective valence effect for morally bad characters because of simulation- and evaluation-driven corrugator activity canceling each other out. As can be seen in **Figure 5A**, this is exactly what we observed.

For morally good characters, the EMG response showed a clear and rapid increase in frowning activity at negative state adjectives (solid black line), but no such increase at positive state adjectives (solid gray line), with the signals diverging from about 300–400 ms onwards. Statistical analysis of average EMG during the entire 1-s segment confirmed that participants frowned significantly more when a good character had a negative emotion than when he or she had a positive emotion (difference_{neg-pos} = 13.70, $t(428.27) = 3.13$, $p = 0.002$, 95% CI [5.10, 22.30]). Trends in the response also differed. The model fitted a flat line at positive affective state adjectives but included a significant linear increase in activation at negative adjectives ($b = 56.40$, $t(76.12) = 2.65$, $p = 0.01$, 95% CI [14.03, 98.78]), as well as a cubic trend (see **Supplementary Section S2d**).

For morally bad characters, the EMG response showed no such differential increase in frowning activity at negative state adjectives (dashed black line), relative to positive state adjectives (dashed gray line). The average segment EMG analysis confirmed that average frowning during this 1-s interval did not statistically differ at negative vs. positive state adjectives (difference_{neg-pos} = 0.72, $t(428.64) = 0.17$, $p = 0.87$, 95% CI [-7.88, 9.32]). As can be seen in **Figure 5A**, modest upward linear trend components improved the overall model, with a marginally significant b in the case of negative adjectives ($b = 9.67$, $t(56.35) = 1.92$, $p = 0.06$, 95% CI [-0.42, 19.75]), but not for positive adjectives ($b = 6.57$, $t(62.27) = 1.06$, $p = 0.30$, 95% CI [-5.86, 19.00]), and no significant difference between the two linear trends (difference_{neg-pos} $b = 3.10$, $t(115.74) = 0.39$, $p = 0.70$, 95% CI [-12.74, 18.95]).

Minimal In- and Outgroup

As can be seen in **Figure 5B**, we did not obtain the expected pattern of results in this part of the design. For minimal ingroup characters (i.e., designated before as a type P person), the EMG response showed a clear and rapid increase in frowning activity at negative state adjectives (solid black line) starting around 300–400 msec, and no such increase at positive state adjectives (solid gray line). However, for minimal outgroup characters (i.e., designated before as a type O person), the exact same result was observed, with a clear and rapid increase in frowning activity at negative state adjectives (dashed black line) starting at around 300–400 msec, and no such increase at positive state adjectives (dashed gray line). This suggests that minimal group membership did not modulate the net differences between responses to adjectives like “angry” and “happy”.

The statistical analysis of average EMG during the entire 1-s segment confirmed that participants frowned significantly more when a character had a negative emotion than when he or she had a positive emotion, for minimal ingroup characters (difference_{neg-pos} $b = 10.29$, $t(427.92) = 2.35$, $p = 0.02$, 95% CI [1.69, 18.89]), as well as for minimal outgroup characters (difference_{neg-pos} $b = 11.76$, $t(428.76) = 2.69$, $p = 0.01$, 95% CI [3.16, 20.36]). Furthermore, there was no difference in average frowning activity at negative state descriptions of minimal ingroup vs. outgroup members (**Figure 5B**, black lines; difference_{ingrp-outgrp} $b = 0.09$, $t(428.82) = 0.02$, $p = 0.98$, 95% CI [-8.51, 8.69]), nor at positive state descriptions of minimal ingroup vs. outgroup members (**Figure 5B**, gray lines; difference_{ingrp-outgrp} $b = 1.56$, $t(428.02) = 0.36$, $p = 0.72$, 95% CI [-7.03, 10.16]).

As clearly evident in **Figure 5B**, the temporal development of the corrugator EMG signal at positive and negative state adjectives also did not vary as a result of minimal group membership. Negative affective state adjectives led to a significant linear increase in corrugator activity both for minimal ingroup members ($b = 42.85$, $t(55.50) = 2.29$, $p = 0.03$, 95% CI [5.34, 80.36]) and for minimal outgroup members ($b = 39.76$, $t(144.53) = 3.47$, $p = 0.001$, 95% CI [17.12, 62.40]), with no significant difference between the two (difference_{ingrp-outgrp} $b = 3.09$, $t(99.49) = 0.14$, $p = 0.89$, 95% CI [-40.45, 46.64]). The model also included a negative cubic trend at negative affective state adjectives for minimal ingroup and outgroup members, but the patterns did not differ (see **Supplementary Section S2d**).

Our multiple-drivers model, and our additional assumption of weaker (but non-zero) group-dependent evaluation in the minimal group case than in the moral group case, had led us to expect that the differential adjective valence effect (e.g. “angry” vs. “happy”) would be smaller with minimal (type P) ingroup characters than with moral (good) ingroup characters, with the corrugator signal to a minimal ingroup character experiencing negative emotion to end up below that to a moral ingroup character experiencing the same emotion (i.e., black solid line in **Figure 5B** lower than black solid line in **Figure 5A**). However, although descriptively the EMG-response pattern is in the right direction, pairwise comparisons showed no significant

difference between these two signals, neither in terms of the 1-s segment average, nor in terms of the linear or cubic trend component (all p 's > 0.63). Also, we had expected the corrugator signal to a minimal ingroup character experiencing positive emotion to end up above that to a moral ingroup character experiencing the same emotion (i.e., gray solid line in **Figure 5B** higher than gray solid line in **Figure 5A**). However, both moral- and minimal ingroup-positive were fitted with a flat line that did not significantly differ in elevation ($p = 0.79$). For a full report of all estimates and comparison see **Supplementary Section S2d**.

All in all, in the morality part of the design, we replicate the core results of our earlier work: corrugator responses to negative and positive emotion adjectives strongly depend on who is experiencing the emotion described. In the minimal-group part, however, the identity of the character does not matter at all, with equally large adjective valence effects for minimal ingroup and minimal outgroup characters.

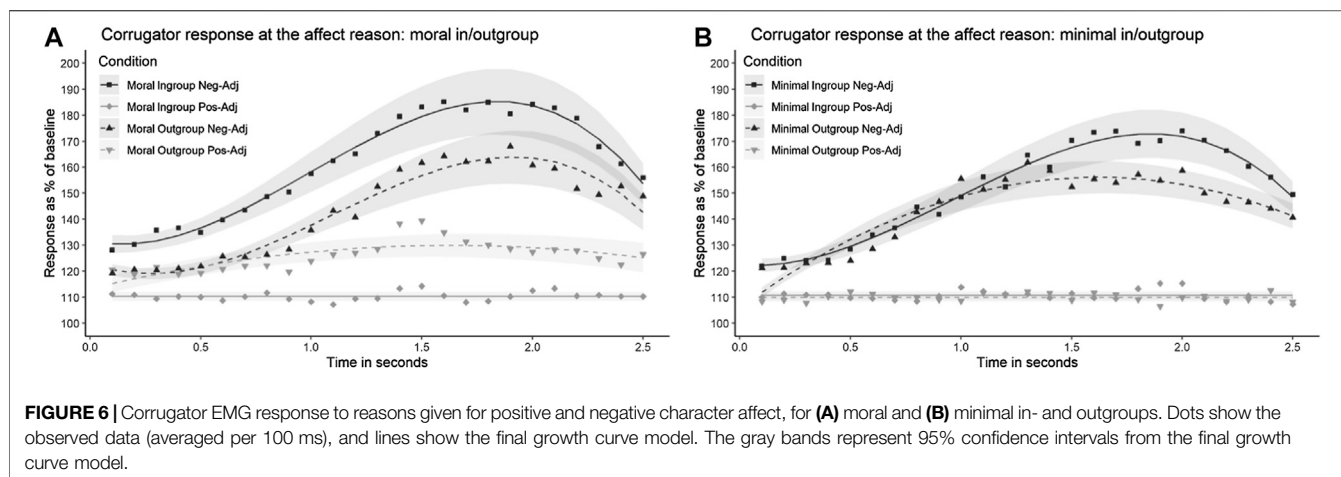
Affect Reason

At the affective reason segment, participants read about events that provided a reason for the character's emotion. The analysis at this segment involves a 2 (grouping dimension: morality vs. minimal group) \times 2 (group: ingroup vs. outgroup) \times 2 (affect reason valence: positive vs. negative) design. **Figure 6** displays the associated corrugator EMG-responses. One striking aspect of the EMG-patterns in **Figures 6A and B** is the renewed phasic corrugator response in all four conditions motivating a character's negative emotion, which suggests that these sentence fragments contained enough information to elicit additional differential corrugator activity. Also, as evident from the entire-epoch **Figure 3**, these new phasic corrugator responses ride on top of relatively stable corrugator differences that emerged at the prior affective state adjective, and that lasted for several more seconds, throughout the intermediate neutral connector phrase (e.g., “when after a few minutes”). Because corrugator activity is expressed as a percentage of the same pre-stimulus baseline at all three critical segments, these longer-lasting state adjective effects are responsible for the pre-existing differences at 0 s in **Figure 6**.

Analysis of the average EMG activity across the entire 2.5-s affect reason segment revealed a significant three-way-interaction of grouping dimension, group, and affect reason valence ($F(8, 255.28) = 41.79$, $p < 0.001$), an interaction that to some extent reflects these earlier adjective-triggered EMG effects. As before, we discuss all further effects for moral and minimal groups separately.

Moral In- and Outgroup

In line with 't Hart et al. (2019), we had expected a renewed phasic corrugator response to events that were the reason for negative, as opposed to positive, emotion, particularly for good characters, but possibly also for bad characters. As can be seen in **Figure 6A**, there is indeed a clear and substantial increase for negative events befalling good characters (solid black line) and a flat-line response for positive events befalling these same characters (solid gray line). A smaller but descriptively comparable response difference emerged for bad characters, with negative events (dashed black



line) eliciting somewhat higher corrugator EMG activity than positive events (dashed gray line).

For good characters (solid lines), average corrugator activation across the segment was indeed significantly higher for negative events than for positive events (difference_{neg-pos} $b = 62.20$, $t(430.05) = 5.20$, $p < 0.001$, 95% CI [38.69, 85.71]). Furthermore, while negative events elicited a significant linear increase in corrugator activity ($b = 41.70$, $t(61.46) = 2.42$, $p = 0.02$, 95% CI [7.20, 76.20]), modulated by significant quadratic and cubic trends (both $p < 0.04$, see **Supplementary Section S2f**), positive events elicited a flat-line EMG response. For bad characters (dashed lines), however, average corrugator activation at negative and positive events was not significantly different (difference_{neg-pos} $b = 21.38$, $t(438.89) = 1.78$, $p = 0.08$, 95% CI [-2.25, 45.01]). Also, although negative events elicited an almost significant linear increase in corrugator activity ($b = 38.64$, $t(61.72) = 1.94$, $p = 0.06$, 95% CI [-1.16, 78.44]) while positive events did not ($b = 4.27$, $t(61.76) = 1.01$, $p = 0.32$, 95% CI [-4.20, 12.74]), the difference between the two linear trends was not significant (difference_{neg-pos} $b = 34.37$, $t(67.30) = 1.69$, $p = 0.10$, 95% CI [-6.25, 75.00]). Both events did elicit a significant quadratic trend ($p = 0.02$ and $p = 0.01$ for negative and positive events respectively, see **Supplementary Section S2f**).

With the two EMG-signals for good characters being much (and significantly) further apart than the two EMG-signals for bad characters, **Figure 6A** could be taken to suggest that readers are again more sensitive to the fate of good characters than to that of bad ones, just as at the adjective. However, the elevated average corrugator response to negative over positive events with moral ingroup characters is to a large extent already present at 0 s, and is as such presumably largely due to spill-over from the earlier adjective effect (see particularly **Figure 3**, and compare the EMG-pattern at segment onset in **Figure 6A** to the EMG-pattern at segment offset in **Figure 5A**). We therefore cannot confidently model this pattern of results as *renewed* differential sensitivity to the fate of good and bad characters. In all, the only informative result in this part of the design is a significant phasic rise-fall response when reading about bad events (happening to good or

bad people alike), and when reading about good events happening to bad people.

Minimal In- and Outgroup

As can be seen in **Figure 6B**, the dominant pattern of results is that of large phasic corrugator responses to negative events befalling both minimal ingroup (“type P”) and outgroup (“type O”) characters, and no responses to positive events. Statistical analysis confirms this. For minimal ingroup characters, average corrugator activation across the segment was higher for negative events than for positive events (difference_{neg-pos} $b = 52.06$, $t(429.90) = 4.35$, $p < 0.001$, 95% CI [28.55, 75.56]). Furthermore, while negative events happening to minimal ingroup characters elicited a significant linear increase in corrugator activity ($b = 40.57$, $t(59.99) = 2.26$, $p = 0.03$, 95% CI [4.67, 76.46]), which was modulated by a significant quadratic and marginally significant cubic trend ($p = 0.03$ and 0.07 , respectively, see **Supplementary Section S2f**), the corrugator response to positive events was modeled as a flat line. For minimal outgroup characters, average corrugator activation across the segment was also higher for negative events than for positive events (difference_{neg-pos} $b = 43.94$, $t(430.28) = 3.67$, $p < 0.001$, 95% CI [20.43, 67.45]). Furthermore, while negative events again elicited a significant linear increase in corrugator activity ($b = 12.38$, $t(61.92) = 2.71$, $p = 0.01$, 95% CI [3.25, 21.52]), which was modulated by a significant quadratic trend ($p = 0.02$, see **Supplementary Section S2f**), the corrugator response to positive events was again modeled as a flat line.

We had speculated that minimal group status might also have an impact on how negative vs. positive events affected the corrugator response. Although **Figure 6B** suggests a somewhat stronger EMG-response to *negative* events befalling minimal ingroup characters than befalling minimal outgroup characters, the statistics do not clearly support this: average corrugator activation over the entire 2.5-s segment did not differ (difference_{ingrp-outgrp} $b = 9.10$, $t(439.04) = 0.76$, $p = 0.45$, 95% CI [-14.53, 32.73]), nor did any of the trends (e.g., linear trend difference_{ingrp-outgrp} $b = 28.18$, $t(67.75) = 1.52$, $p =$

0.13, 95% CI [-8.77, 65.14]). EMG-responses to positive events befalling minimal ingroup vs. minimal outgroup characters were fitted with a flat line whose elevation did not differ either (difference_{ingroup-outgroup} $b = 0.99$, $t(421.35) = 0.08$, $p = 0.93$, 95% CI [-22.40, 24.38]).

DISCUSSION

When processing language, do we *simultaneously* use our emotion systems to simulate somebody else's described emotion *and* to evaluate, i.e., have our own emotions about, what is described? We explored the viability of a multiple-drivers model for language-driven emotion ('t Hart et al., 2018; 't Hart et al., 2019; van Berkum et al., in press) by "downtuning" the force of character-dependent emotional evaluation via a minimal-groups paradigm, such that corrugator EMG responses would reveal character-independent emotion simulation to a larger extent. Also, we aimed to replicate the findings of 't Hart et al. (2018), 't Hart et al. (2019), generalizing those earlier morality-based observations to a situation where characters were simply *declared* as good or bad, rather than shown to be so earlier in a story. As for morality, we indeed replicated the core result of our earlier studies: substantially more frowning to negative emotion adjectives than to positive ones when the character having the emotion was seen as morally good, but *not* when he or she was seen as morally bad. However, and in contrast to our expectations, defining characters as belonging to a *minimal* (rather than a moral) in- or outgroup did not matter to how much more readers frowned to negative as opposed to positive emotion adjectives. We first discuss the EMG-results per segment, and then turn to a more general discussion.

Processing Character Descriptions

In our study, introducing some unknown fictional character as a member of a minimal in- or outgroup did not elicit any differential frowning. As for *morally defined* groups, however, things were different: declaring some unknown fictional character as "really bad" led to a small but significant phasic increase in frowning, whereas declaring a character as "really good" did not affect the corrugator. It is perhaps tempting to relate this to differences at the level of situation modeling (see **Figure 1**), i.e., of imagining a concrete bad character in some real or imaginary context (with a silhouette providing extra input). However, because isolated negative words are known to elicit more frowning than positive words (e.g., Larsen et al., 2003; Kunkel, 2018; see van Berkum et al., in press, for review), this effect may very well also—or exclusively—hinge on automatic responses associated with the retrieval of negative vs. positive words ("bad" vs. "good"). Either way, it is interesting to compare the very modest current effect to the very large corrugator responses to descriptions morally bad and good character *behavior* in our earlier two studies. In 't Hart et al. (2018), 't Hart et al. (2019), phasic corrugator increases were some 50–90% higher at peak relative to baseline, when participants read about a main character committing a concrete moral transgression (e.g., deliberately speeding up to soak a pedestrian in the rain) than

when reading about that character displaying morally good behavior (e.g., deliberately slowing down to *not* soak the pedestrian). In the current study, however, seeing a silhouette simply described as really bad generated a phasic corrugator increase which was only some 10% higher at peak relative to baseline, as compared to a silhouette described as really good. Although adequately controlled within-experiment comparisons are required to explore the matter further, this comparative observation could be taken to suggest that describing a concrete bad action in some detail is frowned upon to a much larger extent than simply defining somebody as a bad person, an interpretation that is in line with the idea that our brains evolved to deal with concrete events and actions, and are as such much more sensitive to narrative than to non-narrative descriptions (e.g., Boyd, 2009; Boyd, 2018).

Processing Character Affect

Our predictions for the impact of character morality on reading a subsequent adjective that described *an emotion of that character* were confirmed. With good characters, readers frowned more at negative affective state adjectives like "angry" than at positive affective state adjectives like "happy", with the difference emerging very rapidly, within only a few hundred milliseconds after adjective presentation. With bad characters, our earlier work had led us to predict that this differential valence effect would be reduced to (close to) zero, which was indeed what we observed in the current study too. Taken together, these EMG-results constitute a direct replication of the 't Hart et al. (2018), 't Hart et al. (2019) findings. Like the original findings, the new findings are compatible with a multiple-drivers account in which the valenced emotional responses associated with language-driven simulation and evaluation align for good characters, but counteract each other in the case of bad characters.

Our current morality-based EMG-findings also extend the morality-based 't Hart et al. (2018), 't Hart et al. (2019) results from a paradigm where characters were described as actually *doing* something good or bad to a paradigm where characters are simply described as *being* good or bad. Note that the size of the EMG-effect at the critical state adjective (a difference at peak of about 30% relative to baseline) is comparable to the corresponding effect at the critical affective state adjective observed by 't Hart et al. (2019); a difference at peak of about 20% relative to baseline. Thus, although declaring rather than showing somebody as bad strongly attenuates the differential EMG-response of readers at the character segment, the *downstream impact of this* on how readers respond to various character emotions is not attenuated by that factor at all.

Furthermore, our findings are in line with other EMG-evidence that the social identity of characters can affect later language-driven processing. In an EMG-study on social unexpectedness, for example, descriptions of moral transgressions generated a larger corrugator response if they were committed by characters previously described in a positive, rather than a negative, way (Bartholow et al., 2001). Also, in an EMG-study involving Italian in- or outgroup politicians (e.g., Berlusconi; Fino et al., 2019) the corrugator responded strongly to negative vs. positive emotional

expression descriptions (e.g., “Berlusconi frowns” vs. “Berlusconi smiles”) if the politician belonged to the participant’s political ingroup, but not if he or she belonged to the participant’s political outgroup. The overall pattern of results in the latter study is actually strikingly similar to the pattern in our current and two earlier studies, with average corrugator EMG-responses to outgroup politicians that are not only indifferent to the characters’ emotional state, but that are also positioned between the very different corrugator signals to negative vs. positive emotions of ingroup politicians. This makes sense: political and moral orientations are strongly related (e.g., see van Berkum et al., 2009; Haidt, 2012), and both are associated with strong in- and outgroups. Still, the stability of this crucial finding across labs and materials is reassuring.

In the minimal-group part of the design, the EMG results here were predicted to be an attenuated version of those in the moral-group part of the design, with an intermediately sized adjective valence effect for both in- and outgroup characters, and *some* group-dependent modulation of this effect. However, although the adjective valence effects for minimal in- and outgroups were indeed of an intermediate magnitude, they also were exactly the same. Under a multiple-drivers account, this suggests that when reading, say, “Mark is angry”, readers not only *simulate* negative emotion at the lexical and/or situation-model level similarly for minimal in- and outgroup characters, but also *evaluate* the unhappy event in the same way. This evaluation may or may not be neutral. Importantly, however, it does not differ as a result of whether a type P or type O person is being angry.

A major goal of the current study was to look for new traces of the “power struggle” between language-driven simulation and evaluation, beyond what is visible when working with moral materials. We tried to do so by reducing the impact of *character-dependent* evaluation while keeping the force of lexical and situation-model simulation intact. But this part of the endeavor did not succeed. The reason may well be that the current minimal-groups manipulation is too subtle, and that when applied to fictional people, the resulting group bias is simply too weak to generate any detectable character-dependent evaluation at the critical emotional state adjective. We return to the implication of this after discussing our findings at the third segment.

Processing Reasons for Character Affect

Although the experimental logic hinged on the EMG results at critical adjectives describing the character’s positive or negative emotion, EMG responses to the later verbal *explanation* for that emotion also provided some information. First of all, the explanations for negative character emotion elicited renewed rise-fall phasic corrugator responses in all four character conditions (of at least an additional 30% relative to the signal at 0 msec), whereas the explanations for positive character emotion elicited zero responses in three out of four cases, and only a small (~10%) phasic increase when the positive emotion involves a bad character. Example explanations for negative character emotion involve such phrases as “(because) her

shares turned out to be worthless”, “(because) he stared at her and ignored her”, “(because) somebody pushed her aside to get in more quickly” and “(because the waitress) responds in a grumpy way and looks angrily at her”. The phasic corrugator effects that these reasons for negative character emotion elicit in the reader can thus be explained in many ways, including frowning on moral transgressions, imagining unpleasant states of affairs, or simulating the negative emotions of secondary characters. It is also conceivable that reading about a reason for negative emotion can briefly boost the situation-model simulation of the main character being in that negative state—after all, knowing that somebody’s anger has a reason that fully justifies it, and that you can identify with, may well deepen one’s mental representation of that anger. Because the affect reason segments were not controlled to allow us to discriminate between these various options, these are all issues for future research.

A second and theoretically more interesting finding is that, at the onset of this affective reason segment, the corrugator activation levels by and large echo those at the end of the affective state adjective segment (compare **Figures 5** and **6**). As can be seen in **Figure 3**, the reason is that the corrugator response to descriptions of character emotion are to a large extent maintained throughout the intervening 3 s, during which people read neutral connector phrases such as “...when after a few minutes...” or “...when he arrives at the station and...”. In the case of moral in- and outgroup characters, this sustained corrugator behavior replicates what we observed at neutral connector phrases in the 't Hart et al. (2019) study. As discussed in our earlier paper, this could be taken to indicate that the emotion simulation induced by phrases like “Mark is angry” is more likely to occur at the level of the situation model (where the character is modeled as angry) than at the—presumably more short-lived—level of simulation as part of retrieving the meaning of the word “angry” from memory. Of course, under the current multiple-drivers account for our morality-based EMG-results at the state adjective, sustained simulation would need to be matched with equally sustained group-dependent evaluation. Also note that the degree of stability over these three intervening seconds is not perfect, which could be taken to indicate dynamic fluctuations in simulation, evaluation, or both. Still, we find it striking that the reader’s emotional state, as indexed by the corrugator, remains relatively stable for several seconds after the critical adjective, not just with moral in- and outgroup characters, but also with minimal ones.

Counteracting Simulation and Evaluation Drivers, or Something Else?

What are the implications for our theoretical model? The first question we should ask is whether the absence of a minimal group effect at the affective state adjective falsifies the multiple-drivers model. We don’t think it does. If emotion simulation and emotional evaluation both drive corrugator EMG, but evaluation is the same for both minimal groups (e.g., people care as much, or as little, about a type-P character’s feelings than about a type-O character’s feelings), no modulation of the adjective valence effect is

to be expected. So, rather than rejecting the multiple-drivers model on these grounds, a more sensible strategy at this point is to look for other techniques that may selectively down- or up-regulate the force of one of the presumed drivers (e.g., using story materials in which “bad” characters commit severe, moderate or mild moral transgressions). Also, our study does replicate the original findings that led us to adopt the multiple-drivers model in the first place, extending the relevant phenomenon to situations where characters are simply declared—rather than shown—to be good or bad. The lack of increased frowning to negative state adjectives like “angry” over positive state adjectives like “happy”, for morally bad characters, can therefore be explained by the same account that we provided for those earlier findings, a tie between lexical and/or situation model simulation pushing corrugator activity up and fairness-based evaluation pushing it down.

As we already pointed out in our earlier publications, a simple account that involves lexical or situation-model *simulation only* cannot explain why the corrugator faithfully tracks the valence of emotion adjectives when the sentence is about a good character, but *not* when it is about a bad character. Also, it is difficult to account for our morality-based results in terms of *evaluation only*. The results in **Figure 5A** might tempt one to infer that readers care about what happens to good, but not *bad*, characters, and that this differential evaluation alone can parsimoniously explain the EMG results. However, this interpretation seems unlikely. Part of the joy of written or streamed fiction comes from caring about what happens to good as well as bad characters. Also, if we would not care about what happens to bad people, gossip would become dysfunctional, and Schadenfreude would not exist. Of course, in a boring lab, things could be different. However, Schadenfreude has also been established in laboratory studies (e.g., Leach and Spears, 2009; Feather and Nairn, 2005; Singer, et al., 2006), and has even been shown to influence corrugator activity (Cikara and Fiske, 2012). More generally, why would the lab context lead people to become indifferent to the fate of bad people, but not good people?

With *simple* simulation-only and evaluation-only accounts dismissed, the multiple-drivers account displayed in **Figure 1** remains an attractive one for our morality-based EMG results, with positive or negative emotional responses associated with language-driven simulation and evaluation aligning for good characters but counteracting each other for bad characters. The explanatory power and flexibility of this multi-factor model is of course also a vulnerability. It is therefore crucial to obtain independent evidence for our assumption that, at least in our materials, simulation and evaluation fully cancel each other out when reading about the emotions of bad people.

Furthermore, although we did not consider them before running the current study, other theoretical explanations for our results may be on the table as well. One possibility is that with immoral characters, readers are somehow less inclined to engage in embodied simulation of what is being described, so less

likely to *simulate* an angry or happy character. This *selective-simulation* idea fits with recent ideas on embodied language processing, where it is becoming clear that language-driven simulation is not an all-or-none concept but depends on all kinds of contextual factors (Willems and Casasanto, 2011; Havas and Matheson, 2013; Zwaan, 2014; Pecher and Zwaan, 2017; Pecher 2018; Winkelman et al., 2018). Identification, or liking, could be one of those factors (Hoeken and Sinkeldam, 2014). As indicated in **Figure 1** and discussed more fully elsewhere (van Berkum et al., in press), we also cannot exclude that *emotional mimicry*, in response to vividly imagined character affect, partly drives emotional facial expressions during language processing. Such mimicry might occur more for good characters than for bad ones either because emotions of the former are simulated to a stronger extent, or because mimicry itself is selective, and more likely to occur with ingroup or otherwise likable characters than with other characters (see Hess and Fischer, 2014, for a review of relevant findings, and Fino et al., 2019, for EMG-results interpreted in terms of language-driven mimicry).

The possibility of selective simulation and/or selective mimicry illustrates the fact that we are dealing with a very complex situation here. Although we currently prefer our multiple-drivers account over post-hoc accounts in terms of selective simulation and/or selective mimicry—if only because it was conceived of *before the experiment*—we acknowledge that our studies are only scratching the surface. Language can lead to emotion in many different ways, and disentangling them will remain a challenge for some time.

LIMITATIONS

We end with some limitations of the current study. First, the multiple-drivers account illustrated in **Figure 1** inevitably introduces several free parameters in our modeling of language-driven facial EMG data. Of course, only *some* of the depicted drivers may actually be at work (i.e., explain EMG variance) at any given time. Furthermore, like so many other workings of the human brain, language-driven emotion may simply be this complex. Nevertheless, the explanatory power of the current model is also a vulnerability, which will need to be addressed in future work. Second, most of our participants were female, with only 6 males in a group of 64 participants. With small empathy-related sex differences in corrugator EMG reported elsewhere van der Graaff et al. (2016), this may matter. Third, we did not use a deliberate strategy to prevent people from guessing that their facial expressions were the object of study (e.g., attach dummy electrodes elsewhere, cf. Fridlund and Cacioppo, 1986). Although corrugator activity is in part automatic (e.g., Cacioppo et al., 1992; Dimberg et al., 2000; Neumann et al., 2005; Tamietto and de Gelder, 2010; van Boxtel, in press), and although it is plausible that participants soon forgot about the electrodes (see Nordin, 1990, for evidence of rapid facial sensory habituation), it may be wise to consider such a strategy in future work. Fourth, in our

growth curve analysis, only linear, quadratic and cubic trends are fitted, and they were constrained to fit the signals in a segment of a predefined duration. Although this worked out reasonably well in our data, the segment constraint obviously imposes limitations on how the data can be modeled—our procedure would not work well, for instance, when most of the segment contained a flat line, with a huge effect in the narrow last bit of the signal only. Fifth, we assessed emotion in terms of valence only—this simplified the research logic, but it also ignores some of the richness of language-induced emotion. Finally, we made relatively simple working assumptions about how characters are perceived (e.g., as good or bad), and about how people evaluate, say, something bad happening to a bad character. We think that given our materials, those assumptions are reasonable. However, people are layered, and so is their response to other people's fate. The study of language-driven human emotion will sooner or later need to take on this additional complexity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Data publication platform of Utrecht University <https://doi.org/10.24416/UU01-YM9VPP>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Linguistics Chamber of the Faculty of

Humanities Ethics assessment Committee at Utrecht University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

B'tH, MS, and JvB designed the study, AvB provided specific EMG expertise for study design and data analysis, B'tH conducted the study, B'tH and MS analyzed the results, and JvB, MS, and B'tH wrote the paper.

FUNDING

Partly supported by NWO Vici grant #277–89–001 to JvB.

ACKNOWLEDGMENTS

We thank Ella Bosch and Eletta Damen for help with running the experiment, Huub van den Bergh for his input during the analysis, and members of the UiL OTS Language and Communication research group for feedback. Results were initially published in a doctoral dissertation ('t Hart, 2017). Correspondence: JvB j.vanberkum@uu.nl.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2021.590077/full#supplementary-material>

REFERENCES

- Anderson, M. L. (2010). Neural Reuse: A Fundamental Organizational Principle of the Brain. *Behav. Brain Sci.* 33, 245–266. doi:10.1017/s0140525x10000853
- Barsalou, L. W. (2008). Grounded Cognition. *Annu. Rev. Psychol.* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639
- Barsalou, L. W. (2009). Simulation, Situated Conceptualization, and Prediction. *Phil. Trans. R. Soc. B* 364 (1521), 1281–1289. doi:10.1098/rstb.2008.0319
- Bartholow, B. D., Fabiani, M., Gratton, G., and Bettencourt, B. A. (2001). A Psychophysiological Examination of Cognitive Processing of and Affective Responses to Social Expectancy Violations. *Psychol. Sci.* 12 (3), 197–204. doi:10.1111/1467-9280.00336
- Boyd, B. (2009). *On the Origin of Stories: Evolution, Cognition, and Fiction*. Harvard University Press. doi:10.2307/j.ctvj9xvk
- Boyd, B. (2018). The Evolution of Stories: from Mimesis to Language, from Fact to Fiction. *Wires Cogn. Sci.* 9 (1), e1444. doi:10.1002/wcs.1444
- Cacioppo, J. T., Bush, L. K., and Tassinari, L. G. (1992). Microexpressive Facial Actions as a Function of Affective Stimuli: Replication and Extension. *Pers Soc. Psychol. Bull.* 18 (5), 515–526. doi:10.1177/0146167292185001
- Cikara, M., and Fiske, S. T. (2012). Stereotypes and Schadenfreude. *Soc. Psychol. Personal. Sci.* 3 (1), 63–71. doi:10.1177/1948550611409245
- Diehl, M. (1990). The Minimal Group Paradigm: Theoretical Explanations and Empirical Findings. *Eur. Rev. Soc. Psychol.* 1 (1), 263–292. doi:10.1080/14792779108401864
- Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious Facial Reactions to Emotional Facial Expressions. *Psychol. Sci.* 11 (1), 86–89. doi:10.1111/1467-9280.00221
- Ekman, P., Hager, J. C., and Friesen, W. V. (1981). The Symmetry of Emotional and Deliberate Facial Actions. *Psychophysiol.* 18, 101–106. doi:10.1111/j.1469-8986.1981.tb02919.x
- Feather, N., and Nairn, K. (2005). Resentment, Envy, Schadenfreude, and Sympathy: Effects of Own and Other's Deserved or Undeserved Status. *Aust. J. Psychol.* 57 (2), 87–102. doi:10.1080/00049530500048672
- Fingerhut, J., and Prinz, J. J. (2018). Grounding Evaluative Concepts. *Phil. Trans. R. Soc. B* 373 (1752), 20170142. doi:10.1098/rstb.2017.0142
- Fino, E., Menegatti, M., Avenanti, A., and Rubini, M. (2016). Enjoying vs. Smiling: Facial Muscular Activation in Response to Emotional Language. *Biol. Psychol.* 118, 126–135. doi:10.1016/j.biopsycho.2016.04.069
- Fino, E., Menegatti, M., Avenanti, A., and Rubini, M. (2019). Unfolding Political Attitudes through the Face: Facial Expressions when Reading Emotion Language of Left-And Right-Wing Political Leaders. *Scientific Rep.* 9 (1), 1–10. doi:10.1038/s41598-019-51858-7
- Foroni, F., and Semin, G. R. (2013). Comprehension of Action Negation Involves Inhibitory Simulation. *Front. Hum. Neurosci.* 7, 1–7. doi:10.3389/fnhum.2013.00209
- Foroni, F., and Semin, G. R. (2009). Language that Puts You in Touch with Your Bodily Feelings. *Psychol. Sci.* 20 (8), 974–980. doi:10.1111/j.1467-9280.2009.02400.x
- Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for Human Electromyographic Research. *Psychophysiol.* 23 (5), 567–589. doi:10.1111/j.1469-8986.1986.tb00676.x

- Frijda, N. H. (2007). *The Laws of Emotion*. Mahwah, NJ: Erlbaum.
- Glenberg, A. M., and Gallese, V. (2012). Action-based Language: A Theory of Language Acquisition, Comprehension, and Production. *cortex* 48 (7), 905–922. doi:10.1016/j.cortex.2011.04.010
- Glenberg, A. M. (2011). How Reading Comprehension Is Embodied and Why that Matters. *Int. Electron. J. Elem. Educ.* 4 (1), 5–18.
- Glenberg, A. M., Webster, B. J., Mouilso, E., Havas, D., and Lindeman, L. M. (2009). Gender, Emotion, and the Embodiment of Language Comprehension. *Emot. Rev.* 1, 151–161. doi:10.1177/1754073908100440
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., and Ditto, P. H. (2011). Mapping the Moral Domain. *J. Personal. Soc. Psychol.* 101 (2), 366–385. doi:10.1037/a0021847
- Greene, J. D. (2014). “The Cognitive Neuroscience of Moral Judgment and Decision-Making,” in *The Cognitive Neuroscience V*. Editor M. S. Gazzaniga (Cambridge, Massachusetts, USA: MIT Press), 1013–1024.
- Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. London, United Kingdom: Penguin Books.
- Havas, D. A., and Matheson, J. (2013). The Functional Role of the Periphery in Emotional Language Comprehension. *Front. Psychol.* 4, 294. doi:10.3389/fpsyg.2013.00294
- Hess, U., and Fischer, A. (2014). Emotional Mimicry: Why and when We Mimic Emotions. *Social Personal. Psychol. Compass* 8, 45–57. doi:10.1111/spc3.12083
- Hoeken, H., and Sinkeldam, J. (2014). The Role of Identification and Perception of Just Outcome in Evoking Emotions in Narrative Persuasion. *J. Commun.* 64, 935–955. doi:10.1111/jcom.12114
- Höfling, T. T. A., Gerdes, A., Föhl, U., and Alpers, G. W. (2020). Read My Face: Automatic Facial Coding versus Psychophysiological Indicators of Emotional Valence and Arousal. *Front. Psychol.* 11, 1388. doi:10.3389/fpsyg.2020.01388
- Künecke, J., Sommer, W., Schacht, A., and Palazova, M. (2015). Embodied Simulation of Emotional Valence: Facial Muscle Responses to Abstract and Concrete Words. *Psychophysiol* 52 (12), 1590–1598. doi:10.1111/psyp.12555
- Kunkel, A. E. (2018). *The Processing Of Moral Transgressions: Investigating The Role Of Affective Evaluations*. Unpublished doctoral dissertation. Tübingen, Germany: Eberhard Karls Universität.
- Larsen, J. T., Norris, C. J., and Cacioppo, J. T. (2003). Effects of Positive and Negative Affect on Electromyographic Activity over Zygomaticus Major and Corrugator Supercilii. *Psychophysiol.* 40 (5), 776–785. doi:10.1111/1469-8986.00078
- Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford: Oxford University Press.
- Leach, C. W., and Spears, R. (2009). Dejection at In-Group Defeat and Schadenfreude toward Second- and Third-Party Out-Groups. *Emotion* 9 (5), 659–665. doi:10.1037/a0016815
- Lee, S., and Potter, R. F. (2018). The Impact of Emotional Words on Listeners' Emotional and Cognitive Responses in the Context of Advertisements. *Commun. Res.* 47 (8), 1155–1180.
- Mirman, D. (2015). Growth Curve Analysis. Philadelphia, PA: Drexel University. Available at: www.danmirman.org (Accessed March 24, 2015).
- Morrison, S., Decety, J., and Molenberghs, P. (2012). The Neuroscience of Group Membership. *Neuropsychologia* 50, 2114–2120. doi:10.1016/j.neuropsychologia.2012.05.014
- Neumann, R., Hess, M., Schulz, S., and Alpers, G. (2005). Automatic Behavioural Responses to Valence: Evidence that Facial Action Is Facilitated by Evaluative Processing. *Cogn. Emot.* 19 (4), 499–513. doi:10.1080/02699930441000364.001
- Nordin, M. (1990). Low-threshold Mechanoreceptive and Nociceptive Units with Unmyelinated (C) Fibres in the Human Supraorbital Nerve. *J. Physiol.* 426, 229–240. doi:10.1113/jphysiol.1990.sp018135
- Pecher, D. (2018). Curb Your Embodiment. *Top. Cogn. Sci.* 10 (3), 501–517. doi:10.1111/tops.12311
- Pecher, D., and Zwaan, R. A. (2017). Flexible Concepts: a Commentary on Kemmerer (2016). *Lang. Cogn. Neurosci.* 32 (4), 444–446. doi:10.1080/23273798.2016.1274413
- Peck, R., and Devore, J. (2008). *Statistics: The Exploration and Analysis of Data*. 6th ed.. London, United Kingdom: Thomson Higher Learning.
- Pulvermüller, F., Shtyrov, Y., and Ilmoniemi, R. (2005). Brain Signatures of Meaning Access in Action Word Recognition. *J. Cogn. Neurosci.* 17 (6), 884–892. doi:10.1162/0898929054021111
- Quené, H., and van den Bergh, H. (2008). Examples of Mixed-Effects Modeling with Crossed Random Effects and with Binomial Data. *J. Mem. Lang.* 59, 413–425. doi:10.1016/j.jml.2008.02.002
- Quené, H., and van den Bergh, H. (2004). On Multi-Level Modeling of Data from Repeated Measures Designs: a Tutorial. *Speech Commun.* 43, 103–121. doi:10.1016/j.specom.2004.02.004
- Scarantino, A. (2014). The Motivational Theory of Emotions. *Moral Psychol. Hum. Agency*, 156–185. doi:10.1093/acprof:oso/9780198717812.003.0008
- Scherer, K. R., and Moors, A. (2019). The Emotion Process: Event Appraisal and Component Differentiation. *Annu. Rev. Psychol.* 70, 719–745. doi:10.1146/annurev-psych-122216-011854
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic Neural Responses Are Modulated by the Perceived Fairness of Others. *Nature* 439, 466–469. doi:10.1038/nature04271
- 't Hart, B. (2017). *That's Frowned upon: Using Facial EMG to Track Evaluation and Simulation during Affective Language Processing*. Utrecht: LOT Dissertation Series.
- 't Hart, B., Struiksma, M. E., Van Boxtel, A., and van Berkum, J. J. A. (2018). Emotion in Stories: Facial EMG Evidence for Both Mental Simulation and Moral Evaluation. *Front. Psychol.* 9, 613. doi:10.3389/fpsyg.2018.00613
- 't Hart, B., Struiksma, M. E., Van Boxtel, A., and van Berkum, J. J. A. (2019). Tracking Affective Language Comprehension: Simulating and Evaluating Character Affect in Morally Loaded Narratives. *Front. Psychol.* 10, 318. doi:10.3389/fpsyg.2019.00318
- Tajfel, H., Billig, M. G., Bundy, R. P., and Flament, C. (1971). Social Categorization and Intergroup Behaviour. *Eur. J. Soc. Psychol.* 1, 149–178. doi:10.1002/ejsp.2420010202
- Tamietto, M., and De Gelder, B. (2010). Neural Bases of the Non-conscious Perception of Emotional Signals. *Nat. Rev. Neurosci.* 11 (10), 697–709. doi:10.1038/nrn2889
- Tappin, B. M., and McKay, R. T. (2016). The Illusion of Moral Superiority. *Soc. Psychol. Personal. Sci.* 8, 1–9. doi:10.1177/1948550616673878
- Tooby, J., and Cosmides, L. (2008). “The Evolutionary Psychology of the Emotions and Their Relationship to Internal Regulatory Variables,” in *Handbook of Emotions*. Editors M. Lewis, J. M. Haviland-Jones, and L. F. Barrett (New York: Guilford), 114–137.
- van Berkum, J. J. A. (2018). “Language Comprehension, Emotion, and Sociality: Aren't We Missing Something?,” in *Oxford Handbook of Psycholinguistics*. Editors S. A. Rueschmeyer and G. Gaskell (Oxford: Oxford University Press), 644–669.
- van Berkum, J. J. A. (in press). “A Survey of Emotion Theories and Their Relevance to Language Research,”. To Appear in *Language and Emotion: An International Handbook*. Editors G. L. Schiewer, J. Altarriba, and B. C. Ng (Berlin: Germany: De Gruyter Mouton).
- van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., and Murre, J. (2009). Right or Wrong? *Psychol. Sci.* 20, 1092–1099. doi:10.1111/j.1467-9280.2009.02411.x
- van Berkum, J. J. A., Struiksma, M. E., and 't Hart, B. (in press). “Using Facial EMG to Track Emotion during Language Comprehension: Past, Present, and Future,” in *Language Electrified: Techniques, Methods, Applications, and Future Perspectives in the Neurophysiological Investigation of Language*. Editors M. Grimaldi, Y. Shtyrov, and E. Brattico (Springer).
- van Boxtel, A. (in press). “Electromyographic (EMG) Responses of Facial Muscles during Language Processing,” in *Mirko Grimaldi, Yury Shtyrov & Elvira Brattico Language Electrified: Techniques, Methods, Applications, and Future Perspectives in the Neurophysiological Investigation of Language* (Springer).
- van Boxtel, A. (2010). “Facial EMG as a Tool for Inferring Affective States,” in *Proceedings of Measuring Behavior 2010, Eindhoven, Netherlands, August 24–27, 2010*. Editors A. J. Spink, F. Grieco, O. E. Krips, L. W. S. Loijens, L. P. J. J. Noldus, and P. H. Zimmerman.
- van der Graaff, J., Meeus, W., de Wied, M., van Boxtel, A., van Lier, P. A. C., Koot, H. M., et al. (2016). Motor, Affective and Cognitive Empathy in Adolescence: Interrelations between Facial Electromyography and Self-Reported Trait and State Measures. *Cogn. Emot.* 30 (4), 745–761. doi:10.1080/02699931.2015.1027665
- Vigliocco, G., Meteyard, L., Andrews, M., and Kousta, S. (2009). Toward a Theory of Semantic Representation. *Lang. Cogn.* 1, 219–247. doi:10.1515/langcog.2009.011
- Vossen, H. G. M., Piotrowski, J. T., and Valkenburg, P. M. (2015). Development of the Adolescent Measure of Empathy and Sympathy (AMES). *Personal. Individual Differences*, 74, 66–71. doi:10.1016/j.paid.2014.09.040

- Willems, R. M., and Casasanto, D. (2011). Flexibility in Embodied Language Understanding. *Front. Psychol.* 2. doi:10.3389/fpsyg.2011.00116
- Winkielman, P., Coulson, S., and Niedenthal, P. (2018). Dynamic Grounding of Emotion Concepts. *Phil. Trans. R. Soc. B* 373 (1752), 20170127. doi:10.1098/rstb.2017.0127
- Winkielman, P., Niedenthal, P., Wielgosz, J., Eelen, J., and Kavanagh, L. C. (2015). Embodiment of Cognition and Emotion. In *APA Handbook of Personality and Social Psychology, Vol. 1. Attitudes and social cognition*. Editors M. Mikulincer, P. R. Shaver, E. Borgida, and J. A. Bargh. Washington, DC: American Psychological Association. 151–175. doi:10.1037/14341-004
- Winter, B. (2020). *Statistics for Linguists: An Introduction Using R*. New York: Routledge.
- Zwaan, R. A. (2014). Embodiment and Language Comprehension: Reframing the Discussion. *Trends Cogn. Sci.* 18, 229–234. doi:10.1016/j.tics.2014.02.008
- Zwaan, R. A., and Pecher, D. (2012). Revisiting Mental Simulation in Language Comprehension: Six Replication Attempts. *PLoS One* 7, 51382. doi:10.1371/journal.pone.0051382
- Zwaan, R. A. (2016). Situation Models, Mental Simulations, and Abstract Concepts in Discourse Comprehension. *Psychon. Bull. Rev.* 23 (4), 1028–1034. doi:10.3758/s13423-015-0864-x
- Zwaan, R. A., Stanfield, R. A., and Yaxley, R. H. (2002). Language Comprehenders Mentally Represent the Shapes of Objects. *Psychol. Sci.* 13, 168–171. doi:10.1111/1467-9280.00430

Conflict of Interest: 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 © 2021 't Hart, Struiksma, van Boxtel and van Berkum. 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.



Face-to-Face Communication in Aphasia: The Influence of Conversation Partner Familiarity on a Collaborative Communication Task

Willemijn Doedens*, Arpita Bose, Lydia Lambert and Lotte Meteyard

ABCD Lab, School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom

OPEN ACCESS

Edited by:

Pia Knoefele,
Humboldt University of Berlin,
Germany

Reviewed by:

Si On Yoon,
The University of Iowa, United States
Tom Sather,
University of Wisconsin-Eau Claire,
United States

*Correspondence:

Willemijn Doedens
w.doedens@donders.ru.nl

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 18 June 2020

Accepted: 27 April 2021

Published: 08 June 2021

Citation:

Doedens W, Bose A, Lambert L and
Meteyard L (2021) Face-to-Face
Communication in Aphasia: The
Influence of Conversation Partner
Familiarity on a Collaborative
Communication Task.
Front. Commun. 6:574051.
doi: 10.3389/fcomm.2021.574051

Aphasia is language impairment due to acquired brain damage. It affects people's ability to communicate effectively in everyday life. Little is known about the influence of environmental factors on everyday communication for people with aphasia (PWA). It is generally assumed that for PWA speaking to a familiar person (i.e. with shared experiences and knowledge) is easier than speaking to a stranger (Howard, Swinburn, and Porter). This assumption is in line with existing psycholinguistic theories of common ground (Clark, 1996), but there is little empirical data to support this assumption. The current study investigated whether PWA benefit from conversation partner (CP) familiarity during goal-directed communication, and how this effect compared to a group of neurologically healthy controls (NHC). Sixteen PWA with mild to severe aphasia, sixteen matched NHC, plus self-selected familiar CPs participated. Pairs were videotaped while completing a collaborative communication task. Pairs faced identical Playmobile rooms: the view of the other's room was blocked. Listeners attempted to replicate the 5-item set-up in the instructor's room. Roles were swapped for each trial. For the unfamiliar condition, participants were paired with another participant's CP (PWA were matched with another PWA's CP based on their aphasia profile). The outcomes were canonical measures of communicative efficiency (i.e. accuracy, time to complete, etc.). Results showed different effects in response to the unfamiliar partner for PWA compared to NHC: In the instructor role, PWA showed faster trial times with the unfamiliar partner, but similar accuracy scores in both conditions. NHC, on the other hand, showed similar trial times across CPs, but higher accuracy scores with the unfamiliar partner. In the listener role, PWA showed a pattern more similar to NHC: equal trial times across conditions, and an improvement in accuracy scores with the unfamiliar partner. Results show that conversation partner familiarity significantly affected communication for PWA dyads on a familiar task, but not for NHC. This research highlights the importance of identifying factors that influence communication for PWA and understanding how this effect varies across aphasia profiles. This knowledge will ultimately inform our assessment and intervention of real-world communication.

Keywords: aphasia, face-to-face communication, common ground, conversation partner familiarity, interaction

INTRODUCTION

One-third of individuals who suffer a stroke will experience aphasia (difficulties speaking and understanding language, reading and writing) (Spaccavento et al., 2013), with detrimental effects on communication and functioning in everyday life (Lam and Wodchis, 2010; Hachoui et al., 2014). When compared against various health conditions (e.g. cancer and Alzheimer's disease) aphasia has the highest impact on quality of life (Hilari et al., 2003; Lam and Wodchis, 2010; Spaccavento et al., 2013). The loss of functional language use affects social, vocational, and emotional well-being (Hilari et al., 2003; Spaccavento et al., 2013), preventing People with Aphasia (PWA) from participating in society and maintaining relationships.

Traditionally, the study of aphasia has focused on impairments of language, with assessment tasks that present isolated language elements (e.g. sounds, words, sentences) in highly controlled lab environments. These studies have been the foundation for the development of reliable assessment instruments and intervention plans targeted at particular profiles of language impairment (Thompson et al., 2008). However, it is generally accepted that such impairment-based performance measures do not reliably predict communication ability in the real world (Holland, 1982; Kolk and Heeschen, 1992; Wilkinson, 1995; Beeke et al., 2007; Davidson et al., 2008; Armstrong et al., 2011). Perhaps because of the complexity of language and communication, the same level of detailed analysis has not been applied to real-world communication for PWA (Leaman and Edmonds, 2019). Providing reliable assessment and evidence-based interventions at the level of communication has, for that reason, remained problematic in aphasiology (Brady et al., 2016). This is a crucial gap in knowledge, as improvement in the ability to communicate in one's own day-to-day environment remains one of the most important long-term goals reported by clinicians and PWA themselves (Thompson et al., 2008).

There is a need for systematic, theoretically driven research on naturalistic communication in aphasia. Recently, we showed how a theoretical framework of situated language use, borrowed from research with neurologically healthy controls (NHC) (Clark, 1996), can be applied to aphasia rehabilitation (Doedens and Meteyard, 2019). It provides a structure along which different components of real-world communication, and their influence on a person's ability to communicate, can be examined systematically. The framework defines communication as being 1) interactive—including at least one other person, 2) multimodal—involving multiple channels of information and 3) contextual—grounded in shared situational, personal and social knowledge.

Here, we will focus on the contextual aspect of communication. One part of contextual information is common ground shared with a conversation partner (CP) – part of which is modulated by the familiarity of that CP. For PWA, questionnaires on communication often distinguish between the ability to communicate with familiar and unfamiliar CPs (e.g. the disability questionnaire of the

Comprehensive Aphasia Test; Howard et al., 2004; or the Aphasia Impact Questionnaire-21; Swinburn et al., 2018). The assumption is often made that it is easier for PWA to speak to a familiar person than speaking to a stranger (Green, 1982; Wirz et al., 1990; Ferguson, 1994; Perkins, 1995; Howe et al., 2008; Laakso and Godt, 2016). The familiarity advantage has also been reported by PWA as an influential factor when it comes to ease of communicating (Dalemans et al., 2010).

Conversation Partner Familiarity in the Control Literature

Conversation partner familiarity is more specifically defined as personal common ground (Clark, 1996). This constitutes a set of past experiences, beliefs and knowledge that are shared between two people. There is a belief that people who know each other well understand each other better (i.e. communication is more accurate) and need fewer words to understand each other (i.e. communication is more efficient) than when strangers communicate (Sillars et al., 1997; Pollmann and Krahmer, 2017). Part of the assumption is that when people know each other well, they are better at taking the other person's perspective (Mead, 1934; Fussell and Krauss, 1989). In line with this, a number of studies have shown that people that are familiar with each other (e.g. friends, couples), feel like they understand and know each other well (Pollmann and Finkenauer, 2009) and overestimate the degree to which they will be understood by their familiar partner (Dixon et al., 1998; Savitsky et al., 2011; Riordan and Trichtinger, 2016). Research with NHC has shown that the presence of shared experiences and knowledge, personal common ground, can facilitate communication (Clark, 1996; Brown-Schmidt, 2009a). In conversation, interlocutors can rely on this shared information as being "given", i.e. not requiring too much further explanation. As a result, CPs can rely on more informal, implicit, abbreviated language in their exchanges (Herrmann, 1983; Hornstein, 1985; Clark, 1996). By relying on the "givenness" of shared information, speakers can produce less complete utterances (Bard et al., 2014), while listeners can use personal common ground to restrict the number of possible interpretations based on the shared knowledge (Brown-Schmidt, 2009b; Skipper, 2014). It has, therefore, been suggested that reliance on shared knowledge can make communication less effortful and more efficient (i.e. requiring less time, fewer words and/or less cognitive energy, Boyle et al., 1994; Clark, 1996; Horton and Gerrig, 2005; Smith et al., 2005; Zwaan, 2016). In line with this, Isaacs and Clark (1987) showed that New Yorkers were more efficient in describing buildings in the city when they spoke to other New Yorkers compared to non-New Yorkers. In a study by Pollmann and Krahmer (2017), familiar and unfamiliar pairs did not differ on communicative efficiency, as measured in number of words and minutes spent on the task. The familiar pairs did show an advantage in accuracy scores, but this was dependent on the task: accuracy scores on a game of Taboo via email (with no non-verbal communication or feedback in the form of clarification questions) did not differ between the two groups. In a face-to-face setting, familiar pairs achieved higher scores than unfamiliar

pairs. Brown-Schmidt (2009a) found a similar effect of task: only performance on an interactive task showed an influence of common ground between conversation partners (participants were manipulated on shared knowledge within the experiment, not on personal familiarity; Brown-Schmidt, 2009b). The authors argued that the type of task and its complexity might have been of influence: the more complex, the greater the need to rely on shared information to complete the task. Interestingly, Pollmann and Krahmer (2017) showed that in addition to the higher accuracy scores, familiar pairs reported higher levels of motivation and enjoyment of the game, suggesting that these factors might influence communicative efficiency and accuracy as well. Finally, on the non-interactive email task, an effect of friendship closeness on accuracy was found for the group of familiar pairs. Andersson and Ronnberg (1997) showed that participants performed better on a word association task when working with friends compared to strangers. Fussell and Krauss (1989) also reported higher levels of accuracy when subjects were asked to interpret a message that was recorded specifically for them by a friend, than when they were asked to interpret a message that was recorded by a stranger. While the difference in accuracy scores between these conditions was significant, it was very small. The authors hypothesized that the traditional referential communication task might not have required participants to rely on personal common ground. Instead, reliance on general, community-wide knowledge would have enabled participants to successfully interpret messages recorded by a stranger (Fussell and Krauss, 1989). Furthermore, the authors suggested that the degree to which familiar pairs know each other (i.e. length of time, level of intimacy) might have mediated this effect, as the familiar pairs in their study had known each other for less than six months (Clark and Schaefer, 1987; Fussell and Krauss, 1989).

The research on the benefit of conversation partner familiarity is, however, inconclusive. Gould et al. (2002) did not find differences across tasks between familiar and unfamiliar pairs. The authors also suggested that the familiarity effect might only be present in particular communicative or experimental contexts, as well as depend on the type of relationship that is studied. Schober and Carstensen (2010) also found no difference between familiar and unfamiliar pairs on their efficiency and accuracy in describing unfamiliar things, such as tangram shapes. While Pollmann and Krahmer (2017) found differences in accuracy between familiar and unfamiliar pairs on a face-to-face task, no differences were found between the groups on efficiency. Finally, it has also been suggested that the existence of shared knowledge between two interlocutors might not necessarily lead to a reliance on that shared knowledge per se. Instead, it might lead the speaker to rely more on their own knowledge. While this can facilitate communication on some topics (when speaking about topics that are part of shared knowledge), it can also lead to greater confusion when communicating about topics that are not part of common ground (Wu and Keysar, 2007; Savitsky et al., 2011). Overall, the research on the effect of conversation partner familiarity on communication efficiency and accuracy remains relatively inconclusive. It is suggested to depend on factors such as the type and complexity of task, the topic of conversation and

whether it requires personally shared knowledge to be understood, the type, length and intimacy of the relationship under study and the motivation of the interlocutors on the task.

Conversation Partner Familiarity in the Aphasia Literature

Only a small number of studies have explored the influence of personal common ground on communication for PWA. Leaman and Edmonds (2019) analyzed and compared the unstructured conversations of eight PWA (most with mild anomic aphasia) with a familiar conversation partner (FCP) and an unfamiliar speech and language therapist (SLT). The authors reported no differences on measures of communicative success, on linguistic measures such as grammaticality (morphological and verb tense/mood errors) and sentence production (correct use of a complete sentence frame and the relevance of lexical items in the frame in the discourse context), or on lexical retrieval behaviors (false starts, repetitions, pauses of 2+ s, etc.). These findings suggest that some linguistic characteristics of conversation for PWA might remain stable across conversation partners.

Kistner (2017) assessed gesture use by twenty PWA (ranging from severe to mild aphasia) and NHC in conversation with FCPs and unfamiliar conversation partners (UFCP). A procedural and a narrative conversational task were used to elicit conversation. UFCPs were SLT students or researchers with knowledge of aphasia. In this study, both PWA and NHC showed an increase in the number of gestures when speaking to the UFCP as compared to the FCP. The authors hypothesized that gesture production increased to help disambiguate meaning or as speech became more complex. With the UFCP, this need increased due to the lack of shared reference. Williams et al. (1994) explored the influence of conversation topic and conversation partner familiarity for 22 PWA and ten NHC on a procedural and story-retell task. The syntactic complexity measures in the study showed no effect of CP familiarity (Williams et al., 1994). On the same dataset, Li et al. (1995) found no significant differences on discourse grammar between conversations with FCPs and UFCPs, except on the description of the setting in the story retell task, where PWA provided more detail with the FCP. The authors suggested PWA might have felt more comfortable or at ease with the familiar CP, which could have facilitated recall of that particular aspect of the story. Finally, case studies by Gurland et al. (1982) and Lubinski et al. (1980) showed that PWA used different communication styles depending on the familiarity of their CP: Gurland et al. (1982) showed a greater number of acknowledgments were produced in conversation with a familiar CP, while with the unfamiliar CP, topic-relevant turns increased. The authors suggested PWA might take on a more “passive, less informative role with the spouse (familiar CP) vs. the clinician (unfamiliar CP)” (Williams et al., 1994). Lubinski et al. (1980) compared the unstructured conversation of one PWA with a familiar (spouse) and a therapy session with an UFCP (in this case, a SLT). The topic of conversation was not controlled for. The number of conversational breakdowns and repairs were assessed: similar types of conversational breakdowns were found with the FCP and

UFCP. The way in which the breakdowns were repaired, however, differed significantly. UFCPs (SLT) tended to gloss over the breakdowns, while FCPs (spouse) actively attempted to repair them collaboratively with the PWA. The authors suggested that one reason for this difference was the different goals each CP had during their conversation with the PWA: the clinician often let the PWA repair the conversational trouble, while the spouse wanted to collaboratively discuss the plans for that day. Ferguson (1994) found no difference in trouble indicating behaviors between FCP and UFCP in a study with eight PWA, where the conversational topic was slightly more aligned. The authors found that the way these troubles were dealt with was different depending on the familiarity of the CP: UFCP more often took on the responsibility of repairing the trouble (i.e. “other-repair”), rather than letting the PWA repair the trouble (i.e. “self-repair”). The authors hypothesized that by not letting PWA repair the trouble as often, UFCPs might have been driven by a desire to avoid potential continued conversational breakdown. The familiarity manipulation might not have been sufficient in this latter study: the role of UFCP was filled by someone who knew the PWA less well compared to the FCP, but still had known the PWA for years.

Confounding Factors in Interactive Communication in Aphasia

In addition to the effect of personal common ground, there are two confounding factors that have been shown to influence the communicative ability of PWA. First, research has shown that communication for PWA is influenced by the extent of knowledge the CP has about aphasia, the language impairment and on potential communication strategies they can use to facilitate communication (Rayner and Marshall, 2003). CPs with knowledge of communicating with PWA have been shown to enable PWA to communicate more effectively and increase the PWA’s level of participation in conversation (Lindsay and Wilkinson, 1999; Pound et al., 2000; Kagan et al., 2001; Simmons-Mackie et al., 2010; Wilkinson and Wielaert, 2012; Nykanen et al., 2013). PWA also specifically self-report the positive impact of communicating with someone who knows about aphasia and what communication strategies to use during conversation (Dalemans et al., 2010; Harmon, 2020).

Second, the sense of comfort and support experienced during communication has been suggested as an important factor for communicative ability (Dalemans et al., 2010; Worrall et al., 2010; Harmon, 2020). Though not exclusively, this sense of comfort and support is often associated with the familiarity of the CP. This line of reasoning suggests that the fear of not being able to express oneself due to the language impairment and subsequently the fear of “losing face” or of being perceived unfavorably because of the communication difficulties, can make communication with an UFCP more effortful and a more negative experience (Harmon, 2020). For PWA, this could potentially result in more errors in their language production, more and longer word searches, or potentially result in avoidance of the interaction with the UFCP resulting in, for example, shorter interactions altogether.

Suggestions to this end have been made in the literature (Li et al., 1995; Kistner, 2017). In a discussion of the use of compensatory communication strategies by PWA, Simmons-Mackie and Damico showed that PWA may vary their communication strategies depending on the goal in a particular context, such as “looking okay”, rather than being maximally communicatively effective (Simmons-Mackie and Damico, 1995). To the knowledge of the authors the sense of being at ease during communication and the influence of conversation partner familiarity has not been explored empirically.

In sum, the existing research suggests that the presence of personal common ground can influence communication for PWA. The existing evidence base is small, but it seems that the effect of conversation partner familiarity might depend on the level at which communication is measured. It seems that lower level linguistic measures such as verb or sentence production could remain stable across different conversation partners, while higher level communication strategies such as the use of gesture or the repair of conversational trouble might vary. More work is needed, however, to assess whether this advantage exists, how it manifests, whether it exists for all types of aphasia, and if it is mediated by other factors such as aphasia severity. It is crucial to control for the influence of other confounding factors such as knowledge of aphasia of the CP, the sense of comfort experienced by the PWA as well as the conversation topic.

The Current Study

The aim of the current study was to investigate whether CP familiarity affects communication for PWA. Participants completed a collaborative task that required communication in two different conditions: once with a FCP, and once with an UFCP. Participants were in two groups: PWA with a NHC conversation partner, and NHC with a NHC conversation partner. To investigate the question of personal common ground we controlled for the potential influence of two confounding factors. Knowledge of aphasia was controlled for by swapping the CPs of pairs of PWA who were matched on their linguistic and communication impairment profiles. Knowledge of aphasia was also tested through a questionnaire. The sense of comfort was taken into account by asking each familiar and unfamiliar pair to indicate the level of comfort they felt while completing the task with their conversation partner. These research questions were part of a bigger pre-registration (<https://osf.io/9xwm7>).

A collaborative task was used to elicit naturalistic communication between the participant pairs. Different versions of this task have been used in previous research with NHC (Clark and Wilkes-Gibbs, 1986; Boyle et al., 1994; Clark, 1996; Clark and Krych, 2004; Howarth and Anderson, 2007; Lysander and Horton, 2012) where naturalistic communication is investigated in a controlled lab setting. This experimental setup made it possible to adhere to the previously described framework of real-world communication and to manipulate variables within that framework (Doedens and Meteyard, 2019), see **Table 1**.

To measure the effect of the experimental manipulation on communicative success for PWA and NHC, a selection of key

TABLE 1 | Different types of self-initiated repairs and clarification requests that were coded.

Type of code	Description	Example
Self-initiated repairs		
Revised repair	The interlocutor repeats the main clause with modifications	"The man goes under the chair. ... no I mean he goes on the chair"
Addition repair	The interlocutor provides additional information to the main clause	"The sofa is in opposite the window ... the small window"
Word finding repair	The interlocutor explicitly has word-finding difficulties (repetitions without revisions, additions or explicit statements of difficulties finding a word are not included)	"The d ... d ... oh what is that word?"
Clarification requests		
Request for elaboration or clarification	The interlocutor asks their CP to provide more information on what has been said. This type of clarification request includes most wh-questions	"Which window?" or "Where?"
Statement of not understanding	The interlocutor indicates that they did not follow what their CP said	"I don't understand" or "Huh?"
Partial or complete repetitions	The interlocutor repeats (part of) a phrase as produced by the CP, sometimes with a questioning intonation, to check if they have understood correctly	CP1: "by the window on the left" CP2: "by the window on the left?"
Insertion	When the CP is speaking the interlocutor inserts a word or phrase that fits into the utterance of the CP. This can happen, for example, when the CP pauses to search for a word. The insertion functions as an evaluation for the interlocutor to assess if they have correctly understood the utterance of the CP.	CP1: "and then the sofa is facing the.." CP2: "The tv cabinet?" CP1: "yes, the tv cabinet"
Indirect request for clarification	The interlocutor asks for a repetition of what has been said, indirectly indicating they (might not) have not fully understood or followed	"Please speak more slowly"

outcome measures was made based on previous literature on CP familiarity with PWA and NHC. Based on research with NHC, measures of trial time and task accuracy were selected. Previous research with PWA suggests that the number of times trouble is identified during conversation, can be indicative of communicative success (e.g., Beeke, 2012). We therefore also included a measure of self-initiated repair (i.e. instances where the "instructor" initiates a self-correction) and other-initiated repair (i.e. instances where the "listener" requests clarification on what has been said) as a measure of communicative success.

Due to the nature of the task, an additional analysis was included (not part of the pre-registration). This analysis aimed to assess the influence of role (instructor or listener) on goal-directed communication. The current study included trials in which PWA and NHC took turns in an "instructor" role, requiring them to actively communicate new information to their CP. Conversely, participants also took on the "listener" role, requiring them to follow instructions from their CP. Previous studies with NHC have assumed no differences in role for measures such as time taken and accuracy (Boyle et al., 1994). Therefore, no difference in roles was expected for NHC for the measures of time and accuracy. However, as PWA present with impairments of language production and comprehension, a difference in performance based on role can be expected. For the number of self-initiated repairs and clarification requests, we expected an effect of role for both groups. Self-initiated repairs are naturally expected to be more frequent when someone speaks more (i.e. the "instructor" role), while Clarification requests are naturally expected to be more frequent when someone is in the "listener" role. Finally, given the inherent variability of the language impairment within the aphasia group, we include a visual representation of the individual difference in scores between conditions (i.e. familiar-unfamiliar), ordered by a standardized measure of aphasia severity. This will provide insight into the spread of

individual data-points within the aphasia group, and how this compares to the NHC group.

Analysis addressed the following research questions:

- (1) What is the effect of speaker role (instructor/listener) on goal-directed communication?
- (2) What is the effect of CP familiarity (personal common ground) on goal-directed, face-to-face communication in aphasia?
- (3) Do PWA differ from NHC in how they respond to CP familiarity during goal directed communication?

Based on the existing literature, it was hypothesized that it will be easier for PWA to complete the task with a FCP than with an UFCP, as evidenced by the familiar pair taking less time, requiring fewer repairs, obtaining higher accuracy scores and fewer requests for clarification. Based on the case study by Lubinski et al. (1980), it could be the case that the number of repairs falls into the category of more lower-level behavior which remains stable across conversation partners. In comparison to NHC, we expect PWA to show a similar direction of the effect of CP familiarity. Due to the presence of the language impairment for PWA, we expect the CP familiarity effect to be greater for PWA compared to NHC, i.e. we expect PWA to have more difficulty adapting to communicating with an UFCP, or to benefit more from communicating with their FCP (see Table 2).

MATERIALS AND METHODS

Ethics Statement

This study was carried out with ethical clearance from the School of Psychology and Clinical Language Sciences, University of Reading (Ref: 2018-093-LM). All participants provided informed consent prior to taking part in the study. Consent

TABLE 2 | Hypotheses for each outcome measure, shown for each research question (RQ).

Outcome measure	Description	Hypotheses			
		RQ 1 (instructor vs listener)		RQ 2 (PWA)	RQ 3* (NHC vs. PWA)
		PWA	NHC		
Trial time	Faster times indicate “better” communication	Instructor ≠ listener	Instructor = listener	Familiar < unfamiliar	NHC < PWA
Task accuracy	Higher accuracy indicates “better” communication	Instructor ≠ listener	Instructor = listener	Familiar > unfamiliar	NHC < PWA
Self-initiated repairs	Fewer repairs indicate “better” communication	Instructor > listener	Instructor > listener	Familiar < unfamiliar	NHC < PWA
Clarification requests	Fewer requests indicate “better” communication	Instructor < listener	Instructor < listener	Familiar < unfamiliar	NHC < PWA

Notes: RQ1: the effect of role, RQ2: the effect of CP familiarity for PWA, RQ3: the difference in effect of CP familiarity between PWA and NHC. * hypotheses are about the difference scores between the familiar and unfamiliar conditions. A larger difference score represents a bigger impact of the experimental manipulation.

and information forms were adapted to aphasia friendly format for the participants with aphasia.

Participants

Sixteen participants with post-stroke aphasia (42–72 years, $M = 60.94$, $SD = 9.41$) and sixteen control participants (NHC, 52–84 years ($M = 64.94$, $SD = 9.66$)) took part in the current study. PWA and controls were matched for age ($t(30) = 1.19$, $p = 0.245$) and years of education ($t(29) = -0.07$, $p = 0.946$). Nine male and seven female PWA were recruited through the Aphasia Research Registry of the School of Clinical Language Sciences, University of Reading (British Academy Grant ARP scheme 190023), as well as through local stroke groups. PWA were at least one-year post-stroke (1–14 years, $M = 7.04$, $SD = 3.85$) and were native speakers of English prior to the stroke. Exclusion criteria were coexisting neurological diagnoses such as dementia and an inability to provide consent due to severe comprehension difficulties. Seven male and nine female NHC were recruited through the older adult research panel at the School of Psychology, University of Reading. Exclusion criteria were a history of neurological illness. All subjects reported normal or corrected-to-normal vision and hearing.

All participants brought along a FCP to take part in the study with them. The PWA self-nominated a FCP who they spoke to regularly. Six male and ten female FCPs agreed to take part (partner, friend or family member between the ages of 22–72 years, $M = 54.12$, $SD = 15.12$, see **Table 3** for more details). All FCPs except those labeled child (only ID 48), ex-partner and friend lived in the same house with the PWA. For NHC, partners were recruited as the FCP (aged range 51–79 years, $M = 64.12$, $SD = 7.57$, see in the **Supplementary Table S1**). All FCPs lived in the same house with their partner. All FCPs reported normal or corrected-to-normal vision and hearing and did not report a history of neurological illness.

All PWA completed the Western Aphasia Battery-Revised (WAB-R; Kertesz, 2009). The aphasia quotient score (AQ) ranged from 11.60–94.2 ($M = 65.88$, $SD = 26.59$), severities ranging from very severe to mild (see **Table 3** for an overview). To obtain a standardized measure of communicative ability, PWA also completed the Scenario Test United Kingdom (Hilari et al., 2018). Scores ranged from 20.25–54 (maximum score = 54, $M = 45.64$, $SD = 8.83$; details shown in **Table 3**). Thirteen out of sixteen PWA had some degree of weakness (hemiparesis) on the right-hand side due to the stroke. All PWA were able to use

their unaffected arm and hand effectively. All PWA were mobile enough to attend the experiment at the University clinic. One PWA attended the clinic in a wheelchair.

All participants without aphasia completed the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005), a cognitive screening tool for mild cognitive impairment. Scores ranged between 17–30 ($M = 27.23$, $SD = 2.49$). Six participants scored below the cut-off score of the test (<26 points; one NHC with score 17; two FCP to NHC with scores 23; three FCP to PWA with scores: 22, 23, 24), suggesting the potential presence of mild cognitive impairment. Due to difficulties in recruiting the PWA subjects, their partners and age- and years of education-matched controls, none were excluded from participation on the basis of their MoCA scores. Following reviewer comments, an additional statistical analysis was run in which these subjects were excluded, as described in the Statistical Analysis section. The MoCA was not administered with the people with aphasia. The heavy reliance of this test on language in its instructions and responses makes this test unsuitable and unreliable for administration with people with existing language processing difficulties.

Procedure

All participants were invited to take part in a study about conversation and different CPs. Background testing with PWA was completed either at the participant's home or at the School of Clinical Language Sciences, University of Reading. All NHC completed background testing at the University of Reading.

For the experimental session, two participants and their respective FCP were invited to the Speech and Language Therapy Clinic at the University of Reading.

Task

The experimental design consists of a collaborative, referential communication task (Clark and Krych, 2004) that allows pairs to interact and communicate freely, replicating a real-life face-to-face communicative setting. Pairs sat across from each other, in front of identical playmobile rooms (see **Figure 1**). The view of the other person's room was blocked by a low barrier. Five items were placed in one room (instructor), while the other room (listener) remained empty with six items placed on the side of the room. Pairs were asked to replicate the setup of the instructor's room in the listener's room. They were asked to communicate as they normally would, including the use of any communication aids. Pen and paper were provided for both

TABLE 3 | Descriptive for PWA and their familiar conversation partners, ordered by WAB-AQ score (lowest to highest). Scenario Test classification is based on the percentiles of the Dutch norm group, solely used here to provide a descriptive classification.

PWA											Familiar conversation partner						
ID	Sex	Age	Years of education	Time post stroke (months)	WAB-AQ	Severity	Class	Scenario test			ID	Sex	Age	Years of education	MoCA	Partner type	Years knowing PWA
								Raw score	Perc	Classification*							
69	M	72	10	71	11.6	Very severe	Global	36	34	Limited	70	F	69	11	22*	Partner	54
43	M	57	12	58	18.3	Very severe	Broca/Global	44.47	49	Okay	44	F	56	13	27	Partner	30
15	F	57	19	56	27.6	Severe	Wernicke	20.25	13	Almost none	16	M	60	17	23*	Partner	40
19	M	65	13	177	56.8	Moderate	Broca	44.47	49	Okay	20	F	60	13	26	Partner	40
45	F	42	19	12	58.2	Moderate	Broca	48	63	Okay	46	F	43	13	29	Friend	10
5	M	72	16	123	61.6	Moderate	Broca	41.63	42	Okay	6	F	55	13	26	Ex-partner	25
41	M	45	11	75	62.5	Moderate	Broca	37	34	Limited	42	F	41	13	30	Partner	20
37	F	68	17	136	69.5	Moderate	Conduction	48	63	Okay	38	M	69	17	27	Partner	50
47	F	70	10	116	72.2	Moderate	Broca	49	68	Okay	48	F	49	15	27	Child	49
78	F	51	15	20	74.1	Moderate	Anomic	44.47	49	Okay	79	M	23	17	28	Child	23
11	M	66	17	56	83.8	Mild	Conduction	53	93	Good	12	M	22	13	28	Grandchild	22
53	M	67	19	42	89.4	Mild	Anomic/Transcortical Sensory	53	93	Good	54	F	72	14.5	27	Partner	33
7	F	51	16	55	90.1	Mild	Anomic	54	100	Good	8	M	63	14	24*	Partner	10
1	M	64	11	133	90.8	Mild	Anomic	53	93	Good	2	F	58	18	28	Partner	26
67	F	65	16	110	93.3	Mild	Anomic	53	93	Good	68	M	66	18.5	27	Partner	47
13	M	63	18	111	94.2	Mild	Transcortical sensory	51	78	Good	14	F	60	18	29	Partner	42

Notes: *Classification refers to the communicative ability of the PWA: "almost no communicative ability", "seriously limited communicative ability", "okay communicative ability in simple situations" and "good communicative ability in simple situations". * indicates a MoCA score below the cut-off (<26).

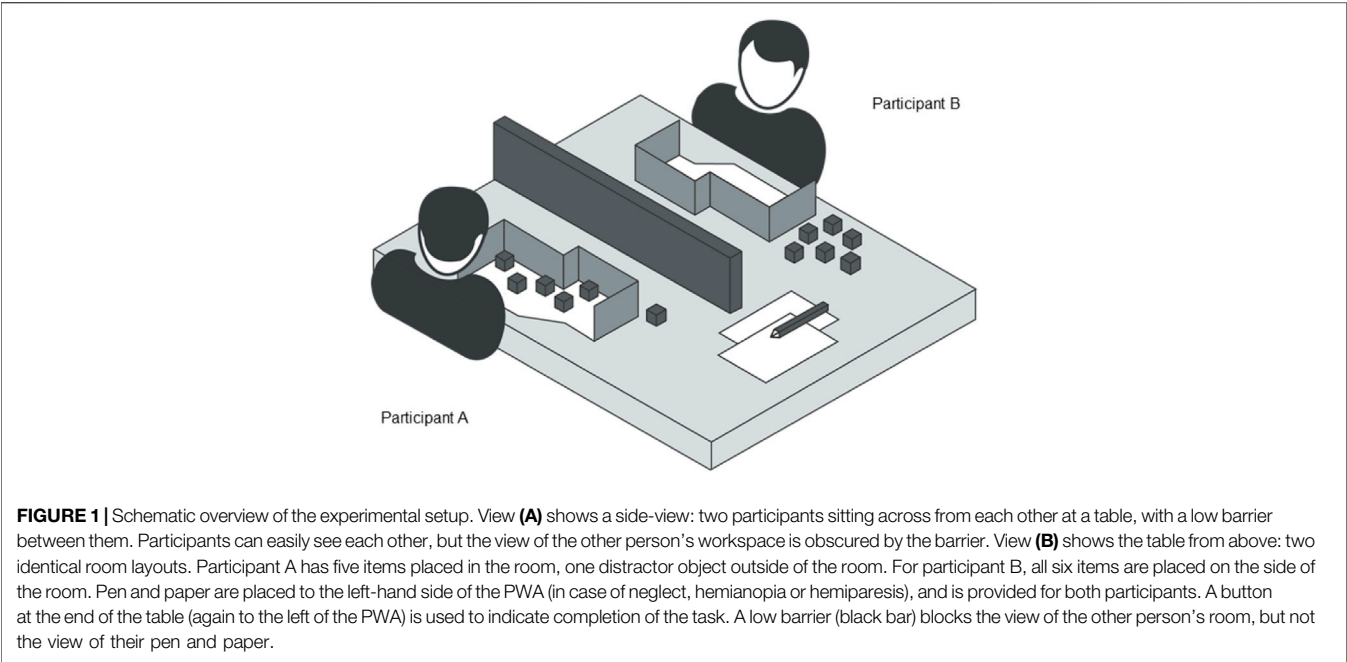


TABLE 4 | Description of the experimental manipulation according to the theoretical framework of face-to-face communication.

Component of the framework	Manipulation in the current experiment
Interactive	Unrestricted interaction with the CP (i.e. no restrictions on giving feedback, asking questions, etc.) Interaction with a single CP
Multimodal	Unrestricted use of all communicative modalities (gesture, facial expressions, body posture, intonation, language) Optional use of pen and paper for drawing and writing (specified as "if you need to, you can use") Added option of communication aid
Common ground	
Personal	Interaction with a familiar CP and with an unfamiliar CP (<i>the main experimental manipulation</i>)
Communal	—
Communicative	Repetition of the same task across 6 trials allowing CPs to build communicative context. Theoretically, this context could have carried over into the unfamiliar condition, where the same task was repeated
Situational	The use of 6 concrete, highly frequent, familiar, and recognisable objects and their physical location in relation to a physical space and each other

participants. Participants were instructed not to show items to their CP or to look over the barrier at the other room. In many ways, the current set-up echoes that of PACE in the aphasia literature (Davis and Wilcox, 1985; Davis, 2005). Aphasia friendly images were used to visually support the instruction for all participants. The experimenter left the room for the duration of the task. When the pair completed the task, they pressed a button. The experimenter then re-entered, took a picture of both rooms, and showed the participants the result. Any paper used was collected by the experimenter and the next trial was set up.

Each pair (familiar and unfamiliar) completed the game six times: For each trial, roles (instructor/listener) were swapped, resulting in three instructor trials and three listener trials for each participant. The starting role was counterbalanced across participants. A different setup of items was used for each trial, the order of which was randomized for each pair.

The experimental manipulations of the current study can be summarized according to the previously described framework of

real-world communication (Doedens and Meteyard, 2019). See **Table 4**.

Materials

An empty Playmobil room with four windows and one door was used for the current experiment. Six Playmobil objects were selected based on psycholinguistic features that have been shown to influence lexical retrieval in PWA (Nickels and Howard, 1995, see in the **Supplementary Table S2** for details). The items were selected based on high levels of concreteness, familiarity and imageability, as well as (roughly) low number of phonemes to facilitate naming of the items as much as possible.

Six different room setups were created by placing five Playmobil items in various configurations across the room (see **Figure 2**). One item (counterbalanced across trials) was a distractor and placed outside of the room. Three additional objects were permanently placed in the same location across all six trials, functioning as reference points for the other objects:

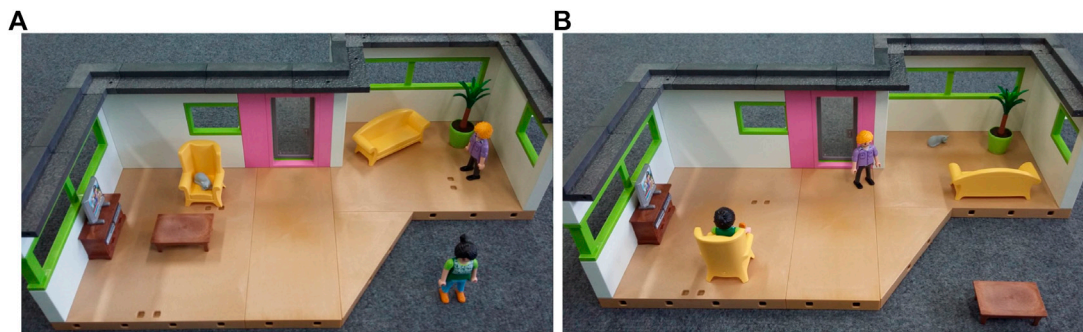


FIGURE 2 | Two examples of item setup in the Playmobil living room. Five items are placed in different locations in the room, one item was always left on the side as the distractor (and did not need to be placed by the listener).

1) a chest of drawers with 2) a television on top and a 3) potted plant in the opposite corner of the room.

Between conditions, the physical appearance (i.e. the color) of the cat and the hair of the woman was changed to incorporate some variation in the stimuli. Two reference objects were also changed: the potted plant was replaced by a different potted plant and the television was replaced by a set of books. The location of all the items remained constant.

Familiarity Manipulation

In the unfamiliar condition, each participant was matched with another participant's FCP. PWA were matched with the FCP of a PWA with a similar aphasia profile based on their WAB-AQ score and their communication score on the Scenario Test (Meulen et al., 2010). This way, PWA were matched with an FCP who was unfamiliar at a personal level, but who had experience communicating with someone with roughly similar communication difficulties. Where possible, PWA were also matched on age and gender (see in the **Supplementary Table S3** for more details). In the control group, NHCs were matched on gender, age and years of education (in order of priority). For the unfamiliar condition, each NHC was paired up with their matched NHC's FCP (see in the **Supplementary Table S4** for details on matching).

At the beginning of each condition, each participant was asked to rate the familiarity of their CP on an aphasia-friendly Likert scale (0 = this person is a stranger, 5 = I know this person extremely well). For both groups, the FCP was rated higher in familiarity (PWA: $M = 3.55$, $SD = 0.62$, NHC: $M = 3.97$, $SD = 0.12$) compared to the UFCP (PWA: $M = 0.52$, $SD = 0.92$, NHC: $M = 0.03$, $SD = 0.12$). The difference in familiarity ratings was significant for both groups (PWA: $t(30) = 10.97$, $p < 0.001$, NHC: $t(30) = 89.09$, $p < 0.001$).

The order of conditions was not counterbalanced: All participants first completed the familiar condition, followed by the unfamiliar condition. The authors decided against counterbalancing the order of conditions to minimize potential anxieties about communicating with an UFCP for the PWA.

Controlling for Knowledge of Aphasia

To control for knowledge of aphasia, all CPs of PWA filled out a questionnaire testing their knowledge of aphasia (factual

knowledge and knowledge on communication strategies as described in Rayner and Marshall, 2003). Knowledge of aphasia was similar for FCP (factual: $M = 10.1$, $SD = 3.02$, strategies: $M = 22.4$, $SD = 1.9$) and UFCP (factual: $M = 10.4$, $SD = 0.98$, strategies: $M = 22.3$, $SD = 2.75$). A paired t -test showed no significant differences between FCPs and UFCPs (factual knowledge: $V = 6$, $p = 0.854$, knowledge of communication strategies: $t(6) = 0.16$, $p = 0.877$)¹.

Sense of Comfort With the CP

The degree of comfort participants felt with their FCP and UFCP during the task was taken into account: At the end of each condition, each participant was presented with a statement ("I feel that my partner and I communicate comfortably together") and a visual 5-point Likert scale (0 = completely disagree, 4 = completely agree). For both PWA and NHC, the degree of comfort they felt with their CP was roughly equal in the familiar (PWA: $M = 3.56$, $SD = 0.51$, NHC: $M = 3.71$, $SD = 0.47$) and unfamiliar condition (PWA: $M = 3.28$, $SD = 0.52$, NHC: $M = 3.53$, $SD = 0.62$). A non-parametric paired t -test showed no significant difference between the degree of comfort participants felt with their FCP and UFCP (PWA: $V = 18$, $p = 0.119$, NHC: $V = 20$, $p = 0.299$).

Coding

All trials were video and audio recorded. Videos of the interactions were coded in ELAN (The Language Archive, 2019). For the purpose of this study, the following measures were coded:

Trial time. All videos were coded for trial time. Trial time was defined as the moment participants started to communicate on a trial (speak, draw, gesture, etc.) until the moment one of the participants pressed the button to signal the experimenter to come into the room.

Task accuracy. Task accuracy was defined as the correct placement of the items in the listener's room as compared to the instructor's room as set up by the experimenter. The setup of the instructor's and listener's room was photographed at the end of each trial. Both images were scored by two independent judges

¹One PWA was excluded from this analysis. The data from this questionnaire is currently inaccessible due to COVID-19 lockdown restrictions on the campus of the University of Reading.

on accuracy (correct/incorrect) of two aspects of the item: its location (in the room and in relation to other objects), its orientation. For the people, two additional aspects were coded: the action that was undertaken by the item (i.e. standing, sitting, etc.) and the positioning of the arms. For all other objects, the action was always coded as correct, resulting in a maximum score of three per item, and four per person (a maximum score of 20 and a minimum score of 4, examples of low, moderate and high accuracy scores are provided in the **Supplementary Figure S1**). In case of doubt due to different angles of the pictures, a grid was superimposed on the floor of each image using Kinovea software (Charmant and Contrib., 2006-2011).

Self-initiated repairs. Self-initiated repairs were defined as instances where a participant explicitly attempted to repair or change their own output (often described as the repair initiation; Wilkinson, 2006; Schegloff et al., 1977). A self-initiated repair was always an explicit correction initiated by the interlocutor themselves, without any prompts from the conversation partner. Three different types of self-initiated repairs were coded, partially based on Perkins (1993) (see **Table 1**). For the word-finding repairs, repetitions of parts of words are expected, but if parts of a word are repeated without revisions, additions or explicit statements of difficulties finding a word, these are not coded as a repair. All self-initiated repairs are coded, regardless of the way in which the repair is resolved (i.e. by the interlocutor themselves, collaboratively with their conversation partner or by the conversation partner). Whether a repair is successful or not was not coded (i.e. whether the correction creates a correct utterance or not, or whether the correct word is produced, or the search is abandoned). Non-verbal instances of self-initiated repairs are also included (e.g. direct gaze at the partner to provide help in a word search, Beeke, 2012). The total number of self-initiated repairs was counted for each trial and participant.

Clarification requests. Clarification requests are defined as instances when one interlocutor indicates to their conversation partner that they have not fully understood what has been said (also described as an “other-initiated” repair; Schegloff et al., 1977). Five types of clarification requests were coded, partly based on Schegloff et al., (1977) (see **Table 1**). Coding included verbal and non-verbal clarification requests such as clear eye gazes and frowns, or clear shrugs directed at the CP. The total number of clarification requests was counted for each trial and participant.

Coding of the latter two outcome measures is expected to be more subjective compared to the first two outcome measures due to the inherent nature of the coding process (Beeke et al., 2007). Self-initiated repairs were coded by a second rater (native English-speaking speech and language therapy student), resulting in a moderate intraclass correlation coefficient ($ICC = 0.74$, $CI = 0.51-0.87$, $p < 0.001$, calculated in R studio using the psych package version 1.9.12.31; (Revelle, 2020).

Statistical Analysis

All outcome measures showed a non-normal distribution and contained outliers. The outcome measures also showed significant differences in variance between groups. Log-linear transformations did not eliminate the problems of normality or

extreme values in the data. To avoid relying on assumptions of normality, a bootstrap procedure was used to obtain a distribution based on resampling of the existing data, from which the test statistic was derived (Wilcox, 2012). Outliers and differences in variance between groups were dealt with by choosing robust analyses based on the median (percentile bootstrap) and 20% trimmed means (bootstrap-t). An alpha threshold of 0.05 was used to determine statistical significance. All analyses were run in R Studio version 1.1.463 (RStudio, 2020). The results from the median analysis are reported in the paper. When there was a difference in outcome, results from both analyses are discussed. For all bootstrapping methods, 10,000 bootstrap samples were used (Rousselet et al., 2019).

First, we ran an omnibus between-by-within-by-within 2 (group: PWA/NHC) \times 2 (role: instructor/listener) \times 2 (condition: familiar/unfamiliar) robust analyses on all outcome measures: of the median (bwwmcpb in Wilcox, 2012) and the 20% trimmed mean (bwwmcp in Wilcox, 2012). We then ran specific follow up comparisons to answer our research questions.

Research question 1: An effect of role (instructor or listener).

Research question 2: An effect of CP familiarity for PWA.

We analyzed each group separately (PWA or NHC). This helps us to identify patterns for each group of participants, and to address whether role and familiarity have an effect on goal directed communication. Two factors were entered into analysis. First, the condition of familiarity (familiar/unfamiliar), as this was our principle experimental manipulation. Second, the role of the participant (instructor/listener). Role was expected to affect the nature of communication in the goal directed communication task for PWA.

Thus, within subjects 2 (role: instructor/listener) \times 2 (condition: familiar/unfamiliar) robust analyses were conducted on all outcome measures: of the median (wmmcpb in Wilcox, 2012), and of the 20% trimmed mean (wmmcpbt in Wilcox, 2012). Planned comparisons were conducted for significant main effects: for a main effect of role, a dependent groups analysis on each level of condition (familiar/unfamiliar) was run on the median and 20% trimmed mean (bootdpci and ydbt, respectively, in Wilcox, 2012). For a main significant effect of condition, the same dependent groups analysis was conducted on each level of role (instructor/listener). The full results of these analyses are reported in the **Supplementary Table S5**. Results of the planned comparisons are reported in the **Supplementary Tables S7–S10**.

Research question 3: An effect of CP familiarity for PWA compared to NHC.

We first accounted for the effect of Role (see above) by splitting data into Instructor or Listener trials. We then completed between-by-within 2 (group: PWA/NHC) \times 2 (condition: familiar/unfamiliar) robust analyses on all outcome measures: of the median (sppba, sppbb and sppbi in Wilcox, 2012) and the 20% trimmed mean (bwtrimbt in Wilcox, 2012). Planned comparisons on significant main effects of group (PWA vs NHC) were conducted with an independent groups analysis (pb2gen in Wilcox, 2012), to test the effect at each level of condition (familiar/unfamiliar). For a main significant effect of condition, a dependent groups analysis (bootdpci and ydbt, as described above,

in Wilcox, 2012) was conducted on each level of group (PWA/NHC). The full results of these analyses are reported in the **Supplementary Table S11**. Results of the planned comparisons are reported in the **Supplementary Tables S13–S16**.

To evaluate the influence of participants who scored below cut-off on the MoCA, all statistical analyses reported above were conducted a second time. In these analyses all the sessions (familiar and unfamiliar) in which one participant within a dyad had a MoCA score below the cut-off were excluded. This resulted in the exclusion of data from three dyads in the familiar and unfamiliar conditions, both for PWA and NHC. The results of the 2×2 analyses are shown in the **Supplementary Tables S6–S12**. Any differences in the outcomes of the $2 \times 2 \times 2$ omnibus are mentioned in the results below.

To assess the individual patterns of behavior, a difference score between conditions was calculated for each role: for each participant, the value of each outcome measure for the familiar condition was deducted from the value of the unfamiliar condition. The difference scores were then plotted by group. This visual representation of individual difference scores by aphasia severity is not part of the formal statistical analysis, due to the small and unequal numbers of subjects within the different groups of aphasia severity.

RESULTS

Trial Time

In the omnibus analysis, the analysis based on the median did not show any significant main effects or interactions. The 20% trimmed mean analysis resulted in a main effect group (PWA vs NHC; estimated mean difference = 363.61 s, $p = 0.026$), with longer trial times for NHC compared to PWA. The main effect of condition was also significant (familiar vs unfamiliar; estimated mean difference = 248.78 s, $p = 0.049$), with longer trial times for the familiar condition. No other main effect or interaction was significant.

The omnibus analysis without the participants with low MoCA scores based on the median did show a main effect of group (PWA vs. NHC, $p = 0.018$). In the trimmed means analysis, the main effect of condition was no longer significant ($p = 0.052$).

Research question 1: An effect of role (instructor or listener).

Research question 2: An effect of CP familiarity for PWA.

PWA

The 2 (role: instructor/listener) \times 2 (condition: familiar/unfamiliar) analysis showed a main effect of role (instructor vs. listener; estimated median difference = 156.65 s, $p < 0.001$), with longer trial times for instructors (median = 332.49, $CI = 259.57, 404.11$) compared to listeners (median = 251.72, $CI = 191.80, 362.96$). Planned pairwise comparisons show that the difference in trial time for instructors vs. listeners holds for both conditions (*familiar*: $p = 0.015$, *unfamiliar*: $p = 0.035$)². For PWA,

total trial times were longer when they were in the instructor role as compared to the listener role. See **Figure 3**.

There was a main effect of condition (familiar vs. unfamiliar; estimated median difference = 167.34 s, $p < 0.001$), with longer trial times in the familiar condition (median = 363.92, $CI = 307.84, 404.11$) compared to the unfamiliar condition (median = 251.28, $CI = 198.96, 277.92$). Planned comparisons show that the difference in trial time between familiar and unfamiliar conditions was significant for the instructor role ($p < 0.001$) and not when PWA take on the listener role ($p = 0.201$). In the instructor role, PWA took less time to complete a trial in the unfamiliar condition compared to the familiar condition. In the listener role, trial times were more equal. See **Figure 3**.

The interaction of role*condition was not significant (estimated median difference = 38.02 s, $p = 0.457$).

NHC

There were no significant effects (*role*: estimated median difference = 173.4 s, $p = 0.014$, *condition*: estimated median difference = 75.75 s, $p = 0.46$, *interaction*: estimated median difference = 21.26 s, $p = 0.76$). For NHC trial times were constant for both roles (instructor/listener) and conditions (familiar/unfamiliar). See **Figure 3**.

Research question 3: An effect of CP familiarity for PWA compared to NHC.

Instructor Trials

There was no significant main effect of group (PWA/NHC, estimated median difference = 78.37 s, $p = 0.199$), with PWA and NHC showing similar overall total trial times for Instructor trials.

There was a significant main effect for condition (familiar vs unfamiliar; estimated median difference = 68.09 s, $p = 0.01$)³, with longer trial times in the familiar condition (median = 384.50, $CI = 343.35, 491.88$) compared to the unfamiliar condition (median = 284.29, $CI = 259.57, 457.81$). Planned comparisons within subjects showed that for PWA, total trial times were faster in the unfamiliar condition compared to the familiar condition (see **Figure 3**). Whilst the main effect of condition was significant, planned comparisons did not show a difference within subjects for the familiar vs unfamiliar conditions for NHC ($p = 0.203$).

The interaction of group * condition was not significant (estimated median difference = -53.52 s, $p = 0.253$).

Listener Trials

There was a main effect of group (PWA vs. NHC; estimated median difference = 144 s, $p = 0.008$). Total trial times were longer for NHC (median = 374.35, $CI = 351.61, 457.16$) compared to PWA (median = 270.22, $CI = 173.27, 365.58$).

²Planned comparisons using the 20% trimmed mean did not show a significant difference between roles ($p = 0.136$) for the unfamiliar condition. The presence of a larger number of outliers will have affected the trimmed means analysis more than the median. We will therefore rely on the median analysis.

³The 20% trimmed means analysis did not show the significant effect for condition ($Q = 3.44$, $Q_{crit} = 4.16$, $p = 0.074$). The distribution of data in the two conditions is slightly different. This in combination with the presence of multiple outliers in the familiar condition will have affected outcome of the median and trimmed mean analyses differently. To avoid the influence of too many outliers, the median analysis will be used here.

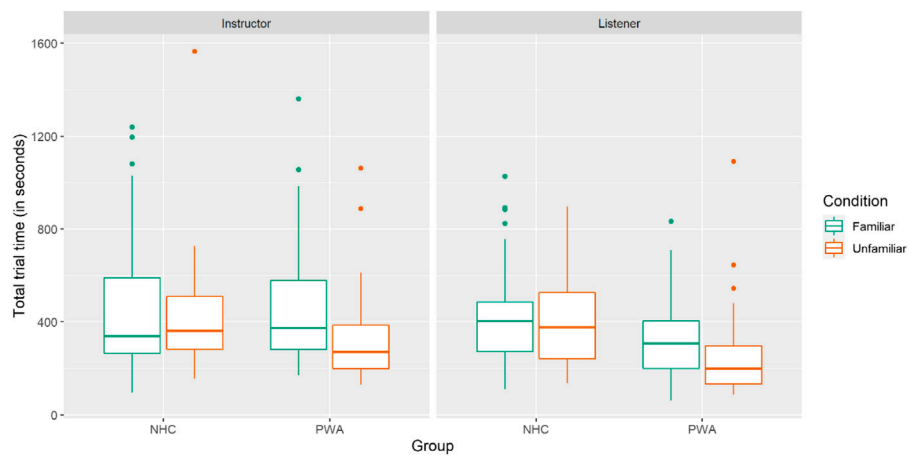


FIGURE 3 | Boxplots showing total trial time by condition and group, for each role (instructor/listener).

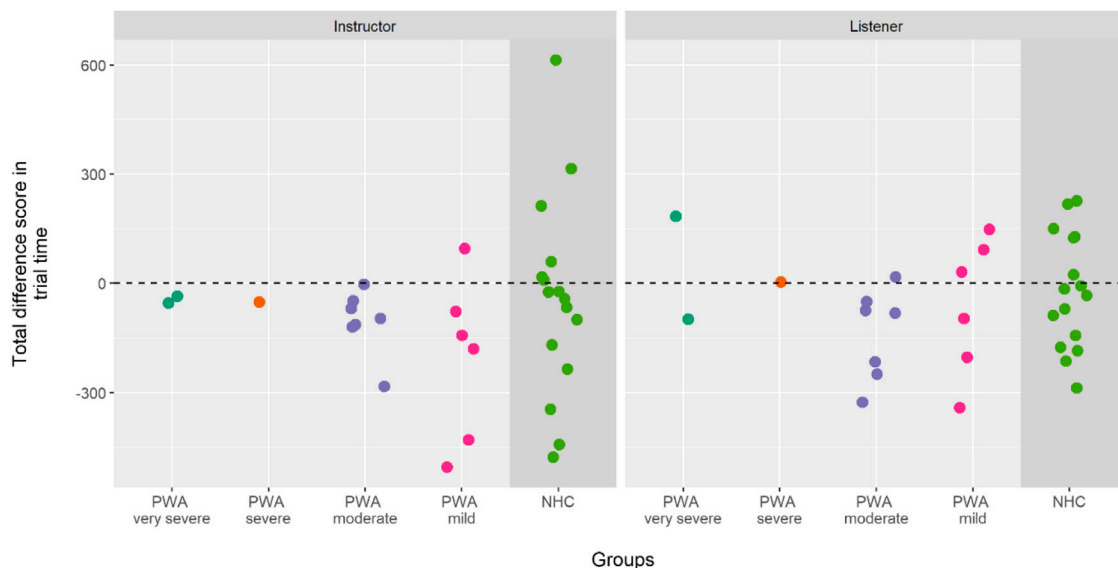


FIGURE 4 | Plot showing individual data points for both groups for difference score between familiar and unfamiliar conditions, by role (PWA grouped by WAB categorization). Zero represents no change in total trial time between conditions, negative values indicate a shorter total trial time in the unfamiliar condition compared to the familiar condition.

Planned comparisons between subjects showed a significant difference in the unfamiliar condition ($p = 0.009$), with trial times for PWA significantly faster than for NHC. The same comparison for the familiar condition was not significant ($p = 0.158$). See **Figure 3**.

The main effect of condition (estimated median difference = 60.4 s, $p = 0.08$) and the interaction of group * condition was not significant (estimated median difference = -53.52 s, $p = 0.399$).

Supplementary Figure S2 shows the changes in total trial times for each group, condition and role by trial. This figure shows a relatively smooth transition in trial times between the final trial of the familiar condition and the first trial of the unfamiliar condition for both groups.

Summary of Results for Trial Time

Total trial times for NHC dyads were slower than PWA dyads (this effect held when participants with low MOCA scores were removed). Total trial times were longer when PWA took on the instructor role, regardless of the familiarity of the CP. In addition, total trial times for PWA were faster for the unfamiliar condition. For NHC, there was no significant difference in trial times in the familiar and unfamiliar conditions, or between the different roles.

Changes at the Level of Individual Dyads

To explore the results descriptively, we plotted the changes in total trial time for each dyad (**Figure 4**). Data for PWA has been

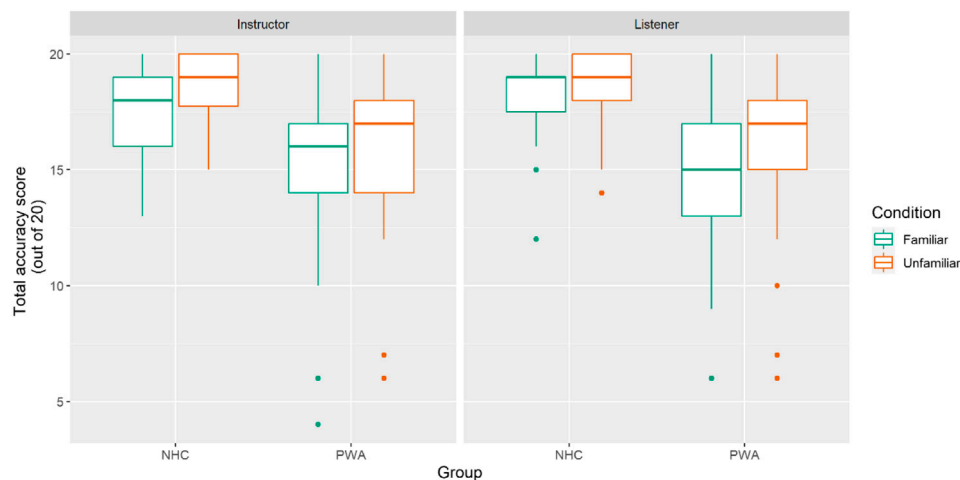


FIGURE 5 | Boxplots showing total accuracy score (out of 20) by condition and group, for each role (instructor/listener).

grouped according the severity of aphasia for the PWA participant. In general, the spread of data points for both groups (PWA or NHC) is greater for the Instructor role. There is a trend that, as aphasia severity decreases (moving left to right along the x axis), the distribution of difference scores increases with more dyads showing faster total trial times in the unfamiliar condition (negative values). Note that this is confounded by there being more data points for moderate to mild PWA. However, it is tentative evidence that for PWA who are less severe, total trial times were likely to be faster for the unfamiliar condition.

Task Accuracy

The omnibus analysis showed a main effect of group (PWA vs NHC; estimated median difference = 9.67; $p < 0.001$), with NHC scoring higher than PWA. There was also a main effect of condition (familiar vs unfamiliar; estimated median difference = -4.33; $p = 0.008$), with accuracy scores higher in the unfamiliar condition compared to the familiar condition. The main effect of role was not significant (instructor vs. listener; $p = 0.707$). No two-way interactions were significant. Finally, the three-way interaction was significant (group by role by condition; estimated median difference = -3; $p = 0.033$), indicating that accuracy scores were different, depending on the group (PWA vs. NHC), role (instructor vs. listener) and condition (familiar vs. unfamiliar). The patterns driving this three-way interaction are explored below.

The omnibus analysis without the participants with low MoCA scores showed the same effects and interactions.

Research question 1: An effect of role (instructor or listener).

Research question 2: An effect of CP familiarity for PWA.

PWA

As in the omnibus analysis, there was a significant main effect of condition (familiar vs. unfamiliar; estimated median difference = -1.67, $p = 0.049$). Task accuracy scores were higher in the unfamiliar condition (median = 16.17, $CI = 15.42, 17.67$) compared to the familiar condition (median = 15, $CI =$

14.08, 17.17). Planned comparisons showed that in the instructor role, PWA did not show a significant change in accuracy scores between familiar and unfamiliar conditions ($p = 0.607$). In the listener role, the difference in accuracy scores between conditions (familiar/unfamiliar) was significant in the trimmed mean analysis ($p = 0.007$, *median analysis*: $p = 0.062$). Accuracy was higher in the unfamiliar condition compared to the familiar condition. It therefore seems that the main effect of condition (familiar vs. unfamiliar) for PWA was driven by the improvement in accuracy scores in the listener role (see **Figure 5**).

There was no significant main effect of role (estimated median difference = 0.67, $p = 0.538$) and no significant interaction of role*condition (estimated median difference = 1.83, $p = 0.167$).

NHC

There was a significant main effect of condition (familiar vs. unfamiliar; estimated median difference = -0.67, $p = 0.015$), with NHC obtaining higher accuracy scores in the unfamiliar condition (median = 18.75, $CI = 18.33, 19.0$) compared to the familiar condition (median = 18.33, $CI = 17.17, 18.67$). Planned pairwise comparisons showed a significant effect of condition for NHC in the instructor role as measured by the 20% trimmed means analysis ($p = 0.043$, *median analysis*: $p = 0.131$), but not for the listener role ($p = 0.182$). As instructors, NHC obtained higher accuracy scores in the unfamiliar condition compared to the familiar condition, driven more by the significant improvement in scores in the instructor role.

There were no significant effects of role (estimated median difference = -0.67, $p = 0.173$) nor an interaction of role*condition (estimated median difference = -0.33, $p = 0.338$).

Research question 3: An effect of CP familiarity for PWA compared to NHC.

Instructor Trials

The 2 (group: PWA/NHC) \times 2 (condition: familiar/unfamiliar) analysis showed a significant main effect of group (PWA vs.

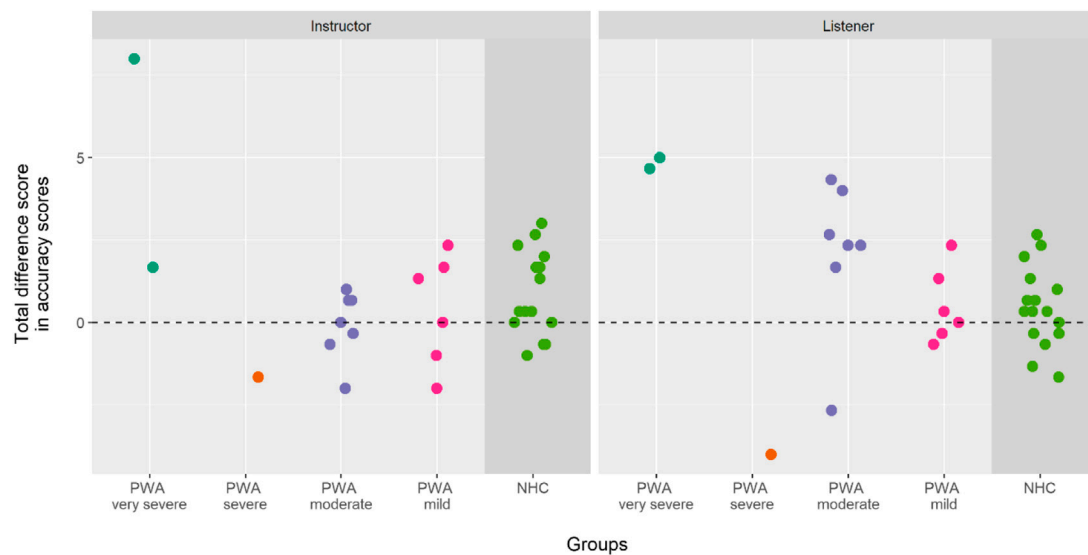


FIGURE 6 | Plot showing individual data points for PWA for difference score between familiar and unfamiliar conditions, by role, categorized by WAB categorization. Zero represents no change in accuracy scores between conditions, negative values indicate a lower accuracy score in the unfamiliar condition compared to the familiar condition.

NHC; estimated median difference = 2, $p = 0.004$), with higher accuracy scores for NHC (median = 18.33, $CI = 17.5, 18.83$) compared to PWA (median = 16.17, $CI = 15.0, 16.75$). Planned pairwise comparisons showed that the effect of group was significant in both conditions (*familiar*: $p = 0.022$; *unfamiliar*: $p = 0.002$). In the instructor role, NHC had significantly higher accuracy scores compared to PWA (see **Figure 5**).

The main effect of condition (familiar/unfamiliar) and the interaction of group*condition were not significant (*condition*: estimated median difference = -0.33 , $p = 0.407$, *interaction*: estimated median difference < -0.01 , $p = 0.95$).

Listener Trials

There was a significant main effect of group (PWA vs. NHC; estimated median difference = 2.83, $p < 0.001$). PWA obtained lower accuracy scores (median = 15.58, $CI = 14.17, 17.0$) compared to NHC (median = 18.42, $CI = 18.33, 19.17$). Planned pairwise comparisons showed that the effect of group was significant in both conditions (*familiar*: $p < 0.001$, *unfamiliar*: $p < 0.001$). In the listener role, NHC had significantly higher accuracy scores compared to PWA. See **Figure 5**.

The main effect of condition (familiar vs. unfamiliar) was significant in the 20% trimmed means analysis⁴ ($Q = 14.09$, $Q_{crit} = 4.36$, $p = 0.002$), with higher accuracy scores in the unfamiliar condition (median = 17.67, $CI = 17.17, 18.67$) compared to the familiar condition (median = 17, $CI = 16, 18.17$). Planned pairwise

comparisons showed that the effect of condition was significant for PWA in the 20% trimmed means analysis ($p = 0.007$, *median analysis*: $p = 0.062$), but not for NHC ($p = 0.182$). In the listener role, PWA had significantly higher accuracy scores in the unfamiliar compared to familiar condition. See **Figure 5**.

The interaction group*condition was not significant (estimated median difference = -1.67 , $p = 0.093$).

Supplementary Figure S3 shows the changes in accuracy scores for each group, condition and role by trial. This figure shows a relatively smooth transition in accuracy scores between the final trial of the familiar condition and the first trial of the unfamiliar condition for both groups, such that there is no clear practice effect across trials. The distributions of accuracy scores differ, accuracy scores become less variable in the unfamiliar condition.

Summary of Results for Accuracy

Overall, NHC always scored higher on task accuracy compared to PWA. When analyzed as separate groups, accuracy scores were higher in the unfamiliar condition for both PWA and NHC. These main effects survived the removal of participants with low MOCA scores.

Changes at the Level of Individual Dyads

The changes in accuracy scores for each dyad are plotted in **Figure 6**. Data for PWA has been grouped according the severity of aphasia for the PWA participant. In general, the spread of data points is greater for PWA than for NHC. Based on aphasia severity, there doesn't seem to be a clear pattern of change in accuracy scores between condition: while the two participants with very severe aphasia have a higher accuracy score in the unfamiliar condition compared to the familiar condition, the opposite is true for the participant with severe aphasia. This is true in the listener and instructor role. The moderate and mild

⁴The main effect of condition was not significant based on the median analysis (estimated median difference = -0.67 , $p = 0.152$). The difference in variance in both conditions could have affected the median less compared to the trimmed mean. In addition to this, a potential ceiling effect means that the median might not reflect the improvements in performance of NHC between conditions as reliably as the trimmed means. We will therefore rely on the trimmed mean analysis here.

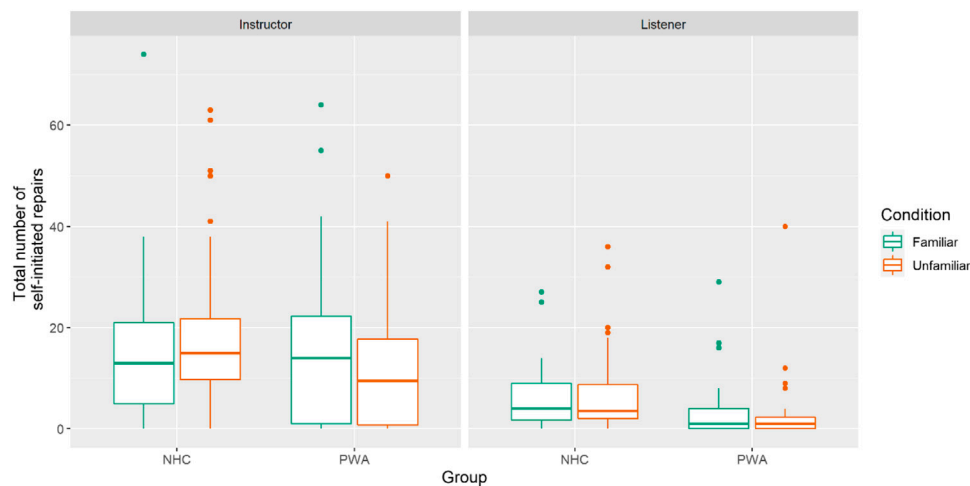


FIGURE 7 | Boxplots showing total number of self-initiated repairs by condition and group, for each role (instructor/listener).

severity groups show a pattern that is more similar to the NHC group, with a tendency to show higher accuracy scores for the unfamiliar condition.

Self-Initiated Repairs

The omnibus analysis showed a significant effect of role (instructor vs listener, estimated median difference = 37.67, $p < 0.001$), with a higher number of self-initiated repairs in the instructor role. No other main effects or interactions were significant.

The omnibus analysis without the participants with low MoCA scores based on the median also showed a main effect of group (PWA vs. NHC, $p = 0.021$), with a greater number of repairs in the instructor role.

Research question 1: An effect of role (instructor or listener).

Research question 2: An effect of CP familiarity for PWA.

PWA

The 2 (role: instructor/listener) \times 2 (condition: familiar/unfamiliar) analysis showed a significant main effect of role (instructor vs. listener; estimated median difference = 17, $p < 0.001$). The number of self-initiated repairs was higher in the instructor role (median = 13, $CI = 1.17, 18$) compared to the listener role (median = 2.08, $CI = 0.17, 4.42$). Planned pairwise comparisons on the effect of role show that the significant difference in number of self-initiated repairs was present in both the familiar ($p < 0.001$) and unfamiliar condition ($p < 0.001$). For PWA, the number of self-initiated repairs was higher when they were in the instructor role compared to the listener role. See **Figure 7**.

There was no significant effect of condition (estimated median difference = 0.5, $p = 0.201$) or of the interaction role \times condition (estimated median difference = -0.17, $p = 0.806$).

NHC

There was a significant main effect of role (instructor vs. listener; estimated median difference = 23, $p < 0.001$). The number of self-initiated repairs was higher in the instructor role (median = 15.25, $CI = 13.17, 23.0$) compared to the listener role (median = 5.75, CI

= 2.5, 9.17). Planned pairwise comparisons on the effect of role show that for NHC the significant difference in number of self-initiated repairs was present in both the familiar ($p = 0.007$) and unfamiliar condition ($p < 0.001$). For NHC, the number of self-initiated repairs was higher when they were in the instructor role compared to the listener role. See **Figure 7**.

There were no significant effects of condition (estimated median difference = 0.33, $p = 0.806$) or interaction of role \times condition (estimated median difference = -2.5, $p = 0.173$).

Research question 3: An effect of CP familiarity for PWA compared to NHC.

Instructor Trials

The 2 (group: PWA/NHC) \times 2 (condition: familiar/unfamiliar) showed no significant effects for group (estimated median difference = 2.25, $p = 0.559$), condition (estimated median difference = 0, $p = 1$) or the interaction group \times condition (estimated median difference = -1, $p = 0.539$). In the instructor role, PWA and NHC self-initiated repairs a similar number of times. The rate of self-initiated repairs was the same in both conditions. See **Figure 7**.

Listener Trials

The 2 (group: PWA/NHC) \times 2 (condition: familiar/unfamiliar) analysis showed a main effect of group (PWA vs. NHC; estimated median difference = 3.25, $p = 0.039$), with a larger number of self-initiated repairs by NHC (median = 5.75, $CI = 2.5, 9.17$) compared to PWA (median = 2.08, $CI = 0.17, 4.42$). Planned pairwise comparisons show that the difference in number of self-initiated repairs did not differ significantly in the familiar condition ($p = 0.133$) or the unfamiliar condition ($p = 0.055$)⁵.

⁵In the unfamiliar condition, the difference in self-initiated repairs between groups was significant in the 20% trimmed means analysis ($p = 0.031$). The presence of a large number of outliers could have inflated the effect of the trimmed means analysis. We will therefore rely on the more conservative median analysis here.

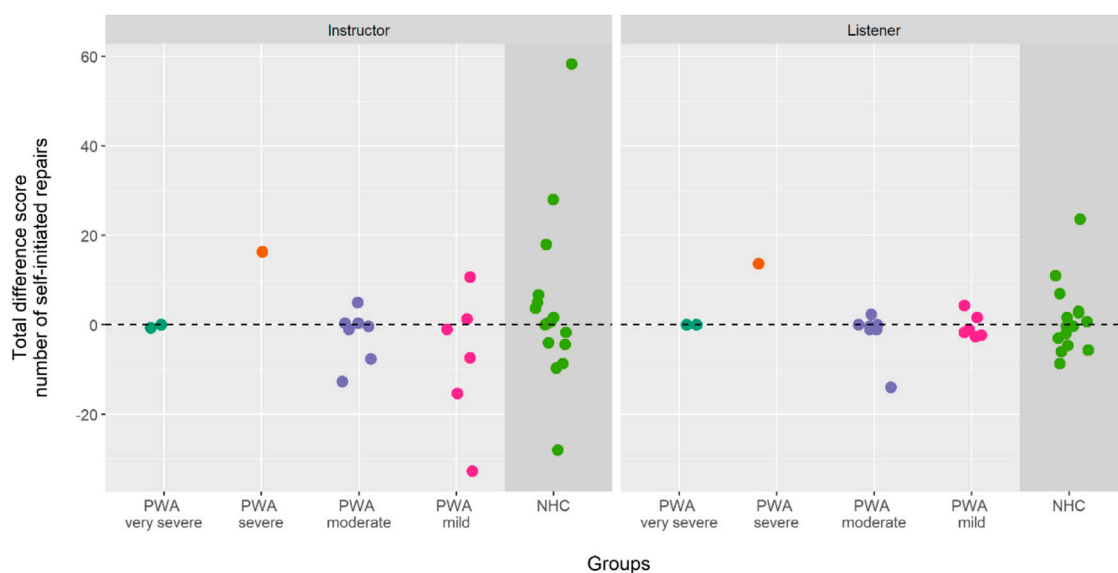


FIGURE 8 | Plot showing individual data points for PWA for difference score between familiar and unfamiliar conditions, by role, categorized by WAB categorization. Zero represents no change in self-initiated repairs between conditions, negative values indicate a smaller number of self-initiated repairs in the unfamiliar condition compared to the familiar condition.

As shown in **Figure 7**, averaged across conditions, NHC show a larger number of self-initiated repairs compared to PWA. This effect disappears when this difference is assessed at the level of each condition (familiar and unfamiliar).

The effect of condition and the interaction were not significant (*condition*: estimated median difference = 0.33, $p = 0.511$, *interaction*: estimated median difference = 0.17, $p = 0.934$).

Supplementary Figure S4 shows the changes in the number of self-initiated repairs for each group, condition and role by trial. This figure shows a relatively smooth transition in the number of self-initiated repairs between the final trial of the familiar condition and the first trial of the unfamiliar condition for PWA and NHC, such that there are no clear practice effects. More statistical analyses at the trial level would need to be conducted to confirm these observations.

Summary of Results for Number of Self-Initiated Repairs

The number of self-initiated repairs depended on the role participants fulfilled: in the instructor role, both PWA and NHC showed a higher number of self-initiated repairs compared to the listener trials, this main effect survived the removal of participants with low MOCA scores. Compared to NHC, PWA produced a similar number of repairs in the instructor role. As listeners, PWA produced fewer self-initiated repairs compared to NHC.

Changes at the Level of Individual Dyads

The changes in number of self-initiated repairs for each dyad is plotted in **Figure 8**. Data for PWA has been grouped according the severity of aphasia for the PWA participant. In general, the spread of data points for both groups (PWA and NHC) is greater for the

instructor role. In the instructor role, there is a trend that as aphasia severity decreases (moving left to right along the x axis), the distribution of difference scores becomes more like the NHC group, with more dyads showing lower number of self-initiated repairs in the unfamiliar condition (negative values). Interestingly, PWA do not show the tendency to increase the number of self-initiated repairs to the extent that NHC do (positive values): PWA tend to show fewer self-initiated repairs in the unfamiliar condition compared to the familiar condition, while NHC show a slightly more equal distribution between decreases and increases in the number of self-initiated repairs. There is tentative evidence that for PWA who are less severe, the number of self-initiated repairs was likely to be smaller for the unfamiliar condition.

Clarification Requests

In the omnibus analysis, there was a significant main effect of group (PWA vs. NHC, estimated median difference = 24.67, $p = 0.002$), with the NHC producing a higher number of clarification requests than PWA. There was a significant main effect of role (instructor vs. listener, estimated median difference = -46.67, $p < 0.001$), with a higher number of clarification requests produced in the listener role as compared to the instructor role. There was a significant main effect of condition in the trimmed means analysis (familiar vs unfamiliar, estimated mean difference = 10.6, $p = 0.033$), with a higher number of clarification requests with the familiar CP than the unfamiliar CP. There was a significant interaction between group and role (estimated median difference = -25.33, $p = 0.001$), with the NHC producing a greater number of clarification requests than PWA when in the listener role, however, this difference was absent for the instructor role (principally because so few clarification requests are made in the instructor role, see **Figure 9** and **Supplementary Figure S5**).

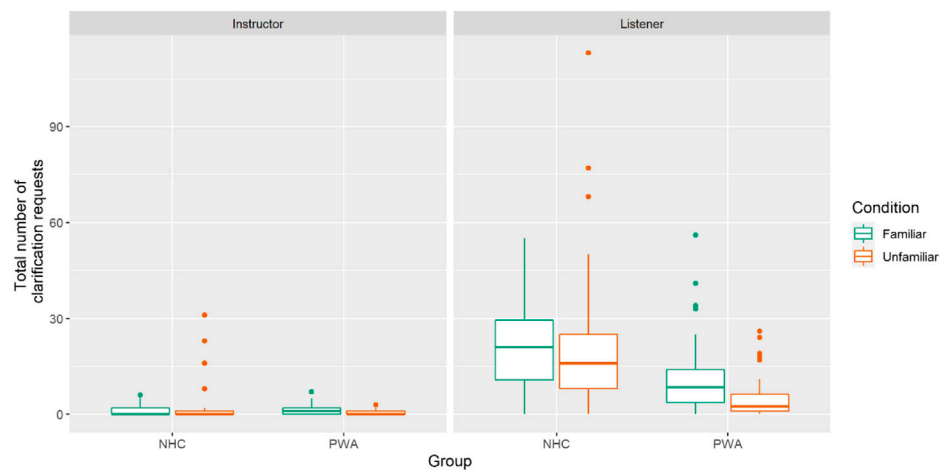


FIGURE 9 | Boxplots showing total number of clarification requests by condition and group, for each role (instructor/listener).

The omnibus analysis without the participants with low MoCA scores based on the median did not show a main effect of condition (familiar vs. unfamiliar, $p = 0.344$). All other effects were as reported above.

Research question 1: An effect of role (instructor or listener).

Research question 2: An effect of CP familiarity for PWA.

PWA

The 2 (role: instructor/listener) \times 2 (condition: familiar/unfamiliar) analysis showed a significant main effect of role (instructor vs. listener; estimated median difference = -10 , $p < 0.001$)⁶. The number of clarification requests was higher when PWA took on the listener role (median = 6.17 , $CI = 3.0, 12.67$) compared to the instructor role (median = 0.75 , $CI = 0.42, 1.33$). Planned comparisons show that for PWA, the difference in number of clarification requests between instructor and listener role was significant in the familiar ($p < 0.001$) and the unfamiliar condition ($p < 0.001$). PWA showed a higher number of clarification requests in the listener role compared to the instructor role. See **Figure 9**.

The main effect of condition was significant (familiar vs. unfamiliar; estimated median difference = 3.33 , $p = 0.010$)⁷, with higher number of clarification requests in the familiar condition (median = 4.42 , $CI = 2.0, 10.5$) compared to the unfamiliar condition (median = 2.17 , $CI = 0.67, 3.25$). Pairwise comparisons resulted in a significant difference between conditions for both the listener ($p = 0.002$) and instructor roles ($p = 0.036$). PWA showed a higher number of

clarification requests in the familiar condition compared to the unfamiliar condition. See **Figure 9**.

The interaction of role \times condition was also significant (estimated median difference = -2.17 , $p = 0.046$)⁸. In the instructor role, there is no difference in number of clarification requests between the familiar and unfamiliar conditions. In the listener role, PWA produced a smaller number of clarification requests in the unfamiliar condition compared to the familiar condition. See **Figure 9**.

NHC

For NHC there was a significant main effect of role (instructor vs. listener; estimated median difference = -34.5 , $p < 0.001$), with more clarification requests produced in the listener role (median = 18.17 , $CI = 13.58, 28.17$) compared to the instructor role (median = 0.75 , $CI = 0.5, 1.08$). Planned pairwise comparisons for the effect of role show that the number of clarification requests between roles is significantly different in both the familiar ($p < 0.001$) and the unfamiliar condition ($p < 0.001$). NHC produced more clarification requests while in the listener role compared to when they were instructors. See **Figure 9**.

There were no significant effects of condition (estimated median difference = 5.33 , $p = 0.244$) or interaction of role \times condition (estimated median difference = -4.5 , $p = 0.388$). For both roles, NHC produced similar numbers of clarification requests in the familiar and unfamiliar conditions. See **Figure 9**.

Research question 3: An effect of CP familiarity for PWA compared to NHC.

Instructor Trials

For the instructor trials the 2 (group: PWA/NHC) \times 2 (condition: familiar/unfamiliar) showed no significant effects for group

⁶The 20% trimmed means analysis did not show a significant effect of role (role: $Q = -11.7$, $p = 0.064$). The variance in the instructor role is close to zero. This will have made the analysis based on the 20% trimmed mean less reliable. We will therefore rely on the outcome of the median analysis here.

⁷The 20% trimmed means analysis did not show a significant main effect of condition ($Q = 4.33$, $p = 0.159$). The same reasoning applies as discussed in footnote 5.

⁸The 20% trimmed means analysis did not show a significant interaction of role \times condition ($Q = -2.67$, $p = 0.112$). The same reasoning applies as discussed in footnote 5.

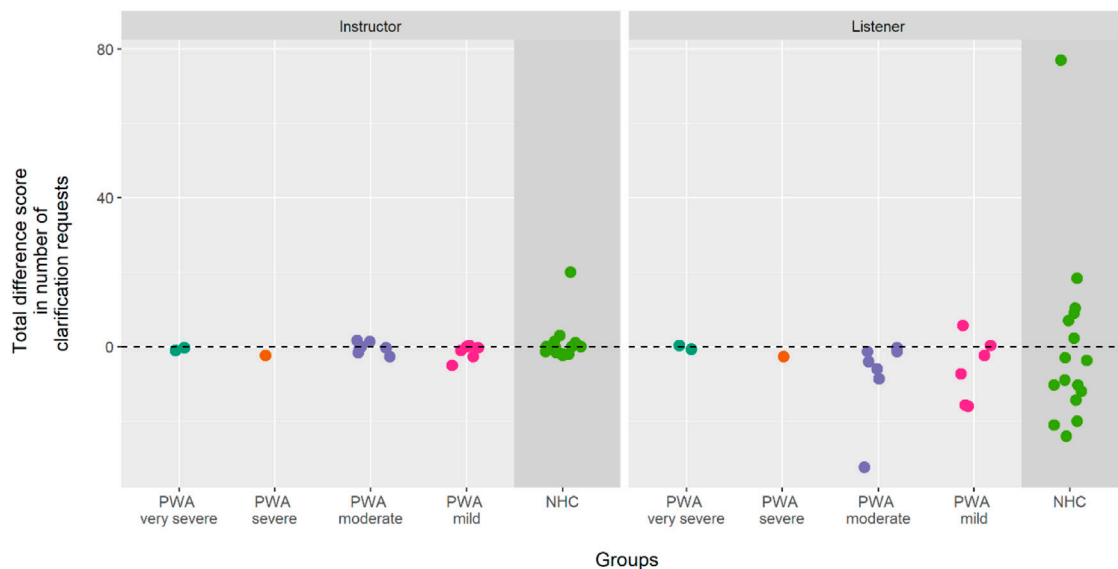


FIGURE 10 | Plot showing individual data points for PWA for difference score between familiar and unfamiliar conditions, by role, categorized by WAB categorization. Zero represents no change in the number of clarification requests between conditions, negative values indicate a smaller number of clarification requests in the unfamiliar condition compared to the familiar condition.

(estimated median difference = -0.17 , $p = 0.657$), condition (estimated median difference = 0.33 , $p = 0.324$)⁹ or the interaction group*condition (estimated median difference = -0.33 , $p = 0.432$).

Listener Trials

The 2 (group: PWA/NHC) \times 2 (condition: familiar/unfamiliar) analysis showed a main effect of group (PWA vs. NHC; estimated median difference = 12.5 , $p = 0.001$), with NHC producing a larger number of clarification requests (median = 18.17 , $CI = 13.58, 28.17$) compared to PWA (median = 6.17 , $CI = 3.0, 12.67$). Planned pairwise comparisons indicated that a significant difference between the two groups existed in both conditions (*familiar*: $p = .032$ ¹⁰, *unfamiliar*: $p < 0.001$). As listeners, NHC showed a higher number of clarification requests compared to PWA in both conditions. See **Figure 9**.

The effect of condition and the interaction were not significant (*condition*: estimated median difference = 3.33 , $p = 0.156$, *interaction*: estimated median difference = 3.83 , $p = 0.454$)¹¹.

⁹The main effect of condition was significant based on the 20% trimmed mean analysis ($Q = 4.74$, $Q_{crit} = 4.38$, $p = 0.042$). The variance for the groups will have been close to zero, which will have made the trimmed means analysis less reliable. We will therefore rely on the median analysis here.

¹⁰In the familiar condition, the trimmed mean analysis showed an insignificant difference between the two groups ($p = 0.755$). Again, the presence of multiple outliers will have inflated the trimmed mean for the PWA group, making the trimmed mean analysis less reliable.

¹¹The main effect of condition was just significant based on the 20% trimmed mean analysis ($Q = 4.29$, $Q_{crit} = 4.27$, $p = 0.049$). As for the instructor trials, the presence of a large number of outliers will probably have inflated the trimmed mean analysis more than the median analysis. To be on the safe side, we will again rely on the more conservative median analysis.

Supplementary Figure S5 shows the changes in the number of clarification requests for each group, condition and role by trial. This figure reflects some potential differences in the number of clarification requests between the final trial of the familiar condition and the first trial of the unfamiliar condition for both groups. The reduction in clarification requests from familiar to unfamiliar conditions may be driven by practice effects, rather than the familiarity of the CP.

Summary of Results for Number of Clarification Requests

The number of clarification requests depended on the role the participants took on: both PWA and NHC asked their conversation partner for clarification more often as listeners compared to when they were instructors. Overall, PWA asked their conversation partner for clarification less often compared to NHC. These effects survived the removal of participants with low MOCA scores. As listeners, PWA asked for clarification less often when working with their unfamiliar conversation partner compared to a familiar conversation partner. In the listener role, NHC did not show a change in number of clarification requests between conditions.

Changes at the Level of Individual Dyads

The changes in number of clarification requests for each dyad are shown in **Figure 10**. Data for PWA has been grouped according the severity of aphasia for the PWA participant. For the instructor role, the change in number of clarification requests was minimal for both groups, and the pattern seems roughly the same across all aphasia severities and groups. In the listener role, there is a trend that as aphasia severity decreases,

TABLE 5 | Results for research questions two and three by outcome measure.

Outcome measure	Description	Role	RQ 2		RQ 3	
			PWA	PWA vs NHC*	Direction main effect of condition **	
					PWA	NHC
Trial time	Faster times indicate "better" communication	Instructor	Fam. > unfam	NHC = PWA	Fam. > unfam	Fam. = unfam
		Listener	Fam. = unfam		Fam. = unfam	Fam. = unfam
Task accuracy	Higher accuracy indicates "better" communication	Instructor	Fam. = unfam	NHC = PWA	Fam. = unfam	Fam. < unfam
		Listener	Fam. < unfam		Fam. < unfam	Fam. = unfam
Self-initiated repairs	Fewer repairs indicate "better" communication	Instructor	Fam. = unfam	NHC = PWA	Fam. = unfam	Fam. = unfam
		Listener	Fam. = unfam		Fam. = unfam	Fam. = unfam
Clarification requests	Fewer requests indicate "better" communication	Instructor	Fam. > unfam	NHC = PWA	Fam. > unfam	Fam. = unfam
		Listener	Fam. > unfam		Fam. > unfam	Fam. = unfam

Notes: Red indicates the outcome is different from the original hypothesis. * hypotheses were about the difference scores between the familiar and unfamiliar conditions. A larger difference score represents a bigger impact of the experimental manipulation. ** in these columns, red indicates a different directional effect in response to the experimental manipulation for PWA compared to NHC.

the distribution of difference scores increases with more dyads showing lower numbers of clarification requests in the unfamiliar condition (negative values). Overall, even the milder severities mostly show more variation in terms of reduction in clarification requests with the UFCP compared to the FCP. NHC show a slightly more equal distribution between decrease and increase in number of clarification requests. These effects are confounded by the uneven spread of data points across aphasia severities.

DISCUSSION

This study examined the effect of conversation partner familiarity on goal-directed, face-to-face communication in aphasia, as part of the contextual component of a theoretical framework of real-world communication. We addressed three research questions.

Research question 1: Is there an effect of role (instructor or listener) during goal-directed communication on the collaborative communication task?

We hypothesized that the type of role (instructor/listener) would affect the outcome measure differently for each group. We predicted that role would have an impact on trial time and accuracy for PWA, but not for NHC. For both groups, we expected an effect of role on the number of self-initiated repairs and clarification requests, due to the nature of these communicative behaviors.

The omnibus analysis showed that overall, NHC showed longer total trial times compared to PWA. There was a significant effect of role for PWA: in the instructor role, PWA took longer to complete a trial compared to when they were in the listener role. For NHC, total trial time was stable across roles.

Overall, PWA obtained lower accuracy scores compared to NHC. For both PWA and NHC, accuracy scores did not significantly differ by role. Planned comparisons on the main effect of condition did show a different pattern of change between the familiar and unfamiliar conditions across the two roles for PWA, which will be discussed in the next section.

The number of self-initiated repairs showed the expected main effect of role: both groups initiated more self-repairs as instructors compared to when they were listeners. Overall, both groups showed equal numbers of self-initiated repairs in the instructor role, while PWA produced fewer repairs compared to NHC in the listener role.

The number of clarification requests also showed the expected main effect of role for both groups. These requests were more frequent in the listener role compared to the instructor role. As listeners, NHC produced more clarification requests compared to PWA.

Overall, these results show that the role participants take on during the task affected the process of goal-directed communication. This is true for PWA on all measures except accuracy. In line with our expectations, role only impacted communication for NHC on the measures of self-initiated repairs and clarification requests.

Research question 2: Do PWA benefit from the familiarity of their conversation partner (personal common ground) during goal directed communication?

For each outcome measure, we tested the hypothesis that it would be easier for PWA to complete the collaborative task with a familiar CP than with an unfamiliar CP. Easier is characterized by the need for less time to complete the task, higher accuracy scores and requiring fewer self-initiated repairs and fewer requests for clarification to reach mutual understanding. The lack of counterbalancing in the design of the current study means that the unfamiliar condition was always presented after the familiar condition. We therefore have to assume that a practice effect is present in the unfamiliar condition. The conclusions we can draw in terms of causality are therefore limited, and we note that omnibus familiarity effects for Total Trial Time and Clarification Requests were no longer significant when participants with low MOCA scores were removed.

The differences between the familiar and unfamiliar condition went against our initial predictions (see **Table 5**). PWA showed shorter total trial times for the unfamiliar condition, higher accuracy for the unfamiliar condition (especially with PWA as listeners) and fewer clarification requests in the unfamiliar condition.

Despite the lack of “familiarity advantage”, it is of interest to note that none of the outcome measures show a change in the “negative” direction during communication with the UFCP (i.e. “worse” communication as evidenced by longer trial times, lower accuracy scores, higher number of self-initiated repairs and clarification requests) as a result of the familiarity manipulation. We expect this to be, at least in part, due to the lack of counterbalancing of conditions, as the unfamiliar condition always came second. If we assume that the familiar condition acted as a practice run, the results suggest that as a group, PWA dyads can show a practice effect (i.e. learning) on a communicative task. Furthermore, on a familiar, practised, concrete task, the communicative ability of PWA dyads are not negatively affected by the lack of personal common ground with their CP during goal-directed communication.

Research question 3: Do PWA differ from NHC in how they respond to conversation partner familiarity?

Finally, we tested whether PWA differ from NHC in how they respond to CP familiarity during goal directed communication. We hypothesized that PWA and NHC would show an overall similar response to the familiarity manipulation on all outcome measures, but that the effect of the experimental manipulation would be greater for PWA compared to NHC, as evidenced by an interaction effect in the group*condition analysis. Results showed no significant interaction effects for any of the outcome measures. When each group was assessed separately for an effect of role and condition, a difference across the familiar and unfamiliar conditions did emerge (see **Table 5**). Due to the experimental design we, again, assume that both groups benefitted from a practice effect in the unfamiliar condition. However, the comparison between performance of both groups in the unfamiliar condition is possible because the practice effect is present for both NHC and PWA.

A comparison of the two groups by role shows that for most outcome measures (five out of eight), PWA and NHC show a different directional response to the change in CP familiarity. NHC showed a stable profile of communicative behavior across the two conditions, apart from an improvement in communicative performance (accuracy scores) as an instructor with an UFCP, which may have come from the practice effect of having the familiar CP condition first. NHC, therefore, generally did not show an effect of CP familiarity in their communicative behavior, nor a significant influence of practice.

In contrast to this, PWA showed a change in communicative behavior between the two conditions as an instructor (time and number of clarification requests) and as a listener (number of clarification requests). As listeners, communicative performance (accuracy) is also affected. In short, PWA show a more widespread change in communicative behavior and performance as a result of the familiarity manipulation compared to NHC. These differences are discussed below.

Familiarity Effect in Aphasia and NHC Instructors

We found that as instructors, PWA showed a different pattern of behavior when working with a FCP compared to an UFCP (shorter trial times, fewer clarification requests with the UFCP, and stable accuracy scores and self-initiated repairs). The stability to the number of self-initiated repairs is in line with

previous studies that have suggested that certain aspects of communication might remain stable across different communicative settings and CPs (Lubinski et al., 1980; Gurland et al., 1982; Leaman and Edmonds, 2019). The higher number of clarification requests with the FCP is also in line with previous research with NHC (Boyle et al., 1994). As suggested by the authors, the unfamiliarity might have discouraged PWA from asking UFCPs for clarification more often. In addition, the experience PWA had gained on the task by the time they worked with the UFCP, could have meant that fewer clarification requests were needed. The stability of the accuracy scores across familiar and unfamiliar CPs, and the reduction in trial time with the UFCP compared to the FCP, suggest that the ability to complete the task in less time with the unfamiliar CP was a result of increased experience and confidence on the task. With the UFCP, PWA were able to achieve the same result (i.e. stable accuracy scores), while putting in less “effort” (i.e. time and number of clarification requests). Differently put, PWA might have been more ‘efficient’ at completing the task with the UFCP compared to the FCP, possibly due to greater experience on the task in the unfamiliar condition. In contrast to this, NHC were shown to put in the same amount of effort (i.e. time, repairs and clarification requests) with both CPs, which resulted in a better outcome with the UFCP (i.e. higher accuracy scores). While both groups had the same amount of practice on the task, a different pattern of behavior is observed in the unfamiliar condition.

There are a number of possible reasons for this difference in effort. Firstly, perhaps PWA felt more comfortable with their FCP compared to the UFCP, resulting in more time and effort spent with the FCP. In line with this, PWA might have felt more comfortable asking for clarification from the FCP compared to the UFCP. The results from our measure of comfort with the CP indicate that at least at the group level, this explanation doesn’t hold, as PWA reported the same level of comfort with both CPs. Another explanation for the reduced time and number of clarification requests is that familiarity of the task reduced the need for more time. The stability of the accuracy scores for PWA, while NHC still improved in the unfamiliar condition (showing a likely practice effect) is perhaps more surprising. It is possible that in the instructor role, PWA dyads reached a ceiling for accuracy and might not have been able to communicate more detail on the task to their CP, even with practice.

Finally, it is possible that as instructors, PWA and NHC differed (consciously or unconsciously) in the criterion they set for achieving mutual understanding. To communicate, interlocutors must continuously achieve mutual understanding together, i.e. they must understand what the other person is saying to continue the conversation (Clark and Wilkes-Gibbs, 1986; Clark and Brennan, 1991; Clark, 1996). Mutual understanding does not have to be perfect for conversation to work. Instead, interlocutors negotiate a criterion of mutual understanding “well enough for current purposes” (Clark, 1996, p. 221). NHC, unrestricted by any communication difficulties, might have set a higher criterion for mutual understanding on the current task (i.e. striving for a higher level of accuracy). This then resulted in similar amounts of effort made in an attempt to achieve higher accuracy scores, regardless of CP familiarity.

For PWA, this process might have unfolded differently. When confronted with the UFCP, PWA might have accepted the level of mutual understanding they had been able to achieve so far (with their FCP) as good enough for current purposes. This might have allowed PWA to strip away any communicative behaviors deemed unnecessary for current purposes (i.e. fewer clarification requests and less time). We can only speculate about the underlying reasons for such a shift. It could have been the desire to avoid unnecessary conversational difficulties (or: avoid “losing face”) with the UFCP (as evidenced by fewer clarification requests initiated by the PWA in the unfamiliar condition) (Simmons-Mackie and Damico, 1995). It could also be that regardless of the CP, PWA tend to strive to minimize communicative (cognitive) effort in light of the good enough accuracy scores more generally.

Listeners

The changes in the number of self-initiated repairs and clarification requests were in line with previous research, as discussed for the instructor role. The increase in accuracy scores with the UFCP, and the stable trial times across CPs go against our predictions and indicates the presence of a practice effect. The NHC group will be used as a reference in the discussion of current findings for PWA.

It seems that as listeners, PWA put in the same amount of “effort” in both conditions (as measured by total trial time), while achieving a better result with the UFCP (i.e. higher accuracy scores). NHC show the same pattern in trial time, but their accuracy scores remain stable. For NHC, this might reflect a ceiling effect rather than a strong behavioral pattern.

The most likely explanation in our view is that PWA benefitted from repeated practice on the task, resulting in better performance on the second half of the trials. Completing the same task with the same set of stimuli a number of times might have created a physical and communicative context (i.e. things that have been discussed within the same conversation become part of common ground) that could have helped restrict the number of possible interpretations for PWA (Skipper, 2014; Doedens and Meteyard, 2019).

Interestingly, while PWA showed shorter trial times with the UFCP when they were instructors, this effect disappeared when they were in the listener role. A potential explanation for this is that those who take on the instructor role are more in control of the way the trial unfolds over time. This would explain why the reduced trial time when PWA are listeners disappears: their CP might have taken the lead, resulting in similar patterns of “effort” as compared to the NHC group and no reduction in overall trial time. Further assessment of the CP role is needed to confirm this interpretation, however. An analysis as reported in this paper, conducted on data from the conversation partners of each PWA when they were in the instructor role, for example, could reveal whether they show a pattern of “effort” across conditions that is similar to NHC or not. Furthermore, insight into the number of turns taken, or the duration of turns for each CP (PWA and their familiar and unfamiliar CPs) could provide more detailed insight into the efforts made by both parties during the task, and how this changed (or not) as a result of the familiarity manipulation.

Aphasia Severity

The inspection of the difference scores on all outcome measures between the familiar and unfamiliar conditions allows us to draw tentative conclusions about the difference in behavioral patterns depending on aphasia severity. Visual inspection of the data shows the tendency for PWA with milder severity to show greater behavioral change as a result of CP familiarity. As might be expected, as aphasia severity decreases the behavioral pattern becomes more like that of the NHC group. Although more research is needed with a larger group of people with severe aphasia, an intuitive interpretation is that less flexibility in communicative behavior is seen for PWA with more severe aphasia, as they have less scope for flexible communication in the first place. More research is needed with a larger group of PWA, divided equally across severities, to draw stronger conclusions about this.

Finally, a limitation to the current study is the fact that a number of the participants scored below the cut-off score on the MoCA, suggesting the potential presence of mild cognitive impairment but perhaps typical for older dyads as sampled here. Previous studies have shown that the presence of mild cognitive impairment can influence performance on a referential communication task, due to impairments in cognitive functioning or impairments in Theory of Mind (Moreau et al., 2015; Moreau et al., 2016). The secondary analyses, excluding the data from the dyads these participants belonged to, showed that main effects of familiarity for Total Trial Times and Clarification requests were no longer significant. Effects of familiarity were already confounded with practice, making it difficult to draw strong conclusions. However, future research should examine the influence of potential cognitive and theory of mind deficits on performance on this task and communication more closely, especially in relation to the older participants, and the ability of conversation partners to provide optimal and flexible communicative support to their conversation partners with aphasia.

CONCLUSION

When communicating about a concrete, practised topic, PWA dyads do not show the often-assumed negative influence of a lack of shared personal common ground. Furthermore, the current results seem to suggest that PWA might be able to carry over the experience on a communicative task across conversation partners. More research is needed, however, to confirm this. It may be the case that in a more complex or abstract task, partner familiarity will have a greater impact on performance for PWA (Fussell and Krauss, 1989).

We found tentative evidence that PWA showed a different response to the presence of an unfamiliar conversation partner compared to NHC (where both groups had the same practice). Based on the current findings, it seems PWA aim to reduce communicative efforts in order to achieve good enough information transfer. This seems specifically the case when PWA are in the “instructor” role. In the listener role, it seems PWA might benefit from the repeated practice on the same task, i.e. building up of common ground within the task, as evidenced by their improved accuracy across conditions. In contrast to PWA, NHC show similar

communicative behaviors across conversation partners. This group seems to strive for the most detailed information exchange, regardless of the familiarity of the CP. In the case of NHC, an improvement in performance suggests NHC might benefit from a building up of experience, or common ground, within the task, regardless of the familiarity of their conversation partner. Especially considering that this task used highly concrete materials that the NHC should have found easy to describe. More research is needed to evaluate the effect of conversation partner familiarity on communicative behaviors and performance in PWA on, for example, an unfamiliar or more complex task. In such a case, the tendency of PWA to minimize communicative efforts with the unfamiliar conversation partner, without having had any practice, could potentially lead to lower performance scores.

CLINICAL IMPLICATIONS

The findings from the current study have clinical implications for treatment and assessment in aphasia rehabilitation. The current study partly supports the existing assumption that conversation partner familiarity affects communication for PWA. Importantly, the outcome on the current task was not negatively affected by the presence of an unfamiliar CP, as shown by equal or improved communicative performance on the task with the unfamiliar conversation partner. We assume these results are at least partly due to a practice effect. However, a positive effect of practice for PWA on a goal-directed communication task, in many ways similar to a setup like PACE (Davis, 2005) for intervention, is something to be celebrated. This research shows that PWA can show different communicative behaviors and communicative purposes, depending on the conversation partner they are communicating with (Simmons-Mackie and Damico, 1995). These findings also have implications for the way communicative behaviors that have been trained in one setting, might generalize (or not) across conversation partners. The results also suggest that PWA with more severe aphasias might be less flexible in adapting to different communicative settings (and therefore might require training on a more generic set of communicative strategies, that work across communication settings and partners). The lower MoCA scores for some CPs also suggest that the ability of the CP to flexibly support and enable the PWA to communicate effectively should be considered during intervention. Although the underlying reasons for the change in communicative behaviors between conversation partners remain unclear, this is important to keep in mind when profiling real-world communicative abilities for PWA.

REFERENCES

- Andersson, J., and Ronnberg, J. (1997). Cued Memory Collaboration: Effects of Friendship and Type of Retrieval Cue. *Eur. J. Cogn. Psychol.* 9, 273–287. doi:10.1080/713752558
- Armstrong, E., Ciccone, N., Godecke, E., and Kok, B. (2011). Monologues and Dialogues in Aphasia: Some Initial Comparisons. *Aphasiology* 25, 1347–1371. doi:10.1080/02687038.2011.577204

DATA AVAILABILITY STATEMENT

The raw data presented in this article are not readily available because of the sensitivity of the video materials. Requests to access the anonymized datasets should be directed to the first author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the PCLS School Ethics Research Committee, School of Psychology and Clinical Languages Sciences, University of Reading. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conceptualisation: WD, AB, LM; Data collection: WD; Data coding: WD, LL, Writing: WD.; Critical review: AB, LM.

FUNDING

The first author was supported by the Magdalen Vernon Studentship from the School of Psychology and Clinical Language Sciences, University of Reading. Recruitment was supported by the Aphasia Research Registry of the School of Clinical Language Sciences, University of Reading (British Academy Grant ARP scheme 190023).

ACKNOWLEDGMENTS

The authors would like to thank Prof. Katerina Hilari at City University London for the use of the English version of the Scenario Test. We would also like to thank Odette Stoutjesdijk, aphasia therapist at the Utrecht Aphasia Center in the Netherlands, for her support. We would also like to thank the reviewers for their insightful feedback. Finally, we would like to thank Lucy Letzer for designing Figure 1.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2021.574051/full#supplementary-material>

- Bard, E. G., Hill, R. L., Foster, M. E., and Arai, M. (2014). Tuning Accessibility of Referring Expressions in Situated Dialogue. *Lang. Cogn. Neurosci.* 29, 928–949. doi:10.1080/23273798.2014.895845
- Beeke, S. (2012). “*Cognitive Pragmatics*”. Editor H.-J. Schmid (Berlin: Mouton de Gruyter).
- Beeke, S., Maxim, J., and Wilkinson, R. (2007). Using Conversation Analysis to Assess and Treat People with Aphasia. *Semin. Speech Lang.* 28, 136–147. doi:10.1055/s-2007-970571

- Boyle, E. A., Anderson, A. H., and Newlands, A. (1994). The Effects of Visibility on Dialogue and Performance in a Cooperative Problem Solving Task. *Lang. Speech* 37, 1–20. doi:10.1177/002383099403700101
- Brady, M. C., Godwin, J., Enderby, P., Kelly, H., and Campbell, P. (2016). Speech and Language Therapy for Aphasia after Stroke. *Stroke* 47, E236–E237. doi:10.1161/Strokeaha.116.014439
- Brown-Schmidt, S. (2009a). Partner-specific Interpretation of Maintained Referential Precedents during Interactive Dialog. *J. Mem. Lang.* 61, 171–190. doi:10.1016/j.jml.2009.04.003
- Brown-Schmidt, S. (2009b). The Role of Executive Function in Perspective Taking during Online Language Comprehension. *Psychon. Bull. Rev.* 16, 893–900. doi:10.3758/pbr.16.5.893
- Charmant, J., and Contrib (2006–2011). *Kinovea*. Retrieved from www.kinovea.org.
- Clark, H., and Brennan, S. E. (1991). “Grounding in Communication,” in *Perspectives on Socially Shared Cognition*. Editors L. B. Resnick, J. M. Levine, and S. D. Teasley (Washington: American Psychological Association).
- Clark, H. H., and Krych, M. A. (2004). Speaking while Monitoring Addressees for Understanding. *J. Mem. Lang.* 50, 62–81. doi:10.1016/j.jml.2003.08.004
- Clark, H. H., and Schaefer, E. F. (1987). Collaborating on Contributions to Conversations. *Lang. Cogn. Process.* 2, 19–41. doi:10.1080/01690968708406350
- Clark, H. H., and Wilkes-Gibbs, D. (1986). Referring as a Collaborative Process. *Cognition*, 22, 1–39. doi:10.1016/0010-0277(86)90010-7
- Clark, H. (1996). *Using Language*. Cambridge, UK: University Press. doi:10.1017/cbo9780511620539
- Dalemans, R. J. P., de Witte, L., Wade, D., and van den Heuvel, W. (2010). Social Participation through the Eyes of People with Aphasia. *Int. J. Lang. Commun. Disord.* 45, 537–550. doi:10.3109/13682820903223633
- Davidson, B., Worrall, L., and Hickson, L. (2008). Exploring the Interactional Dimension of Social Communication: A Collective Case Study of Older People with Aphasia. *Aphasiology* 22, 235–257. doi:10.1080/02687030701268024
- Davis, G. A. (2005). PACE Revisited. *Aphasiology* 19, 21–38. doi:10.1080/0268703044000598
- Davis, G. A., and Wilcox, M. J. (1985). *Adult Aphasia Rehabilitation: Applied Pragmatics*. San, Diego, CA: College-Hill Press.
- Dixon, R. A., Gagnon, L. M., and Crow, C. B. (1998). “Memory Distortions and Their Prevention,”. Editors M. J. Intons-Peterson and D. L. Best (Lawrence Erlbaum Associates Publishers).
- Doedens, W. J., and Meteyard, L. (2019). Measures of Functional, Real-World Communication for Aphasia: a Critical Review. *Aphasiology* 34, 492–514. doi:10.1080/02687038.2019.1702848
- Ferguson, A. (1994). The Influence of Aphasia, Familiarity and Activity on Conversational Repair. *Aphasiology*, 8, 143–157. doi:10.1080/02687039408248647
- Fussell, S. R., and Krauss, R. M. (1989). Understanding Friends and Strangers: The Effects of Audience Design on Message Comprehension. *Eur. J. Soc. Psychol.* 19, 509–525. doi:10.1002/ejsp.2420190603
- Gould, O. N., Osborn, C., Krein, H., and Mortenson, M. (2002). Collaborative Recall in Married and Unacquainted Dyads. *Int. J. Behav. Develop.* 26, 36–44. doi:10.1080/01650250143000292
- Green, G. (1982). Assessment and Treatment of the Adult with Severe Aphasia: Aiming for Functional Generalisation. *Aust. J. Hum. Commun. Disord.* 10, 11–23. doi:10.3109/asl2.1982.10.issue-1.02
- Gurland, G. B., Chwat, S. E., and Gerber Wollner, S. (1982). Establishing a Communication Profile in Adult Aphasia: Analysis of Communicative Acts and Conversational Sequences.
- Hachoui, H. E., Visch-Brink, E. G., Lingsma, H. F., Sandt-Koenderman, M. W., Dippel, D. W., Koudstaal, P. J., et al. (2014). Nonlinguistic Cognitive Impairment in Poststroke Aphasia: a Prospective Study. *Neurorehabil. Neural Repair* 28, 273–281. doi:10.1177/1545968313508467
- Harmon, T. G. (2020). Everyday Communication Challenges in Aphasia: Descriptions of Experiences and Coping Strategies. *Aphasiology* 34, 1270–1290. doi:10.1080/02687038.2020.1752906
- Herrmann, T. (1983). *Speech and Situation: A Psychological Conception of Situated Speaking*. Berlin/Heidelberg: Springer-Verlag. doi:10.1007/978-3-642-61790-4
- Hilari, K., Galante, L., Huck, A., Pritchard, M., Allen, L., and Dipper, L. (2018). Cultural Adaptation and Psychometric Testing of the Scenario Test UK for People with Aphasia. *Int. J. Lang. Commun. Disord.* 53, 748–760. doi:10.1111/1460-6984.12379
- Hilari, K., Wiggins, R., Roy, P., Byng, S., and Smith, S. (2003). Predictors of Health-Related Quality of Life (HRQL) in People with Chronic Aphasia. *Aphasiology* 17, 365–381. doi:10.1080/02687030244000725
- Holland, A. L. (1982). Observing Functional Communication of Aphasic Adults. *J. Speech Hear. Disord.* 47, 50–56. doi:10.1044/jshd.4701.50
- Hornstein, G. A. (1985). Intimacy in Conversational Style as a Function of the Degree of Closeness between Members of a Dyad. *J. Personal. Soc. Psychol.* 49, 671–681. doi:10.1037/0022-3514.49.3.671
- Horton, W. S., and Gerrig, R. J. (2005). The Impact of Memory Demands on Audience Design during Language Production. *Cognition* 96, 127–142. doi:10.1016/j.cognition.2004.07.001
- Howard, D., Swinburn, K., and Porter, G. (2004). *Comprehensive Aphasia Test*. Routledge. Psychology Press.
- Howarth, B., and Anderson, A. H. (2007). Introducing Objects in Spoken Dialogue: The Influence of Conversational Setting and Cognitive Load on the Articulation and Use of Referring Expressions. *Lang. Cogn. Process.* 22, 272–296. doi:10.1080/01690960600632796
- Howe, T. J., Worrall, L. E., and Hickson, L. M. H. (2008). Observing People with Aphasia: Environmental Factors that Influence Their Community Participation. *Aphasiology* 22, 618–643. doi:10.1080/02687030701536024
- Isaacs, E. A., and Clark, H. H. (1987). References in Conversation between Experts and Novices. *J. Exp. Psychol. Gen.* 116, 26–37. doi:10.1037/0096-3445.116.1.26
- Kagan, A., Black, S. E., Duchan, J. F., Simmons-Mackie, N., and Square, P. (2001). Training Volunteers as Conversation Partners Using “Supported Conversation for Adults with Aphasia” (SCA). *J. Speech Lang. Hear. Res.* 44, 624–638. doi:10.1044/1092-4388(2001/051)
- Kertesz, A. (2009). *Western Aphasia Battery - Revised*. Pearson Education.
- Kistner, J. (2017). The Use of Gestures in the Conversations of People with aphasia. PhD Thesis, City: London: University of London. doi:10.1055/s-0037-1604632
- Kolk, H., and Heeschen, C. (1992). Agrammatism, Paragrammatism and the Management of Language. *Lang. Cogn. Process.* 7, 89–129. doi:10.1080/01690969208409381
- Laakso, M., and Godt, S. (2016). Recipient Participation in Conversations Involving Participants with Fluent or Non-fluent Aphasia. *Clin. Linguistics Phonetics* 30, 770–789. doi:10.1080/02699206.2016.1221997
- Lam, J. M. C., and Wodchis, W. P. (2010). The Relationship of 60 Disease Diagnoses and 15 Conditions to Preference-Based Health-Related Quality of Life in Ontario Hospital-Based Long-Term Care Residents. *Med. Care* 48, 380–387. doi:10.1097/mlr.0b013e3181ca2647
- Leaman, M. C., and Edmonds, L. A. (2019). Conversation in Aphasia across Communication Partners: Exploring Stability of Microlinguistic Measures and Communicative Success. *Am. J. Speech Lang. Pathol.* 28, 359–372. doi:10.1044/2018_ajslp-17-0148
- Li, E. C., Williams, S. E., and Volpe, A. D. (1995). The Effects of Topic and Listener Familiarity on Discourse Variables in Procedural and Narrative Discourse Tasks. *J. Commun. Disord.* 28, 39–55. doi:10.1016/0021-9924(95)91023-z
- Lindsay, J., and Wilkinson, R. (1999). Repair Sequences in Aphasic Talk: a Comparison of Aphasic-Speech and Language Therapist and Aphasic-Spouse Conversations. *Aphasiology* 13, 305–325. doi:10.1080/026870399402118
- Lubinski, R., Duchan, J., and Weitzner-Lin, B. (1980). *Analysis of Breakdowns and Repairs in Aphasic Adult Communication*. Clinical Aphasiology Paper.
- Lysander, K., and Horton, W. S. (2012). Conversational Grounding in Younger and Older Adults: The Effect of Partner Visibility and Referent Abstractness in Task-Oriented Dialogue. *Discourse Process.* 49, 29–60. doi:10.1080/0163853x.2011.625547
- Mead, G. H. (1934). *Mind, Self, and Society from the Standpoint of a Social Behaviorist*. Chicago: University of Chicago Press.
- Meulen, I., Sandt-Koenderman, W. M., Duivenvoorden, H. J., and Ribbers, G. M. (2010). Measuring Verbal and Non-verbal Communication in Aphasia: Reliability, Validity, and Sensitivity to Change of the Scenario Test. *Int. J. Lang. Commun. Disord.* 45, 424–435. doi:10.3109/13682820903111952
- Moreau, N., Rauzy, S., Bonnefoi, B., Renié, L., Martinez-Almoyna, L., Viallet, F., et al. (2015). Different Patterns of Theory of Mind Impairment in Mild Cognitive Impairment. *Jad* 45, 581–597. doi:10.3233/JAD-143021

- Moreau, N., Rauzy, S., Viallet, F., and Champagne-Lavau, M. (2016). Theory of Mind in Alzheimer Disease: Evidence of Authentic Impairment during Social Interaction. *Neuropsychology* 30, 312–321. doi:10.1037/neu0000220
- Nasreddine, Z. S., Phillips, N. A., Bäckström, V. r., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool for Mild Cognitive Impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi:10.1111/j.1532-5415.2005.53221.x
- Nickels, L., and Howard, D. (1995). Phonological Errors in Aphasic Naming: Comprehension, Monitoring and Lexicality. *Cortex* 31, 209–237. doi:10.1016/s0010-9452(13)80360-7
- Nykanen, A., Nyrkko, H., Nykanen, M., Brunou, R., and Rautakoski, P. (2013). Communication Therapy for People with Aphasia and Their Partners (APPUTE). *Aphasiology* 27, 1159–1179. doi:10.1080/02687038.2013.802284
- Perkins, L. (1993). “The Impact of Cognitive Neuropsychological Impairments on Conversational Ability in Aphasia,” (University of Newcastle upon Tyne: Department of Speech, 1993). PhD Thesis.
- Perkins, L. (1995). Applying Conversation Analysis to Aphasia: Clinical Implications and Analytic Issues. *Int. J. Lang. Commun. Disord.* 30, 372–383. doi:10.3109/13682829509021449
- Pollmann, M. M. H., and Finkenauer, C. (2009). Investigating the Role of Two Types of Understanding in Relationship Well-Being: Understanding Is More Important Than Knowledge. *Pers. Soc. Psychol. Bull.* 35, 1512–1527. doi:10.1177/0146167209342754
- Pollmann, M. M. H., and Krahmer, E. J. (2017). How Do Friends and Strangers Play the Game Taboo? A Study of Accuracy, Efficiency, Motivation, and the Use of Shared Knowledge. *J. Lang. Soc. Psychol.* 37, 497–517. doi:10.1177/0261927x17736084
- Pound, C., Parr, S., Lindsay, J., and Woolf, C. (2000). *Beyond Aphasia, Therapies for Living with Communication Disability*. London: Routledge.
- Rayner, H., and Marshall, J. (2003). Training Volunteers as Conversation Partners for People with Aphasia. *Int. J. Lang. Commun. Disord.* 38, 149–164. doi:10.1080/1368282021000060308
- Revelle, W. (2020). *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston. Retrieved from <https://CRAN.R-project.org/package=psych>.
- Riordan, M. A., and Trichtinger, L. A. (2016). Overconfidence at the Keyboard: Confidence and Accuracy in Interpreting Affect in E-Mail Exchanges. *Hum. Commun. Res.* 43, 1–24. doi:10.1111/hcre.12093
- Rousselet, G. A., Pernet, C. R., and Wilcox, R. R. (2019). A Practical Introduction to the Bootstrap: a Versatile Method to Make Inferences by Using Data-Driven Simulations. doi:10.31234/osf.io/h8ft7
- RStudio (2020). *RStudio*. Boston: Integrated Development Environment for; R. Boston. Retrieved from <http://www.rstudio.com/>.
- Savitsky, K., Keysar, B., Epley, N., Carter, T., and Swanson, A. (2011). The Closeness-Communication Bias: Increased Egocentrism Among Friends versus Strangers. *J. Exp. Soc. Psychol.* 47, 269–273. doi:10.1016/j.jesp.2010.09.005
- Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language* 53, 361. doi:10.2307/413107
- Schober, M. F., and Carstensen, L. L. (2010). “Expressing oneself/Expressing One’s Self: Communication, Cognition, Language, and Identity,”. Editor E. Morsella (London, England: Taylor & Francis).
- Sillars, A., Shellen, W., McIntosh, A., and Pomegranate, M. (1997). Relational Characteristics of Language: Elaboration and Differentiation in Marital Conversations. *West. J. Commun.* 61, 403–422. doi:10.1080/10570319709374587
- Simmons-Mackie, N. N., and Damico, J. S. (1995). Communicative Competence in Aphasia: Evidence from Compensatory Strategies. *Clin. aphasiology* 23, 95–105.
- Simmons-Mackie, N., Raymer, A., Armstrong, E., Holland, A., and Cherney, L. R. (2010). Communication Partner Training in Aphasia: a Systematic Review. *Arch. Phys. Med. Rehabil.* 91, 1814–1837. doi:10.1016/j.apmr.2010.08.026
- Skipper, J. I. (2014). Echoes of the Spoken Past: How Auditory Cortex Hears Context during Speech Perception. *Phil. Trans. R. Soc. B* 369, 20130297–20130319. doi:10.1098/rstb.2013.0297
- Smith, S. W., Noda, H. P., Andrews, S., and Jucker, A. H. (2005). Setting the Stage: How Speakers Prepare Listeners for the Introduction of Referents in Dialogues and Monologues. *J. Pragmatics* 37, 1865–1895. doi:10.1016/j.pragma.2005.02.016
- Spaccavento, S., Craca, A., Del Prete, M., Falcone, R., Colucci, A., Di Palma, A., et al. (2013). Quality of Life Measurement and Outcome in Aphasia. *Ndt* 27, 27. doi:10.2147/ndt.s52357
- Swinburn, K., Best, W., Beeke, S., Cruice, M., Smith, L., Pearce Willis, E., et al. (2018). A Concise Patient Reported Outcome Measure for People with Aphasia: the Aphasia Impact Questionnaire 21. *Aphasiology* 33, 1035–1060. doi:10.1080/02687038.2018.1517406
- The Language Archive (2019). Nijmegen: Max Planck Institute for Psycholinguistics, T. L. ELAN (Version 5.7) <https://archive.mpi.nl/ta/elan>. Version 5.7
- Thompson, C. K., Worrall, L., and Martin, N. (2008). Approaches to Aphasia Treatment. In *Aphasia Rehabilitation. The Impairment and its Consequences*, 3–24.
- Wilcox, R. R. (2012). *Introduction to Robust Estimation and Hypothesis Testing*. 3rd ed. Cambridge, MA: Elsevier Science & Technology.
- Wilkinson, R. (2006). Applying Conversation Analysis to Aphasic Talk: From Investigation to Intervention. *Revue française de linguistique appliquée* XI, 99. doi:10.3917/rfla.112.0099
- Wilkinson, R. (1995). *Case Studies in Clinical Linguistics*. London: Whurr.
- Wilkinson, R., and Wielaert, S. (2012). Rehabilitation Targeted at Everyday Communication: Can We Change the Talk of People with Aphasia and Their Significant Others within Conversation?. *Arch. Phys. Med. Rehabil.* 93, S70–S76. doi:10.1016/j.apmr.2011.07.206
- Williams, S. E., Li, E. C., Della Volpe, A., and Ritterman, S. I. (1994). The Influence of Topic and Listener Familiarity on Aphasic Discourse. *J. Commun. Disord.* 27, 207–222. doi:10.1016/0021-9924(94)90001-9
- Wirz, S. L., Skinner, C., and Dean, E. (1990). *Revised Edinburgh Functional Communication Profile*. Tucson, Ariz: Communication Skill Builders. doi:10.4000/couillises.1612
- Worrall, L., Sherratt, S., Rogers, P., Howe, T., Hersch, D., Ferguson, A., et al. (2010). What People with Aphasia Want: Their Goals According to the ICF. *Aphasiology* 25, 309–322. doi:10.1080/02687038.2010.508530
- Wu, S., and Keysar, B. (2007). The Effect of Information Overlap on Communication Effectiveness. *Cogn. Sci.* 31, 169–181. doi:10.1080/03640210709336989
- Zwaan, R. A. (2016). Situation Models, Mental Simulations, and Abstract Concepts in Discourse Comprehension. *Psychon. Bull. Rev.* 23, 1028–1034. doi:10.3758/s13423-015-0864-x

Conflict of Interest: 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 © 2021 Doedens, Bose, Lambert and Meteyard. 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.



Ladies First: Gender Stereotypes Drive Anticipatory Eye-Movements During Incremental Sentence Interpretation

Ernesto Guerra^{1*}, Jasmin Bernotat^{2,3}, Héctor Carvacho⁴ and Gerd Böhner³

¹ Center for Advanced Research in Education, Institute of Education, Universidad de Chile, Santiago, Chile, ² Center for Cognitive Interaction Technology, CITEC, Bielefeld University, Bielefeld, Germany, ³ Department of Psychology, Bielefeld University, Bielefeld, Germany, ⁴ School of Psychology, Pontificia Universidad Católica de Chile, Santiago, Chile

OPEN ACCESS

Edited by:

Ramesh Kumar Mishra,
University of Hyderabad, India

Reviewed by:

Paolo Canal,
University Institute of Higher Studies
in Pavia, Italy
Tandra Ghose,
University of Kaiserslautern, Germany

*Correspondence:

Ernesto Guerra
ernesto.guerra@ciae.uchile.cl

Specialty section:

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

Received: 30 July 2020

Accepted: 27 April 2021

Published: 23 June 2021

Citation:

Guerra E, Bernotat J, Carvacho H
and Böhner G (2021) Ladies First:
Gender Stereotypes Drive
Anticipatory Eye-Movements During
Incremental Sentence Interpretation.
Front. Psychol. 12:589429.
doi: 10.3389/fpsyg.2021.589429

Immediate contextual information and world knowledge allow comprehenders to anticipate incoming language in real time. The cognitive mechanisms that underlie such behavior are, however, still only partially understood. We examined the novel idea that gender attitudes may influence how people make predictions during sentence processing. To this end, we conducted an eye-tracking experiment where participants listened to passive-voice sentences expressing gender-stereotypical actions (e.g., “The wood is being painted by the florist”) while observing displays containing both female and male characters representing gender-stereotypical professions (e.g., florists, soldiers). In addition, we assessed participants’ explicit gender-related attitudes to explore whether they might predict potential effects of gender-stereotypical information on anticipatory eye movements. The observed gaze pattern reflected that participants used gendered information to predict who was agent of the action. These effects were larger for female- vs. male-stereotypical contextual information but were not related to participants’ gender-related attitudes. Our results showed that predictive language processing can be moderated by gender stereotypes, and that anticipation is stronger for female (vs. male) depicted characters. Further research should test the direct relation between gender-stereotypical sentence processing and implicit gender attitudes. These findings contribute to both social psychology and psycholinguistics research, as they extend our understanding of stereotype processing in multimodal contexts and regarding the role of attitudes (on top of world knowledge) in language prediction.

Keywords: gender stereotypes, language comprehension, anticipatory eye movements, explicit beliefs, eye tracking

INTRODUCTION

Humans frequently generate expectations about what will happen in the near future (e.g., thinking of tonight’s dinner) or even in the immediate subsequent moment. When processing language, listeners are often capable to predict the word upcoming in the next seconds or even milliseconds. Indeed, anticipatory processing (or prediction) in language and cognition has received special attention in recent years, partly as a consequence of extensive experimental evidence showing that

listeners and readers can make online language predictions (see DeLong et al., 2014; Huettig, 2015 for reviews), but also triggered by a heated debate around mounting inconsistent evidence on the pervasiveness of this phenomenon (DeLong et al., 2017; Nieuwland et al., 2018; Huettig and Guerra, 2019; Nicenboim et al., 2020). Thus, while some level of prediction appears to occur, it is much less clear which cognitive mechanisms support this behavior and what is its role for language and cognition (see Friston, 2010; Hauk, 2016).

Experimental studies on language-mediated visual attention have shown that during spoken sentence comprehension, prediction can be triggered by unfolding linguistic cues and knowledge about the world, allowing comprehenders to anticipate to-be-mentioned visual referents in real time (see Knoeferle and Crocker, 2007; Borovsky et al., 2012). Interestingly, other studies have shown that listeners can rapidly make inferences based on speakers' voices and related social stereotypes, such that comprehension is more difficult when subsequent language does not match the activated stereotype (van Berkum et al., 2008). Research suggests that information about people is frequently processed using stereotypes about social groups. For instance, people might attribute certain characteristics as intrinsic to a particular gender, e.g., women have lower mathematics abilities than men (Spencer et al., 1999; Dovidio et al., 2005; Jost and Kay, 2005).

Indeed, stereotypes about gender do affect sentence comprehension. Carreiras et al. (1996) asked participants to read sentence pairs such as "The footballer wanted to play in the match. He had been training very hard during the week." In their study, the gender pronoun was manipulated to match (e.g., He) vs. mismatch (e.g., She) the gender stereotype associated with the role name (e.g., footballer), which resulted in faster vs. slower reading times, respectively (see also Duffy and Keir, 2004; Kreiner et al., 2008). These findings have been interpreted as suggesting that gender-stereotypical information triggers inferences based on people's world knowledge (e.g., Kreiner et al., 2008). However, research in social psychology suggests otherwise: Although stereotypes constitute basic cognitive structures for categorization, they are in fact culturally transmitted and thus they reflect social biases rather than plausibility (see Brown, 2011).

Previous research on language prediction has certainly shown the relevance of real-world plausibility knowledge. In a study by Kamide et al. (2003), participants saw a visual display with two characters (e.g., tailor, plumber) and four objects (e.g., sewing machine, fabric, sink, and pipe). Participants' eye movements were recorded as they inspected the displays and listened to recorded materials. Spoken sentences such as "The tailor will cut the fabric" would allow precise prediction of the ensuing object only if participants were able to use their world knowledge (about tailors) in addition to the lexical information provided by the verb (cf. Altmann and Kamide, 1999). Results showed that participants anticipated *fabric* instead of *pipe* when hearing "The tailor will cut the. ..." In turn, they preferred to look at the *pipe* upon hearing "The plumber will repair the. ..."

Thus, comprehenders can combine long-term memory representations (world knowledge) with incoming lexical (e.g.,

the verb) and visual (e.g., referents) inputs in real time to make predictions about what will be mentioned next (see also Knoeferle and Crocker, 2007; van Berkum et al., 2008; Borovsky et al., 2012). With sentences like those from Kamide et al. (2003), the top-down influence of world knowledge on verbal and visual information processing provides the comprehenders with sufficient constraints to unequivocally anticipate the correct object, both when the tailor cuts the fabric and when the plumber repairs the pipe. In such a case, certainly, comprehenders retrieve world knowledge about how occupations are associated with particular actions.

Research on the role of gender in language, however, has been somewhat different. First, although previous studies (e.g., Carreiras et al., 1996; Duffy and Keir, 2004; Kreiner et al., 2008) have looked into how gender-associated professions are integrated with a preceding linguistic context, they have not yet addressed the question of whether stereotypes are relevant for making predictions. This is the first question we address in the present study. Secondly, most studies have not directly contrasted stereotypes about women with stereotypes about men but have rather reported the overall processing costs of encountering counter-stereotypical information (cf., Cacciari and Padovani, 2007; Siyanova-Chanturia et al., 2012). It is, thus, unclear whether there are any specific biases for female- or male-stereotypical occupations and how they are integrated into preceding linguistic (and non-linguistic) context, another issue we will examine in our experiment.

Gender Stereotypes Beyond World Knowledge

In the language comprehension literature, the use of gender stereotypes has often been treated as part of world knowledge (e.g., Kreiner et al., 2008). In social psychology, by contrast, studies suggest that people derive their attitudes and stereotypes *from* world knowledge (see Locksley et al., 1980). For instance, Koenig and Eagly (2014) showed that gender stereotypes can be built based on the observation of gender roles: Characteristics of roles that are occupied predominantly by women are attributed to women (e.g., nurses tend to be women, thus women are good at caring for people). A previous study (Hoffman and Hurst, 1990) showed that stereotypes emerge as rationalizations of existing role-related distributions even in the presence of extensive individuating information that is uncorrelated with either the roles or the stereotyped groups. Even more, after stereotypes are established, they can be impervious to real-world information and plausibility.

In a more recent study, Cao and Banaji (2016) tested whether gender-stereotypical associations could be overridden when factual information was provided. In three experiments, participants were presented with two names (e.g., Jonathan, Elizabeth) and were told that one of them was a doctor and the other a nurse. Participants' beliefs about these two characters were assessed using an explicit and an implicit measure both before and after individuating facts were provided. Beliefs at the explicit level were evaluated by asking participants who the doctor was and who the nurse. At the implicit level, participants' beliefs were

measured using an Implicit Association Test (IAT, Greenwald et al., 1998), which measured the strength of association between each individual—Jonathan vs. Elizabeth—and the attribute of doctor vs. nurse. In all three experiments, gender stereotypes operated at the explicit and implicit levels before participants knew individuating facts; when asked about the characters' professions, participants strongly preferred Jonathan to be the doctor and Elizabeth the nurse. Similarly, response times in the IAT matched the beliefs that Jonathan was in fact the doctor and Elizabeth the nurse. More importantly, after individuating factors were presented to the participants, their explicit beliefs were updated and participants responded in accordance with this new information, whereas their implicit beliefs continued to reflect the gender bias.

This leads to the question of whether the use of gendered information could be based on sexist attitudes and not just world knowledge. If only world knowledge drives the processing of gendered information in language, then predictions based on this information should be balanced for both genders. Thus, stereotypes about women should have the same status as stereotypes about men. However, if stereotypes about women are based on sexist attitudes (that is, giving women a lower status than men in society, see Fiske, 1993; Jost and Banaji, 1994), then gender-stereotypical language processing should work differently for stereotypes about women than for stereotypes about men.

The Current Study

We ask whether listeners make use of gender stereotypes (both visually and verbally derived) to make predictions about the agent (that is, the doer or initiator of an action expressed by a verb; see Kroeger, 2005) of verbally conveyed actions. Moreover, if listeners do so, are female and male gender stereotypes treated in the same way? Furthermore, is gender-driven anticipation behavior related to individual differences in explicit beliefs about gender? These questions appear to be particularly important, considering that accounts of prediction during sentence processing have not integrated people's attitudes, but have treated any informational biases as part of real-world knowledge.

To address these questions, we constructed male- and female-stereotypical visual and spoken materials (i.e., occupations and actions, respectively) and combined them to generate six experimental conditions. In each trial, participants saw a display with a male and a female character who, depending on the experimental condition, represented a male-stereotypical occupation (e.g., soldier) or a female-stereotypical occupation (e.g., florist). These were combined with one of three types of sentence in German: sentences conveying female-stereotypical actions (e.g., "The wood is being painted by the florist_{female}," "*Das Holz wird angemalt von der Floristin*"), or conveying male-stereotypical actions (e.g., "The wood is being cut by the florist_{female}," "*Das Holz wird gehackt von der Floristin*"), or a neutral sentence (e.g., "The wood is being stored by the florist_{female}," "*Das Holz wird gelagert von der Floristin*"), which served as a control condition. We predicted that if gender stereotypes guide anticipatory eye movements, we would

observe preferential looks toward the stereotypical character before it is referred to. Moreover, we predicted that the effect of the linguistic stereotype (i.e., actions) on the anticipatory eye-movements would be moderated by the visual stereotype, namely the occupation that the characters in the visual display represent.

Finally, finding a relation between anticipatory eye movements and participants' individual scores on explicit attitudes about gender would provide further support for the idea that the use of gender-stereotypical information during predictive language processing is based on sexist attitudes.

MATERIALS AND METHODS

Participants

Fifty-one German native speakers (16 men; $M_{age} = 23.57$, age range: 18–36 years) from the University community, with normal or corrected-to-normal vision, participated. All participants gave informed consent and were paid either 2€ or given course credits and chocolates for their participation. The number of participants was based on sample sizes previously reported in the literature on predictive eye movements during spoken sentence comprehension (see, e.g., Altmann and Kamide, 1999; Kamide et al., 2003; Borovsky et al., 2012; Huettig and Guerra, 2019).

Experimental Materials

A pre-test on gender associations of different occupations and actions was carried out to generate the verbal and visual materials. Norming data are presented in **Supplementary Material**. Forty-two unique displays were constructed to serve as the visual context in the eye-tracking experiment. Each of them presented a female and a male character representing the same occupation (e.g., a female and a male florist) and two objects, e.g., toothbrush and wooden logs (see **Figure 1**). A total of 14 occupations were used in the experiments, half of which were stereotypically associated with women (e.g., hairdresser, flight attendant), and half stereotypically associated with men (e.g., firefighter, guard).

For each of the 14 occupations, three passive sentences in German were constructed to match the displays, resulting in a total of 42 trials. The sentences always described an action that one of the characters acted upon one of the objects. The action was stereotypically associated with men (e.g., "The wood is being cut by the florist_{female}," see **Figure 1A1**), women (e.g., "The wood is being painted by the florist_{female}," see **Figure 1A2**), or neutral (e.g., "The wood is being stored by the florist_{female}," see **Figure 1A3**), depending on the critical verb. As can be seen in these examples, the agent (e.g., the florist_{female}), or critical noun, was kept the same within each visual context. To balance the stimuli, participants listened to the same number of trials with female and male agents across items.

Additionally, 28 filler trials were also constructed, each with a unique visual context (also with two characters and two objects) and a passive sentence (e.g., "The cell phone is being charged by the firefighter," "*Das Handy wird aufgeladen von dem Feuerwehrmann*"). Characters and objects were repeated

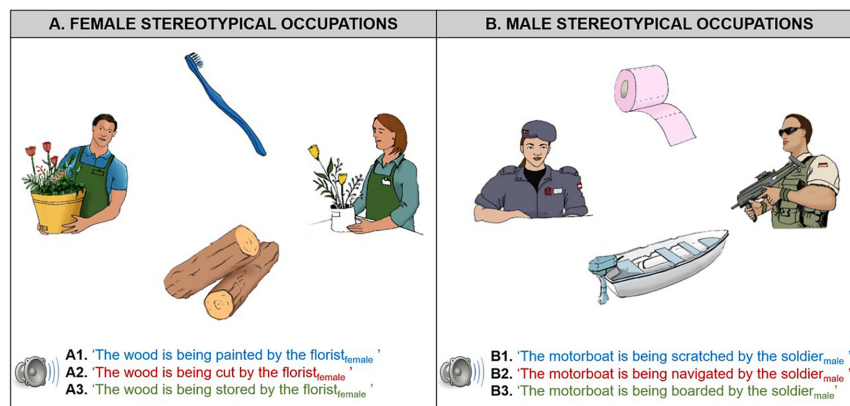


FIGURE 1 | Visual context example as a function of visually derived gender stereotypes (**A**, female-stereotypical occupations; **B**, male-stereotypical occupations), as well as the corresponding verbally conveyed gender stereotypes (1. female-stereotypical actions in blue font; 2. male-stereotypical occupations in red font; 3. neutral in green font). On each trial, participants saw a visual context with either female- or male-stereotypical occupations and heard one out of three possible types of sentences. The objects (e.g., wood, motorboat) and the characters (e.g., florist_{female}, soldier_{male}) were always kept the same within an experimental item, while the sentential verb determined the gender stereotypical action.

among fillers, but none of the verbs from the experimental items were used. After each filler trial, a *yes-or-no* question about the sentences or the pictures was presented (e.g., “Is the tablet being charged by the fireman?” “*Wird das Tablet aufgeladen von dem Feuerwehrmann?*”), which served as a check on participants’ engagement on the task (see **Supplementary Material** for a full list of experimental items and filler materials).

Experimental Design

Our experimental design combined two levels of visually derived gender-stereotypical occupations (i.e., female-stereotypical vs. male-stereotypical occupations) and three levels of verbally conveyed gender-stereotypical actions (i.e., female-stereotypical vs. male-stereotypical vs. neutral actions). We implemented a 2×3 repeated-measures Latin square design, which crossed all six experimental conditions in six experimental lists. Each participant saw one list with every visual context and heard a sentence in one experimental condition. Thus, every participant saw 42 experimental trials, with seven trials per condition. Finally, the same 28 filler trials were also presented on each experimental list. Trial presentation was pseudo-randomized for each participant. In that way, the first trial was always a filler and no more than two experimental items were presented consecutively.

Attitudinal Measures

To assess participants’ explicit attitudes about gender, we used two standardized scales; the Normative Gender Role Orientation scales (NGRO, Athenstaedt, 2000) and the Ambivalent Sexism Inventory (ASI, Glick and Fiske, 1996; German-language version by Eckes and Six-Materna, 1999). The NGRO consists of 29 statements expressing participants’ attitudes toward normative gender roles in society (e.g., “Women are less interested in politics than men,” “*Frauen sind weniger an Politik interessiert*

als Männer”). Participants indicated their agreement with each item on a scale from 1 = *does not apply* to 7 = *applies*). The ASI consists of two 11-item sub-scales, benevolent sexism (ASI-BS; e.g., “No matter how accomplished he might be, a man is not truly complete as a person without the love of a woman,” “*Egal, wie erfolgreich ein Mann auch sein mag, ohne eine Frau, die ihn liebt, fehlt ihm etwas ganz Wichtiges*”) and hostile sexism (ASI-HS; e.g., “Women are too easily offended,” “*Frauen sind zu schnell beleidigt*”). Participants indicated their agreement with each item on a scale from 0 (*I totally disagree*) to 5 (*I totally agree*). The internal consistencies for NGRO (Cronbach- α = 0.89), ASI-BS (Cronbach- α = 0.90), and ASI-HS (Cronbach- α = 0.89) in our sample were high.

Procedure

After giving informed consent, participants first completed the ASI and then the NGRO scale. This took about 10 min. Then, the eye-tracking experiment began, which also took about 10 min to complete. Participants sat comfortably at approximately 70 cm from the computer screen and rested their chins on the eye tracker’s head support. Their eye movements were recorded using an Eyelink 1000 Plus Desktop Mount (SR Research) as they inspected a visual display and listened to linguistic materials through standard computer speakers on each trial. They were instructed to pay attention to what they were hearing and to what they were seeing, which is sometimes called “look-and-listen” studies (see Huettig et al., 2011). They also learned that, occasionally, a *yes-or-no* question about the sentences or the pictures would have to be answered. Before the beginning of the experiment, a default calibration procedure was carried out. On every trial, a participant began fixating a cue in the center of the screen, allowing the experimenter to initiate the trial (or re-calibrate whenever necessary). The visual display was presented for 3 s before the spoken sentences were presented.

DATA ANALYSIS AND RESULTS

Accuracy

Accuracy of responses to the engagement-check items was computed by assigning a zero to incorrect responses and a one to correct responses, and then calculating the mean for each participant. This yields the percentage of correct responses per participant. Participants' accuracy on comprehension questions was at ceiling for most participants ($M = 91\%$; range 79–100%), showing that they engaged in the experiment.

Eye-Tracking Data

Data Analysis

To examine participants' gaze behavior, four areas of interest (AOI) corresponding to the four displayed pictures were defined. Next, a trial-based summary of fixations was produced (Data Viewer software, SR research). This fixation report provided the duration and location of all eye fixations on each trial, which allowed the individualization of the fixation falling into the different interest areas. Subsequently, we used the R Project software (R Core Team, 2020) to further divide our data into time steps of 100 ms. To achieve that, we first inspected all fixations per participant and trial in time steps of 1 ms, where a value of 1 was given to the interest area fixated by the participant at each time step, and a value of 0 to all other areas. Afterward, we aggregated these short time windows by averaging 100 ms again at the participant, trial, experimental condition, and AOI levels. Finally, the average proportion at the participant level and the corresponding 95% confidence intervals (adjusted for within-subjects designs; see Morey, 2008) were calculated for each interest area on each experimental condition for each 100-ms time step. The average onset of the critical verb was 1719.26 ms ($SD = 228.75$ ms) before the onset of the critical noun. Thus, the extent of the time window of analysis goes from 2,000 ms before the onset of the critical noun to 1,000 ms after the onset of the critical noun. This 3,000 ms time window allowed us to determine whether participants exhibited any agent preference before the onset of the critical verb, between the onset of the critical verb and the onset of the critical noun (expected predictive effect), and after the onset of the critical noun (expected referential effect).

Inferential analysis was conducted through non-parametric cluster analysis based on random permutations of conditions (see Barr et al., 2014; Kronmüller et al., 2017; Kronmüller and Noveck, 2019). To do so, we first calculated the log-ratio (see Arai et al., 2007) between the proportion of fixation toward the female and male agents on each condition per participant per trial. Thus, positive log-transformed values represent a preference for the female agent in the visual context, whereas negative values reflect a preference for the male agent in the visual context, independently of experimental condition.

Cluster-based randomization analysis was performed in two stages. We first identified the clusters of interest, defined as a large epoch composed by consecutive 100 ms time windows with reliable effects. We assessed the log-ratio difference between the agents (i.e., female vs. male), as well as the difference between the log-ratio and a chance or zero distribution (i.e., a vector

of zeros reflecting no object preference; see Barzy et al., 2020) for each gender-stereotypical action condition (i.e., female, male, and neutral) independently. Statistical significance ($p < 0.05$) for each 100-ms time window was assessed through mixed-effect linear regressions on our dependent variable (i.e., log-ratio) with stereotypical action (i.e., female vs. male) as fixed effect and random intercepts for participants and items, for each time window and visually derived gender-stereotypical occupation condition separately.

The second stage involved creating three null-hypothesis distributions of t -values, achieved by randomly permutating the values or labels that distinguish different levels of a factor (e.g., female- and male-stereotypical actions). Permutations are based on 2000 iterations in which every 100 ms time window is tested with the labels scrambled in the simulated experiments. Thus, no relation between experimental condition and data remains, providing the null-hypothesis t distributions. The first null-hypothesis distribution was generated by randomly rearranging the visual condition labels (i.e., female- and male-stereotypical occupations), which allowed us to compare the preference for the female agent against the preference for the male agent on each verbal condition individually. To generate the other two null-hypothesis distributions, we first created a chance (or zero) distribution given that log-ratio around zero expressed no object preference (see Barzy et al., 2020). We then created the second and the third null distributions by randomly permutating the female label in the visual condition with the zero label (from a chance distribution) and by permutating the male label in the visual condition with the zero label, respectively. Once the t distributions were computed, we aggregated the t -values at the cluster and iteration levels and then identified the largest absolute summed t -value per iteration and summed them for each cluster level. Finally, we assessed the statistical significance of clusters by comparing the sum of largest t -values of each empirically obtained cluster with the corresponding distribution of largest t -values generated in the simulation. A cluster was considered significant by two-tailed test if it was below percentile 2.5 in that distribution (see Chan et al., 2018).

Results

Figure 2 presents the time-course graphs with the mean proportion of fixations and corresponding adjusted confidence intervals for the critical time window in all six conditions of the 2×3 design. The proportions of fixations (and corresponding CIs) show that neutral sentences afforded no prediction and participants preferred one of the two characters only after the onset of the critical noun. In turn, when participants heard a sentence that described an action stereotypically associated with women, they preferred to look at the female character *before* the agent of that action was mentioned. These predictive patterns, however, differed depending on the visually derived gender-stereotypical occupations: Participants preferred the female agent 1,200 ms before the onset of the critical noun when they inspected a display with two characters representing an occupation stereotypically associated with women, but only 800 ms before the onset of the critical noun when they saw a display with two

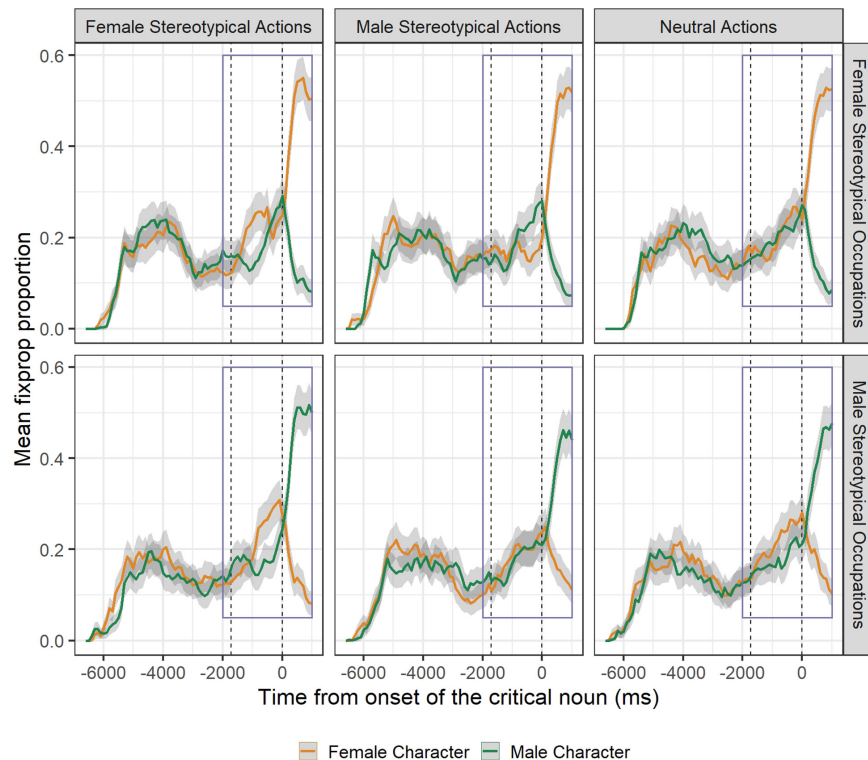


FIGURE 2 | Mean proportion of fixation in all experimental conditions. In all panels, orange lines represent proportion of fixations toward the female character, and green lines represent fixations toward the male character. Gray areas around the main lines represent corresponding 95% CIs adjusted for within-subject designs. The first vertical dashed line represents average onset of the critical verb, and the second line marks the exact onset of the critical noun. The purple squares within the figures mark the window of analysis.

characters representing an occupation stereotypically associated with men.

When participants were presented with a visual context with female-stereotypical occupations and heard male-stereotypical action sentences, we again observed that participants anticipated a character based on the stereotypical action (i.e., the male character). However, the effect appeared only 500 ms before the onset of the critical noun. Finally, when participants were presented with a visual context with male-stereotypical occupations and heard male-stereotypical action sentences, no preferential looks were observed before the onset of the critical noun, resembling the neutral categories.

The results from the cluster-based randomization analysis are consistent with what can be directly inferred from the fixation-proportion time-course plots. In trials where participants heard a neutral sentence, significant clusters appeared only 200 ms after the onset of the critical noun (female agent preference: Observed sum $t = 88.11$; male agent preference: Observed sum $t = 62.24$, contrast between occupation conditions: Observed sum $t = 129.93$, all $p < 0.001$), independently of whether the agents represented an occupation typically associated with women or men.

By contrast, when participants heard a sentence conveying an action stereotypically associated with women, they exhibited a clear preference for the female agent in the visual context.

This preference was significant from 1,200 to 600 ms (observed sum $t = 17.7$, $p < 0.001$) before the onset of the critical noun when participants were presented with a female-stereotypical occupation in the visual context. When they saw a male-stereotypical occupation in the visual context, their preference was somewhat delayed, beginning 800 ms and lasting until 100 ms before the onset of the critical noun (observed sum $t = 22.21$, $p < 0.001$). The differences on how female-stereotypical actions operated when visually situated in a male- vs. female-stereotypical occupation visual context was confirmed by two significant clusters that identified differences between agents' preference in the two distinctive visual contexts (from -1300 to -1000 ms, observed sum $t = 9.58$, $p < 0.01$, and from -400 to -100 ms, observed sum $t = 12.43$, $p < 0.001$). After the onset of the critical noun, participants exhibited a clear preference for the mentioned agent (All clusters started at 200 ms after the noun onset; female agent preference: Observed sum $t = 98.11$; male agent preference: Observed sum $t = 90.14$, contrast between occupation conditions: Observed sum $t = 161.66$, all $p < 0.001$).

When participants heard a sentence expressing an action stereotypically associated with men, they attempted to anticipate the agent of the action (i.e., exhibited more looks to the male rather than the female character before any of them is mentioned), only if the visual context depicted two characters representing occupations typically associated with women.

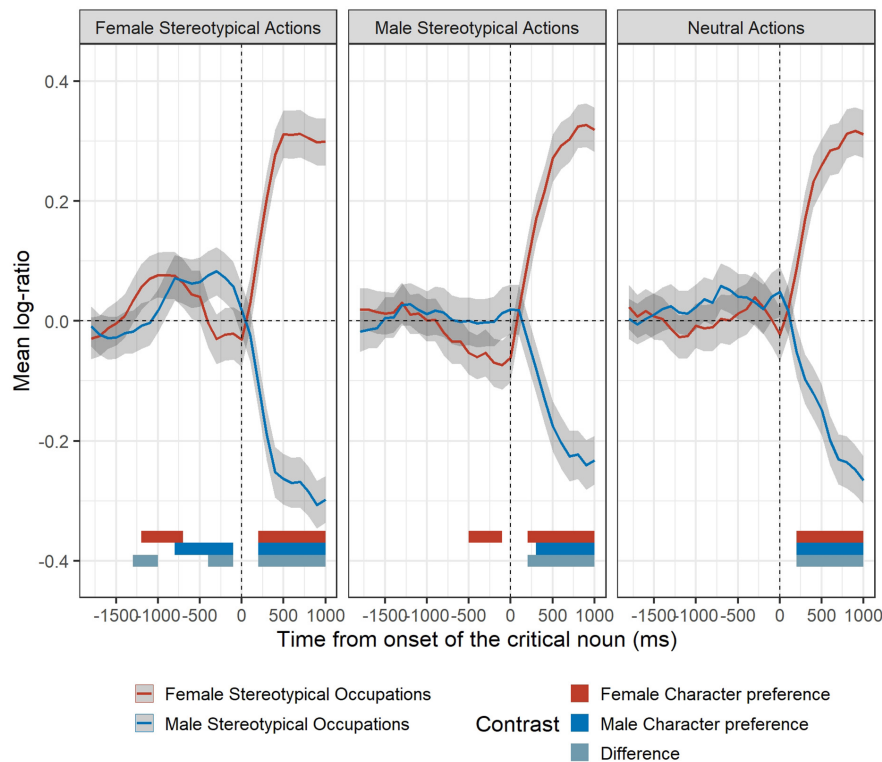


FIGURE 3 | Mean log-ratio between the female and male agents (positive and negative values reflect preference for the female and male character, respectively), as a function of the action experimental condition (neutral actions, female-stereotypical actions, male-stereotypical actions), and aggregated in time steps of 100 ms. Plots are time-locked to the onset of the critical noun. Occupation experimental conditions (female-stereotypical occupations, male-stereotypical occupations) are represented by different line colors. Shaded areas represent within-subject adjusted 95% confidence levels, calculated at the participant level. Bottom horizontal bars visually represent the extent (in milliseconds) of identified significant cluster for each contrast and action experimental condition.

A significant cluster reflecting preference for the male agent was identified between 500 and 100 ms before the onset of the critical noun (observed sum $t = 15.59$, $p < 0.001$). This suggests that male-stereotypical action sentences have a different effect on anticipatory eye movements when situated in a male- vs. female-stereotypical occupation visual context. After the onset of the critical noun, participants preferred to look at the agent mentioned from 200 ms (observed sum $t = 96.04$, $p < 0.001$) in the visual female-stereotypical occupation context and from 300 ms in the visual male-stereotypical occupation context (observed sum $t = 62.75$, $p < 0.001$). The difference between visual contexts also emerged at 200 ms after critical noun onset (observed sum $t = 135.59$, $p < 0.001$). **Figure 3** offers a visual depiction of the cluster analysis on log-ratio between female and male agents in each experimental condition. In particular, the horizontal bars at the bottom show the extension in time of clusters identified as significant for each experimental condition.

Attitudes and Anticipatory Eye Movements

Data Analysis

Participants' scores on the NGRO and the two ASI scales were z-standardized and used as predictors in an ordinary least squares

(OLS) regression for the dependent variable derived from eye movements (i.e., mean log-ratio) aggregated per participant and cluster of interest (i.e., significant clusters occurring before the onset of the critical noun). As we had identified three clusters of interest—that is, a window of time during which we see significant differences between conditions—the condition to which each cluster pertained (i.e., female-stereotypical action in a female-stereotypical occupation visual context, female-stereotypical action in a male-stereotypical occupation visual context, and male-stereotypical action in a female-stereotypical occupation visual context) was also introduced in the regression model as a non-ordinal three-level fixed effect.

As can be seen in **Figure 3**, the male-stereotypical action in a female-stereotypical occupation condition produced anticipation toward the male character, and thus the mean log-ratio values tended to be negative (whereas the female-stereotypical action condition triggered looks to the female character and thus mean log-ratio values tended to be positive). Therefore, participants' mean log-ratio values on that cluster were multiplied by -1 for a more informative comparison between clusters. Our OLS regression uses the data from each scale (by participant), experimental condition, and their interaction to jointly predict the mean log-ratio for each cluster of interest. We conducted three equivalent models (identical predictors), each with a

TABLE 1 | Ordinary least square regression results. Within the table (in bold font), the experimental condition of the cluster that served as intercept in each model.

	Estimate	Se	<i>t</i>	<i>Pr(> t)</i>
Female-stereotypical action and occupation				
(Intercept)	0.070	0.021	3.405	<0.001***
NGRO	−0.033	0.029	−1.139	0.257
ASI (BS)	−0.005	0.025	−0.209	0.834
ASI (HS)	0.008	0.031	0.259	0.796
Male-stereotypical action and female-stereotypical occupation				
(Intercept)	0.062	0.021	3.009	0.003**
NGRO	0.029	0.029	1.003	0.318
ASI (BS)	−0.008	0.025	−0.316	0.752
ASI (HS)	0.003	0.031	0.089	0.929
Male-stereotypical action and occupation				
(Intercept)	0.069	0.021	3.378	<0.001***
NGRO	0.000	0.029	0.009	0.993
ASI (BS)	−0.040	0.025	−1.582	0.116
ASI (HS)	−0.007	0.031	−0.215	0.830

****p* < 0.001; ***p* < 0.01.

different cluster as reference group, by changing only the contrast coding for the cluster predictor levels.

Results

Contrary to our expectations, we observed no main effects of the attitude measures on anticipatory eye movements on any of the clusters, no differences between clusters, and no interaction effects (all *t*-values < |2|). **Table 1** presents a subset of the results from the OLS regression analysis. Concretely, we present the estimates and corresponding statistics for the effect of each scale on each cluster, which we achieved conducting the regression model with different contrast codings, setting the intercept to each cluster. The distributions of the scales and clusters, as well as their intercorrelations, are presented in **Supplementary Material**.

DISCUSSION

Substantial experimental evidence has shown that gender stereotypes do indeed moderate processing time during sentence understanding (Carreiras et al., 1996; Duffy and Keir, 2004; Kreiner et al., 2008). However, we identified a couple of open issues in the context of gender stereotypes. First, no previous experiments had investigated whether gender stereotypes are relevant for predictive language processing as reflected in anticipatory eye movements. Second, we noticed that in the literature on language comprehension, gender-stereotype effects have been interpreted as comprehenders' use of their world knowledge. Literature in social psychology, however, suggests that stereotypes and world knowledge are not equivalent (see, e.g., Cao and Banaji, 2016). Instead, whereas world knowledge is routinely used by people to derive (gender) stereotypes and attitudes, factual counter-stereotypical information might be integrated only at an explicit level but not an implicit level, despite plausibility.

We addressed these questions using a well-established eye-tracking paradigm (Altmann and Kamide, 1999), which allowed us to identify whether participants would anticipate a visual referent based on stereotypical information and to dissociate the moment-by-moment effects of gender stereotypes about women and men during language comprehension. The results showed that participants used gender stereotypes in language and the scene in real time to predict the agent of the sentence. Interestingly, this anticipation was not symmetrical for female and male stereotypes. Analysis of the gaze patterns' time course during language comprehension revealed earlier and longer predictive eye movements when sentences conveyed female stereotypes than when sentences conveyed male stereotypes. Moreover, verbally conveyed gender stereotypes (i.e., actions) interacted with the visually derived stereotypes (i.e., occupations represented by the characters). Anticipatory eye-movements to the female character occurred 300 ms earlier when female-stereotypical action sentences were accompanied by female-stereotypical occupations (e.g., florist) than when those sentences were presented together with male-stereotypical occupations (e.g., soldier). Similarly, when male-stereotypical action sentences were presented together with female-stereotypical occupations we observed relatively late anticipatory eye-movements toward the male character, already 300 ms later than when participants heard female-stereotypical action sentences and were looking at male-stereotypical occupations.

These findings are consistent with a view in which sexism moderates predictive sentence processing. Gender stereotypes rapidly triggered anticipatory eye movements, in particular when visually and linguistically derived representations were consistent, and less so when they were in conflict. Moreover, these effects appear to have had a distinct time course for female- and male-stereotypical information. Although not predicted, we conjecture that this asymmetry in our findings is consistent with a view of women (as opposed to men) being the main targets of sexism. Indeed, negative stereotypes of women (see Glick and Fiske, 1996) are more pervasive and central for male-dominated societies (see e.g., Sidanius and Pratto, 2001; but cf. Eagly and Mladinic, 1994). Consequently, if sexism drives anticipatory eye-movements, it might be responsible for the earlier and larger effect for female-stereotypical action sentences relative to male-stereotypical action sentences that we observed.

The pattern observed for male-stereotypical action sentences when presented together with male-stereotypical occupations seems, at first glance, more intriguing. Although the visual depictions and sentences were stereotypically consistent (both male-biased), no predictive effect was observed at all. We argue, however, that this finding as well may be seen as consistent with the notion that sexist bias can drive prediction during language processing. It is also consistent with literature on gender stereotypes outside of the language-processing domain: When processing gendered information, participants are biased to routinely check whether such information is aligned with female stereotypes (see Glick and Fiske, 1996). From this viewpoint, in our task the listener did not evaluate who was the character that was more likely to be the agent of the action described. Rather, the listener may have assessed (in real time) how likely it is that

the female character would be performing the action described. Consequently, when the male-stereotypical action sentences were presented together with the female character impersonating a male-stereotypical occupation (and thus, not conforming with the gender stereotype), she was not less likely to perform a male-stereotypical action relative to the male character, and thus we observed no anticipatory eye movements.

This mechanism is also consistent with the advantage observed for the female-stereotypical action sentences; a female-stereotypical occupation is perceived as being highly likely to perform a female-stereotypical action, whereas a counter-stereotypical female character is perceived as less so. Finally, this mechanism could also explain the delayed predictive effect found in the male-stereotypical action sentence with female-stereotypical occupation; as the sentence unfolded, listeners discarded the female character as the likely agent of the action, and only then they predicted the male-character to be more likely.

In this context, it would be reasonable to expect that the anticipatory preference for agents at the individual level should be related to individual differences in NGRO and ASI; however, we failed to find such effects (see **Table 1**). It is important to note that the NGRO and ASI scales assess explicit beliefs, that is, participants consciously reported their agreement with statements that are related to normative gender roles and sexist attitudes toward women. By contrast, anticipatory eye movements toward objects in a visual context during spoken language comprehension may be described as indicating an implicit attitude, at least to the extent that the representations activated by language automatically direct overt attention to related depicted objects. Thus, although explicit beliefs and attitudes about gender appeared to have no effect on predictive eye movements, we cannot rule out that such behavior would be related to gender bias as assessed by implicit measures of sexism (e.g., the IAT; see Glick and Fiske, 1996; Rudman et al., 2001). If such a correlation were found in future research, it would corroborate our interpretation. Also, given the unexpected nature of the asymmetry in predictive eye-movements caused by target gender, replication studies specifically addressing this aspect would be welcome.

In sum, we showed that predictive language processing is moderated by gender stereotypes. Importantly, we also found that this prediction is stronger for female (vs. male) depicted characters, consistent with a view in which sexism affects sentence processing incrementally. These findings contribute to both social psychology and psycholinguistics research, insofar as they extend our understanding of stereotype processing in multimodal contexts and with regard to the role of attitudes (on top of world knowledge) in language prediction.

REFERENCES

- Altmann, G. T., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1
- Arai, M., van Gompel, R. P. G., and Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cogn. Psychol.* 54, 218–250. doi: 10.1016/j.cogpsych.2006.07.001

DATA AVAILABILITY STATEMENT

All data and analysis scripts are available online in <https://osf.io/asdfn>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Bielefeld University Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EG, JB, and GB developed the study concept. EG and JB implemented the experiment and analyzed the data. JB collected the data. EG and HC drafted the manuscript and it was revised by GB and JB. All authors interpreted the results and approved its final version for submission.

FUNDING

This research was funded by the Cognitive Interaction Technology Excellence Cluster, CITEC, German Research Council (DFG), the Centre for Social Conflict and Cohesion Studies (Grant ANID/FONDAP/15130009), the Interdisciplinary Center for Intercultural and Indigenous Studies (Grant ANID/FONDAP/15110006), and the research grants ANID/FONDECYT #11161074 to HC and ANID/FONDECYT #11171074 to EG. Supported from ANID/PIA/Basal Funds for Centers of Excellence FB0003 to EG was also gratefully acknowledged.

ACKNOWLEDGMENTS

We are grateful to Tatiana Villablanca and Claire Vandenbossche for producing all visual stimuli.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.589429/full#supplementary-material>

- Athenstaedt, U. (2000). Normative geschlechtsrollenorientierung: entwicklung und validierung eines fragebogens. *Zeitschr. Different. Diagnost. Psychol.* 21, 91–104. doi: 10.1024//0170-1789.21.1.91
- Barr, D. J., Jackson, L., and Phillips, I. (2014). Using a voicetoput a nameto a face: the psycholinguistics of proper name comprehension. *J. Exp. Psychol. Gen.* 143, 404–413. doi: 10.1037/a0031813
- Barzy, M., Black, J., Williams, D., and Ferguson, H. J. (2020). Autistic adults anticipate and integrate meaning based on the speaker's voice: evidence from

- eye-tracking and event-related potentials. *J. Exp. Psychol. Gen.* 149, 1097–1115. doi: 10.1037/xge0000705
- Borovsky, A., Elman, J. L., and Fernald, A. (2012). Knowing a lot for one's age: vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *J. Exper. Child Psychol.* 112, 417–436. doi: 10.1016/j.jecp.2012.01.005
- Brown, R. (2011). *Prejudice: Its social psychology*. Hoboken, NJ: John Wiley & Sons.
- Cacciari, C., and Padovani, R. (2007). Further evidence of gender stereotype priming in language: semantic facilitation and inhibition in Italian role nouns. *Appl. Psychol.* 28, 277–293. doi: 10.1017/s0142716407070142
- Cao, J., and Banaji, M. R. (2016). The base rate principle and the fairness principle in social judgment. *Proc. Natl. Acad. Sci.* 113, 7475–7480. doi: 10.1073/pnas.1524268113
- Carreiras, M., Garnham, A., Oakhill, J., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from english and spanish. *Q. J. Exper. Psychol.* 49, 639–663. doi: 10.1080/027249896392531
- Chan, A., Yang, W., Chang, F., and Kidd, E. (2018). Four-year-old Cantonese-speaking children's online processing of relative clauses: a permutation analysis. *J. Child Lang.* 45, 174–203. doi: 10.1017/S0305000917000198
- DeLong, K. A., Troyer, M., and Kutas, M. (2014). Pre-processing in sentence comprehension: sensitivity to likely upcoming meaning and structure. *Lang. Linguis. Compass* 8, 631–645. doi: 10.1111/lnc3.12093
- DeLong, K. A., Urbach, T. P., and Kutas, M. (2017). Is there a replication crisis? perhaps. Is this an example? no: a commentary on ito, martin, and nieuwland (2016). *Lang. Cogn. Neurosci.* 32, 966–973. doi: 10.1080/23273798.2017.1279339
- Dovidio, J. F., Glick, P. E., and Rudman, L. A. (2005). *On the nature of prejudice: Fifty years after Allport*. Hoboken, NJ: Blackwell Publishing.
- Duffy, S. A., and Keir, J. A. (2004). Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge. *Mem. Cogn.* 32, 551–559. doi: 10.3758/bf03195846
- Eagly, A. H., and Mladinic, A. (1994). Are people prejudiced against women? some answers from research on attitudes, gender stereotypes, and judgments of competence. *Eur. Rev. Soc. Psychol.* 5, 1–35. doi: 10.1080/14792779543000002
- Eckes, T., and Six-Materna, I. (1999). Hostilität und benevolenz: eine skala zur erfassung des ambivalenten sexismus. *Zeitschr. Sozialpsychol.* 30, 211–228. doi: 10.1024/0044-3514.30.4.211
- Fiske, S. T. (1993). Controlling other people: the impact of power on stereotyping. *Am. Psychol.* 48, 621–628. doi: 10.1037/0003-066x.48.6.621
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Glick, P., and Fiske, S. T. (1996). The ambivalent sexism inventory: differentiating hostile and benevolent sexism. *J. Person. Soc. Psychol.* 70:491. doi: 10.1037/0022-3514.70.3.491
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Person. Soc. Psychol.* 74:1464.
- Hauk, O. (2016). Preface to special issue “prediction in language comprehension and production”. *Lang. Cogn. Neurosci.* 31, 1–3. doi: 10.1080/23273798.2015.1102300
- Hoffman, C., and Hurst, N. (1990). Gender stereotypes: perception or rationalization? *J. Person. Soc. Psychol.* 58:197. doi: 10.1037/0022-3514.58.2.197
- Huetting, F. (2015). Four central questions about prediction in language processing. *Brain Res.* 1626, 118–135. doi: 10.1016/j.brainres.2015.02.014
- Huetting, F., and Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Res.* 1706, 196–208. doi: 10.1016/j.brainres.2018.11.013
- Huetting, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* 137, 151–171. doi: 10.1016/j.actpsy.2010.11.003
- Jost, J. T., and Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *Br. J. Soc. Psychol.* 33, 1–27. doi: 10.1111/j.2044-8309.1994.tb01008.x
- Jost, J. T., and Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: consequences for specific and diffuse forms of system justification. *J. Person. Soc. Psychol.* 88:498. doi: 10.1037/0022-3514.88.3.498
- Kamide, Y., Altmann, G. T., and Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J. Mem. Lang.* 49, 133–156. doi: 10.1016/s0749-596x(03)00023-8
- Knoeferle, P., and Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: evidence from eye movements. *J. Mem. Lang.* 57, 519–543. doi: 10.1016/j.jml.2007.01.003
- Kronmüller, E., and Noveck, I. (2019). How do addressees exploit conventionalizations? From a negative reference to an ad hoc implicature. *Front. Psychol.* 10, 1–10. doi: 10.3389/fpsyg.2019.01461
- Kronmüller, E., Noveck, I., Rivera, N., Jaume-Guazzini, F., and Barr, D. (2017). The positive side of a negative reference: the delay between linguistic processing and common ground. *Royal Soc. Open Sci.* 4:160827. doi: 10.1098/rsos.160827
- Koenig, A. M., and Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: observations of groups' roles shape stereotypes. *J. Person. Soc. Psychol.* 107, 371–392. doi: 10.1037/a0037215
- Kreiner, H., Sturt, P., and Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: evidence from eye-movements. *J. Mem. Lang.* 58, 239–261. doi: 10.1016/j.jml.2007.09.003
- Kroeger, P. R. (2005). *Analyzing grammar: An introduction*. Cambridge: Cambridge University Press.
- Locksley, A., Borgida, E., Brekke, N., and Hepburn, C. (1980). Sex stereotypes and social judgment. *J. Person. Soc. Psychol.* 39, 821–831. doi: 10.1037/0022-3514.39.5.821
- Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutorial Quant. Method Psychol.* 4, 61–64. doi: 10.20982/tqmp.04.2.p061
- Nicenboim, B., Vasisith, S., and Rösler, F. (2020). Are words pre-activated probabilistically during sentence comprehension? evidence from new data and a bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia* 142:107427. doi: 10.1016/j.neuropsychologia.2020.107427
- Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., et al. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife* 7:e33468.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/>
- Rudman, L. A., Greenwald, A. G., and McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: self and ingroup share desirable traits. *Person. Soc. Psychol. Bull.* 27, 1164–1178. doi: 10.1177/0146167201279009
- Sidanius, J., and Pratto, F. (2001). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge: Cambridge University Press.
- Siyanova-Chanturia, A., Pesciarelli, F., and Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS One* 7:e48712. doi: 10.1371/journal.pone.0048712
- Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *J. Exper. Soc. Psychol.* 35, 4–28.
- van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.* 20, 580–591. doi: 10.1162/jocn.2008.20054

Conflict of Interest: 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 © 2021 Guerra, Bernotat, Carvacho and Bohner. 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.



Attriters and Bilinguals: What's in a Name?

Federico Gallo^{1,2*}, Keerthi Ramanujan¹, Yuri Shtyrov^{1,3} and Andriy Myachykov^{1,4}

¹ Centre for Cognition and Decision Making, Institute for Cognitive Neuroscience, HSE University, Moscow, Russia, ² Centre for Neurolinguistics and Psycholinguistics, Vita-Salute San Raffaele University, Milan, Italy, ³ Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark, ⁴ Department of Psychology, Northumbria University, Newcastle upon Tyne, United Kingdom

The use of language as a universal tool for communication and interaction is the backbone of human society. General sociocultural milieu and specific contextual factors can strongly influence various aspects of linguistic experience, including language acquisition and use and the respective internal neurolinguistic processes. This is particularly relevant in the case of bilingualism, which encompasses a diverse set of linguistic experiences, greatly influenced by societal, cultural, educational, and personal factors. In this perspective piece, we focus on a specific type of linguistic experience: non-pathological first-language (L1) attrition—a phenomenon that is strongly tied to immersion in non-L1 environments. We present our view on what may be the essence of L1 attrition and suggest ways of examining it as a type of bilingual experience, in particular with relation to its neurocognitive bases.

OPEN ACCESS

Edited by:

Ramesh Kumar Mishra,
University of Hyderabad, India

Reviewed by:

Arturo Hernandez,
University of Houston, United States
Meesha Alecia Warmington,
The University of Sheffield,
United Kingdom

*Correspondence:

Federico Gallo
fgallo@hse.ru

Specialty section:

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

Received: 01 May 2020

Accepted: 14 June 2021

Published: 15 July 2021

Citation:

Gallo F, Ramanujan K, Shtyrov Y
and Myachykov A (2021) Attriters and
Bilinguals: What's in a Name?
Front. Psychol. 12:558228.
doi: 10.3389/fpsyg.2021.558228

Keywords: bilingualism, L1 attrition, bilingual experience, L2 immersion, sociocultural changes

INTRODUCTION

Globalization and mobility are increasingly establishing themselves as defining features of current world. Reports of the United Nations highlight a steady yearly growth in migration; in the last 20 years, the phenomenon doubled in magnitude, with the number of migrants reaching the figure of 260 million in 2017 (United Nations, 2018). As a result of such mobility, alongside changes in educational requirements, and internationalization of the job market, more than half of the world population is currently estimated to be bilingual (e.g., Grosjean, 2010). For instance, bilingual citizens constitute 21.5% of the grand total in the United States population (American Community Survey, 2015), 17.5% in Canada (Canada Census Program, 2011), and 54% in the European Union, where over 90% of the bilingual population reached peaks in some countries (Eurobarometer Report “Europeans and their languages”, 2012). A unique challenge for the bilingual mind is the simultaneous storage and management of two or more linguistic codes, which have been shown to be in a constant interaction with each other. In fact, a considerable amount of literature has investigated and demonstrated the influence of the first-language (L1) on the second one (L2; e.g., Dijkstra and Van Heuven, 2002). The equally important and plausible effects of the L2 on the native language have received much less attention. Here, we focused on one specific type of language experience that is closely associated with the latter—*L1 attrition*—the non-pathological, gradual decrease of native language performance that takes place alongside with, and even without (Baladzaeva and Laufer, 2018), increase in L2 proficiency (Köpke and Schmid, 2004). Research on L1 attrition emerged in the early 1980s, and at present, almost 40 years later, still occupies a relative niche in the field of bilingualism. Nonetheless, a great progress has been made in the last two decades, due to a remarkable effort from attrition researchers to affine definitions and develop tools to investigate the phenomenon and its underlying mechanisms.

In this short opinion piece, after briefly defining the concept of L1 attrition and reviewing its putative underlying mechanisms, we advocated for the continuation and reinforcement of a trend put forward by attrition researchers in the last 15 years, unifying the characters of the bilingual and the attriter within the shared theoretical framework of crosslinguistic interaction. Our goal is to appeal to “traditional bilingualism” researchers to firmly establish the attrition phenomenon in empirical investigations as well as in theoretical platforms, as accounting for “the other side of the coin” can help to shed light on the neural and cognitive phenomena of the bilingual mind.

WHAT IS L1 ATTRITION?

Among the many existing characterizations of this phenomenon, described by Köpke (2004) as a “terminological jungle,” L1 attrition has been broadly defined as “any of the phenomena that arise in the native language of a sequential bilingual as the consequence of the co-activation of language, crosslinguistic transfer or disuse” (Schmid and Köpke, 2017a, p. 637)¹. Guided by this view of L1 attrition, there has been an extensive characterization and a detailed analysis of how the linguistic behaviors of bilingual attriters, particularly those who are captured in productive language and “offline” tasks, contrast with those of monolingual speakers. Examples of such studies include research in L1 accent/phonology attrition (Bergmann et al., 2016; de Leeuw et al., 2018), analyses of morphosyntactic reconfigurations (Karayayla and Schmid, 2019), and the changes in L1 fluency and complexity induced by L2 exposure (Schmid and Jarvis, 2014; Bergmann et al., 2015a). From this viewpoint, linguistic behavioral deviations from monolingual standards are considered as indicative of L1 attrition. While there is extensive evidence of L1 attrition arising from L2 interference (see, Köpke and Schmid, 2004; Schmid and Köpke, 2017a), some studies also report L1 attrition occurring in the absence of an L2. For example, Laufer and Baladzhaeva (2015) and Baladzhaeva and Laufer (2018) investigated lexical, grammatical, and morphosyntactic L1 attrition in a sample of Russian immigrants in Israel with no knowledge of Hebrew as L2, comparing them with immigrant Russian/Hebrew bilinguals and Russian monolinguals still residing in Russia. Their results suggested that L1 linguistic behavior is susceptible to change even without the explicit knowledge of an intervening linguistic system. While L2 interference might still have made a contribution to L1 attrition in the latter case *via* a possible passive exposure to Hebrew spoken within the bilingual Russian immigrant population, the lack of explicit knowledge of L2 in the attriting population indicates a potentially intricate and complex nature of the attrition phenomenon.

This intricate phenomenological nature begs the following question: *What is L1 attrition?* To answer this question, we (1) raised two more questions of why and how it occurs, reviewing

recent findings and (2) endorsed a research strategy that builds upon and might contribute to recent developments aimed at unifying attrition and bilingual research fields. We acknowledged that, in posing these questions, there is an unavoidable circularity problem—each question rests on the assumption that L1 attrition is already defined despite this being the very thing that one hopes to achieve (we discussed a possible solution in section “Who is the L1 attriter—a bilingual by another name?”). Nevertheless, the answers to these interconnected questions may contribute toward qualifying the phenomenon of L1 attrition, and equally importantly they may help us understand how L1 attrition relates (and, essentially, belongs) to the general phenomenon of bilingualism.

Question 1: Why Does Attrition Occur?

First, there appears to be a certain selectivity of L1 attrition effects, i.e., there is considerable interindividual variation both in the severity of its “symptoms” and in the types of linguistic structures and systems it affects (Schmid, 2014). Thus, an alternative question, and perhaps one that is more specific would be, “*When* does attrition occur?”—i.e., under what specific conditions are L1 attrition effects most likely to appear?

It may be informative to begin by considering a situation in which L1 attrition disappears. A return (even if temporary) to the L1-dominant/native environment induces a rapid reversal of L1 attrition effects (e.g., Chamorro et al., 2016b; Gargiulo and van de Weijer, 2018; Köpke and Genevskaya-Hanke, 2018). On the one hand, this phenomenon suggests a potential role by the relative *quantity* and *quality* of contact with L1 as reimmersion in the L1-dominant environment brings better, more frequent opportunities to use L1 of an individual. On the other hand, it reinforces the role of L2 in inducing and driving changes in L1 as contact with L2 is naturally reduced upon a return to the L1 environment. In fact, experimental evidence has pointed at quantity (e.g., de Bot et al., 1991; Isurin, 2007; Opitz, 2013; Bergmann et al., 2016; Chamorro et al., 2016b; Kasparian et al., 2017; Schmid and Yilmaz, 2018; Karayayla and Schmid, 2019) and quality (e.g., Schmid, 2007; de Leeuw et al., 2010; Schmid and Dusseldorp, 2010; de Leeuw et al., 2012; Yilmaz and Schmid, 2012) of L1 exposure and at L2 interference (e.g., Ben Rafael, 2001; Dussias, 2004; Hutz, 2004; Ventureyra et al., 2004; Ribbert and Kuiken, 2010; Schmid and Jarvis, 2014; Chamorro et al., 2016a; de Leeuw et al., 2018) as factors contributing to the presence and the severity of L1 attrition.

Besides these main causal factors, however, research has also highlighted a number of other key variables that can shape the L1 attrition experience of an individual, including the length of residence in an L2-dominant country, age of migration, attitude toward L1 and L2, communal/social identity and affiliation, extent of social integration, socioeconomic status, and age. Since an extensive and detailed review of experimental findings goes beyond the scope of this opinion, we redirected the reader to a recent work by Schmid et al. (2019) and references therein for a thorough account of existing evidence. It is important, however, to note that these factors are not exclusive to L1 attrition, but that they also contribute to variation in the bilingual experience *per se*. Nevertheless, taking together both the selectivity and reversibility

¹Here, guided by our abovementioned aim of generalizing L1 attrition to bilingualism research, we decided to adopt this rather broad definition as a starting point, but note that other definitions have been formulated. For an extensive review, see Köpke (2004).

of L1 attrition effects, it seems that whatever L1 attrition is, it does not appear to involve the actual erasure of tacit linguistic knowledge or representations. This brings us to the next question.

Question 2: How Does Attrition Occur?

Considering what we know thus far, one might ask: Is L1 attrition a stand-alone phenomenon, resulting (perhaps temporarily) from a qualitative and quantitative reduction in L1 use (with or without L2 exposure) or does it emerge due to the words, categories, and rules of L2 interfering with those of L1? Exploring what L1 attrition looks like in the mind and the brain of those experiencing it may help us find possible answers to this question.

Changes in observable linguistic behavior (measured using free speech and “offline” grammaticality judgments) and productive language suggest a possible change in the underlying neuro-psycho-linguistic processes that support such behaviors. The detailed analyses and descriptions of L1 attrition tell us how it may manifest at the behavioral end point (see text footnote 1 for examples of such studies). However, behavioral investigations alone can only hint at *neurocognitive mechanisms* underlying L1 attrition. Electrophysiological and neuroimaging techniques permit simultaneous investigations of both ends of the brain-behavior loop. Thus, examining “L1 attrition” from the point of view of how L1 language processing mechanisms have (or have not) changed might potentially bridge the existing brain-behavior knowledge gap in these studies (see Kasparian and Steinhauer, 2017b; Smith, 2019, for a similar view). In fact, such studies have emerged in recent years (see Schmid et al., 2019, for a comprehensive collection), and they reveal very interesting findings. For instance, using electrophysiological measures (ERPs—event-related potentials), Kasparian et al. (2017) and Kasparian and Steinhauer (2016, 2017a) have shown that certain forms of L1 morphosyntactic processing in L1 attriters are indeed different from that of non-attriting monolinguals.

One advantage of studying L1 attrition using neurophysiological and neuroimaging techniques is that it can help uncover neurocognitive features possibly characterizing attrition even in instances where it does not manifest in external linguistic behaviors. For example, some studies show that attriting bilinguals perform no differently than the non-attriting controls in offline behavioral sentence judgments, while the ERP data show specific group differences: for instance, where German monolinguals exhibit a posterior P600 effect to verb form violations, attriters show a biphasic N400-P600 pattern (Bergmann et al., 2015a). Similarly, Italian attriters display a more temporally distributed late P600 in response to anomalous sentences, while non-attriting peers exhibit only small P600 effects (Kasparian and Steinhauer, 2016). Taken together, these findings suggest that despite offline linguistic performance parity, underlying neurocognitive computations in L1 attriters proceed somewhat differently. In fact, investigating L1 attrition by capturing internal neurolinguistic and neurocognitive processes will be particularly useful in understanding the various attrition experiences (e.g., Laufer and Baladzhayeva, 2015), going beyond their apparent external presentations.

WHO IS THE L1 ATTRITER—A BILINGUAL BY ANOTHER NAME?

The two questions posited above motivate a very important third question: *Who are L1 attriters?* Are they defined by (1) external circumstances, e.g. immigration to an L2 environment, (2) their apparent linguistic behaviors, e.g., changes/reconfigurations in L1 production or offline performance deviations involving L1 morphosyntax, or (3) a specific set of internal neurocognitive states and processes? Equally important question is that is attrition really that separable or distinct from bilingualism *per se*?

It is recalled that Schmid and Köpke (2017a) regarded L1 attrition to be the “effect of the second language on the first” and, by relabeling the crosslinguistic influence of L2 on L1 as L1 attrition, they highlighted that “all bilinguals are attriters” (Schmid and Köpke, 2017b). Here, we shifted the focus to the complementary argument that “all attriters are first and foremost bilinguals” to highlight the point that L2 → L1 effects are non-separable from bilingualism and that “attrition” is indeed a sub-phenomenon of bilingualism. In fact, the key point of L1 attrition research is that not only the L1 influences the L2 but the L2 also affects the L1. A different terminological choice must not create barriers between overlapping research fields—this dynamic interaction of two languages has long been regarded as one of the most defining features of bilingualism (Kroll et al., 2012, 2015). Crosslinguistic interplay (L2 ↔ L1), something that monolinguals obviously lack, is arguably the greatest influence behind the distinctive organization and functioning of the bilingual mind and brain (Hernandez et al., 2005, 2019a; Hernandez, 2013; Li et al., 2014; Bialystok, 2017; Hayakawa and Marian, 2019). In fact, there is currently a substantial amount of empirical evidence demonstrating the specific ways in which L1 and L2 of the bilinguals are affected by one another (Kroll et al., 2012; Coderre, 2015). While the effect of a dominant L1 on a later acquired, relatively less proficient L2 might be expected and considered natural, there is also substantial evidence for an effect of the L2 on the L1. Various types of bilinguals, even in communicative contexts requiring exclusive use of their L1, exhibit certain behaviors due to unavoidable L2 influence that their monolingual peers do not—these include slower word production, decreased accuracy, lower semantic fluency, and increased tip-of-tongue states in their L1 (Bialystok et al., 2012; Costa and Sebastián-Gallés, 2014; Kroll et al., 2015). These are accepted as a natural outcome of housing two *interacting* language systems and are hardly ever labeled as instances of attrition. It is possible that the linguistic behaviors considered to amount to L1 attrition could be due to *changes in the L1–L2 dynamic*, changes in the extent to which the two languages influence one another, brought on by experiential changes including migration to different sociolinguistic environments, and not just purely due to the effects of an L2 on the L1, which persist regardless in *all* bilinguals. Similarly, Hernandez et al. (2019b, p. 260) suggested that the bilingual system is non-linear and dynamic (Hernandez et al., 2019a) and that attrition indicates its reconfiguration(s) and repurposing to suit new contexts. This is something that attrition research has been acknowledging for

a long time (Schmid and Köpke, 2007). Thus, as also proposed by Schmid and Köpke (2007), the relevance of L1 attrition for the theories of bilingual development is reinforced by the fact that L1 attrition appears not to be intrinsically distinct from bilingualism, but rather a feature of the latter.

To determine whether or not L1 attrition is qualitatively distinct from bilingualism as such, one would need to determine features that are attributable to L1 attrition alone. However, in nearly all attrition studies thus far², the L1 attriters are bilinguals, while the non-attriting controls are mostly monolinguals. Thus, the comparisons are not really between “attriters” and “non-attriters” but between those whose L1 surely receives some influence from the L2 and those whose L1 is free from it. Consequently, the findings of group-level linguistic and behavioral differences taken as an indication of L1 attrition can be seen as consequences of bilingualism and fall within the typical range of behaviors that bilinguals exhibit owing to their specific linguistic situation (e.g., see Bergmann et al., 2015b).

This interpretation is equally applicable to the existing ERP studies where the key ERP differences between attriters and monolingual controls might actually reflect a bilingualism artifact. For example, Bergmann et al. (2015a) acknowledged that the biphasic N400-P600 pattern in their bilingual attriters may reflect the relative linguistic difference between the L1 and L2 and is similar to findings reported by Sabourin and Stowe (2008), who examined how L1–L2 similarity alters language processing in (non-attriting) bilinguals. Similarly, Kasparian and Steinhauer (2016) noted that the larger P600 response that only their attriter group exhibited is possibly suggestive of increased conflict-monitoring and “re-checking” while processing anomalous lexico-semantic components in sentences. The fact that bilinguals, compared with monolinguals, rely on the increased extent of conflict monitoring while processing either language to control and manage crosslinguistic influence is empirically well established (Bialystok, 2017; Calabria et al., 2018). However, if the attriting samples in both studies were to be compared with a non-attriting bilingual group that has managed to maintain L1 use (with or without migration), it would tell us whether or not the specific P600 response (signaling increased conflict monitoring) or the biphasic responses (modulated by linguistic distance) are unique to the attriters. In this case, it could qualify as a specific L1-attrition marker; otherwise, if it is shared with bilingual controls, this would signify a bilingualism effect arising due to the presence of two interacting linguistic systems but not a distinct marker of attrition. To the best of our knowledge, no such studies have been done to date.

Thus, we supported the view that L1 attrition, characterized by a relative increase in frequency, quantity, quality, diversity of L2 use, and exposure and a possible concomitant decrease in all these aspects for the L1, be included within the spectrum of bilingual experiences. We further encouraged researchers to view attriters as a group within the *bilingual spectrum* and

sought a better understanding of how distinct they may be, if at all. Further neurocognitive and behavioral comparisons between *attriting and non-attriting bilinguals* (e.g., Major, 2010; Schmid, 2014; Miller and Rothman, 2020) are needed to continue unveiling *what* L1 attrition really is and determine whether there are indeed linguistic behaviors and associated neurocognitive processes that are distinct and separable enough from the range of crosslinguistic bilingual effects to be characterized as L1 attrition. Findings from extant L1 attrition research also seem to reinforce the notion of “heterogeneous outcomes of heterogeneous bilingual experiences.” For example, increasing length of immersion, L2 exposure, and L2 proficiency has been shown to modulate L1 ERP responses among bilingual “L1 attriters” themselves (e.g., Kasparian and Steinhauer, 2017a; Miller and Rothman, 2020). This is in line with recent findings demonstrating that variations in the aforementioned bilingual experiential factors have a discernible impact on the linguistic neurobiology, neurocognition, and behavior of the bilinguals (e.g., Gullifer et al., 2018; DeLuca et al., 2019; Gallo et al., 2021; Sulpizio et al., 2020).

Finally, we addressed the circularity problem mentioned earlier: to study L1 attrition, we already needed to have an idea of *what* it is and where to find it. In order to break free from the risk of circularity, i.e., defining the population of interest in terms of the concept of interest, one needs to determine and adopt attrition-free criteria for identifying the population of interest. Based on the well-documented evidence of L1 attrition effects, we stated with reasonable confidence that bilinguals immersed in an L2-dominant sociocultural environment due to migration are clearly the population of interest here—they are most likely to exhibit behavioral and neurocognitive characteristics that might differ from their bilingual peers elsewhere, especially those in L1-dominant environments. Rather than immediately labeling the migrant bilinguals as L1 attriters, we could qualify and quantify their dual-language experience accompanying migration, acculturation, and immersion in an L2-dominant region as just another variant of the bilingual experience. Comparing this L1-attrition or disuse type of bilingual experience to other experiences on the bilingual spectrum, such as the L1-maintenance experience of non-migrant bilinguals in L1-dominant regions or of those who have managed to maintain active dual-language use despite migration, might generate enough empirical evidence that would allow us to deduce whether or not L1 attrition is wholly distinct and orthogonal from bilingualism.

CONCLUSION

Attrition experience is highly typical for bilinguals who have migrated to non-L1 environments, which motivates understanding of the L1 attrition phenomenon as a distinct type of bilingual experience. It is therefore important to provide comparisons involving migrant bilingual populations with suitable bilingual controls (besides monolingual ones) on linguistic behaviors as well as on the underlying neurocognitive processes. Studying L1 attrition from the vantage point of

²With the exception, of a small set of attriting monolinguals who nonetheless, in the interpretation of the same experimenters reporting this evidence, are most probably experiencing indirect attrition via interaction with attriting bilinguals (Laufer and Baladzhava, 2015; Baladzhava and Laufer, 2018).

bilingualism would not only contribute toward enriching and informing our current understanding of bilingualism itself, but it will also offer a richer perspective on how sociocultural factors shape linguistic behaviors and processes.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

REFERENCES

- American Community Survey (2015). Available online at: <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk> (accessed April 25, 2020).
- Baladzhaeva, L., and Laufer, B. (2018). Is first language attrition possible without second language knowledge? *IRAL Int. Rev. Appl. Linguist. Lang. Teach.* 56, 103–136. doi: 10.1515/iral-2016-0066
- Ben Rafael, M. (2001). *Contact de Langues: Le Français Parlé des Francophones Israéliens*. Ph.D. dissertation. Tel Aviv: University of Tel Aviv.
- Bergmann, C., Meulman, N., Stowe, L. A., Sprenger, S. A., and Schmid, M. S. (2015a). Prolonged L2 immersion engenders little change in morphosyntactic processing of bilingual natives. *Neuroreport* 26, 1065–1070. doi: 10.1097/WNR.0000000000000469
- Bergmann, C., Nota, A., Sprenger, S. A., and Schmid, M. S. (2016). L2 immersion causes non-native-like L1 pronunciation in German attriters. *J. Phon.* 58, 71–86. doi: 10.1016/j.wocn.2016.07.001
- Bergmann, C., Sprenger, S. A., and Schmid, M. S. (2015b). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychol.* 161, 25–35.
- Bialystok, E. (2017). The bilingual adaptation: how minds accommodate experience. *Psychol. Bull.* 143, 233–262. doi: 10.1037/bul0000099
- Bialystok, E., Craik, F. I. M., and Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends Cogn. Sci.* 16, 240–250. doi: 10.1016/j.tics.2012.03.001
- Calabria, M., Costa, A., Green, D. W., and Abutalebi, J. (2018). Neural basis of bilingual language control. *Ann. N. Y. Acad. Sci.* 1426, 221–235. doi: 10.1111/nyas.13879
- Canada Census Program (2011). Available online at: <http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-314-x/98-314-x2011001-eng.cfm> (accessed April 25, 2020).
- Chamorro, G., Sorace, A., and Sturt, P. (2016a). What is the source of L1 attrition? The effect of recent L1 re-exposure on Spanish speakers under L1 attrition. *Bilingualism* 19, 520–532.
- Chamorro, G., Sturt, P., and Sorace, A. (2016b). Selectivity in L1 attrition: differential object marking in Spanish near-native speakers of English. *J. Psycholinguist. Res.* 45, 697–715. doi: 10.1007/s10936-015-9372-4
- Coderre, E. L. (2015). “The neuroscience of bilingualism: cross-linguistic influences and cognitive effects,” in *Neuroscience in Intercultural Contexts*, eds J. E. Warnick and D. Landis (New York, NY: Springer), 73–116. doi: 10.1007/978-1-4939-2260-4_4
- Costa, A., and Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nat. Rev. Neurosci.* 15, 336–345. doi: 10.1038/nrn3709
- de Bot, K., Gommans, P., and Rossing, C. (1991). “L1 loss in an L2 environment: Dutch immigrants in France,” in *First Language Attrition*, eds H. W. Seliger and R. M. Vago (Cambridge: Cambridge University Press), 87–98. doi: 10.1017/cbo9780511620720.006
- de Leeuw, E., Mennen, I., and Scobbie, J. M. (2012). Singing a different tune in your native language: first language attrition of prosody. *Int. J. Biling.* 16, 101–116. doi: 10.1177/1367006911405576
- de Leeuw, E., Schmid, M. S., and Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism* 13, 33–40. doi: 10.1017/s1366728909990289
- de Leeuw, E., Tusha, A., and Schmid, M. S. (2018). Individual phonological attrition in Albanian-English late bilinguals. *Bilingualism* 21, 278–295. doi: 10.1017/S1366728917000025
- DeLuca, V., Rothman, J., Bialystok, E., and Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proc. Natl. Acad. Sci. U.S.A.* 116, 7565–7574. doi: 10.1073/pnas.1811513116
- Dijkstra, T., and Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Bilingualism* 5, 175–197. doi: 10.1017/s1366728902003012
- Dussias, P. E. (2004). Parsing a first language like a second: the erosion of L1 parsing strategies in Spanish-English bilinguals. *Int. J. Biling.* 8, 355–371.
- Eurobarometer Report “Europeans and their languages” (2012). Available online at: <https://epale.ec.europa.eu/en/resource-centre/content/eurobarometer-europeans-and-their-languages-2012> (accessed April 25, 2020).
- Gallo, F., Novitskiy, N., Myachikov, A., and Shtyrov, Y. (2021). Individual differences in bilingual experience modulate executive control network and performance: behavioral and structural neuroimaging evidence. *Bilingualism* 24, 293–304. doi: 10.1017/S1366728920000486
- Gargiulo, C., and van de Weijer, J. (2018). *Recovery effects in L1 attrition: a study on anaphora resolution in Italian*. Abstract from UIC Bilingualism Forum 2018. Chicago, IL: University of Illinois at Chicago.
- Grosjean, F. (ed.) (2010). *Bilingual*. Cambridge, MA: Harvard University Press.
- Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., et al. (2018). Bilingual experience and resting-state brain connectivity: impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia* 117, 123–134. doi: 10.1016/j.neuropsychologia.2018.04.037
- Hayakawa, S., and Marian, V. (2019). Consequences of multilingualism for neural architecture. *Behav. Brain Funct.* 1:6. doi: 10.1186/s12993-019-0157-z
- Hernandez, A., Li, P., and MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends Cogn. Sci.* 9, 220–225. doi: 10.1016/j.tics.2005.03.003
- Hernandez, A. E. (2013). *The Bilingual Brain*. Oxford: Oxford University Press.
- Hernandez, A. E., Claussenius-Kalman, H. L., Ronderos, J., Castilla-Earls, A. P., Sun, L., Weiss, S. D., et al. (2019a). Neuroemergentism: a framework for studying cognition and the brain. *J. Neurolinguist.* 49, 214–223. doi: 10.1016/j.jneuroling.2017.12.010
- Hernandez, A. E., Claussenius-Kalman, H. L., Ronderos, J., Castilla-Earls, A. P., Sun, L., Weiss, S. D., et al. (2019b). Neuroemergentism: response to commentaries. *J. Neurolinguist.* 49, 258–262. doi: 10.1016/j.jneuroling.2018.06.001
- Hutz, M. (2004). “Is there a natural process of decay? A longitudinal study of language attrition,” in *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, eds M. S. Schmid, B. Köpke, M. Keijzer, and L. Weilemar (Amsterdam: John Benjamins Publishing), 189–206. doi: 10.1075/sibil.28.11hut
- Isurin, L. (2007). Teachers’ Language: L1 Attrition in Russian-English Bilinguals. *Mod. Lang. J.* 91, 357–371. doi: 10.1111/j.1540-4781.2007.00585.x
- Karayayla, T., and Schmid, M. S. (2019). First language attrition as a function of age at onset of bilingualism: first language attainment of Turkish-English Bilinguals in the United Kingdom. *Lang. Learn.* 69, 106–142. doi: 10.1111/lang.12316
- Kasparian, K., and Steinhauer, K. (2016). Confusing similar words: ERP correlates of lexical-semantic processing in first language attrition and late

AUTHOR CONTRIBUTIONS

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

FUNDING

The present study has been supported by Russian Science Foundation Grant (Project No. 19-18-00550) to the HSE University.

- second language acquisition. *Neuropsychologia* 93, 200–217. doi: 10.1016/j.neuropsychologia.2016.10.007
- Kasparian, K., and Steinhauer, K. (2017a). On missed opportunities and convenient “truths.” *Linguist. Approaches Biling.* 7, 709–714. doi: 10.1075/lab.00010.kas
- Kasparian, K., and Steinhauer, K. (2017b). When the second language takes the lead: neurocognitive processing changes in the first language of adult Attriters. *Front. Psychol.* 8:389. doi: 10.3389/fpsyg.2017.00389
- Kasparian, K., Vespignani, F., and Steinhauer, K. (2017). First Language Attrition Induces Changes in Online Morphosyntactic Processing and Re-Analysis: an ERP Study of Number Agreement in Complex Italian Sentences. *Cogn. Sci.* 41, 1760–1803. doi: 10.1111/cogs.12450
- Köpke, B. (2004). “Attrition is not a unitary phenomenon: on different possible outcomes of language contact situations,” in *Bilingual Socialization and Bilingual Language Acquisition: Proceedings from the Second International Symposium on Bilingualism*, eds A. M. L. Suarez, F. Ramallo, and X.-P. Rodriguez-Yanez (Vigo: Servizo de Publicacions da Universidade de Vigo), 1331–1347.
- Köpke, B., and Genevskaya-Hanke, D. (2018). First language attrition and dominance: Same or different? *Front. Psychol.* 9:1963. doi: 10.3389/fpsyg.2018.01963
- Köpke, B., and Schmid, M. S. (2004). “Language attrition: the next phase,” in *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, eds M. Schmid, B. Köpke, M. Keijzer and L. Weilemar (Amsterdam: John Benjamins), 1–43.
- Kroll, J. F., Dussias, P. E., Bice, K., and Perrotti, L. (2015). Bilingualism, mind, and brain. *Annu. Rev. Linguist.* 1, 377–394. doi: 10.1146/annurev-linguist-030514-124937
- Kroll, J. F., Dussias, P. E., Bogulski, C. A., and Kroff, J. R. V. (2012). Juggling two languages in one mind: what bilinguals tell us about language processing and its consequences for cognition. *Psychol. Learn. Motiv.* 56, 229–262.
- Laufer, B., and Baladzhaeva, L. (2015). First language attrition without second language acquisition. *ITL Int. J. Appl. Linguist.* 166, 229–253. doi: 10.1075/itl.166.2.02lau
- Li, P., Legault, J., and Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. *Cortex* 58, 301–324. doi: 10.1016/j.cortex.2014.05.001
- Major, R. C. (2010). First language attrition in foreign accent perception. *Int. J. Biling.* 14, 163–183. doi: 10.1177/1367006910363063
- Miller, D., and Rothman, J. (2020). You win some, you lose some: comprehension and event-related potential evidence for L1 attrition. *Bilingualism* 23, 1–15. doi: 10.1017/S1366728919000737
- Opitz, C. (2013). A dynamic perspective on late bilinguals’ linguistic development in an L2 environment. *Int. J. Biling.* 17, 701–715. doi: 10.1177/1367006912454621
- Ribbert, A., and Kuiken, F. (2010). L2-induced changes in the L1 of Germans living in the Netherlands. *Bilingualism* 13, 41–48. doi: 10.1017/S1366728909990320
- Sabourin, L., and Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Lang. Res.* 24, 397–430. doi: 10.1177/0267658308090186
- Schmid, M. E., Köpke, B. E., Cherciov, M. C., Karayayla, T. C., Keijzer, M. C., De Leeuw, E. C., et al. (2019). *The Oxford Handbook of Language Attrition*. Oxford: Oxford University Press.
- Schmid, M. S. (2007). “The role of L1 use for L1 attrition,” in *Language Attrition. Theoretical Perspectives*, eds B. Köpke, M. S. Schmid, M. Keijzer, and S. Dostert (Amsterdam: John Benjamins Publishing), 135–153. doi: 10.1075/sibil.33.10sch
- Schmid, M. S. (2014). The debate on maturational constraints in bilingual development: a perspective from first-language attrition. *Lang. Acquis.* 21, 386–410. doi: 10.1080/10489223.2014.892947
- Schmid, M. S., and Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: the impact of extralinguistic factors. *Second Lang. Res.* 26, 125–160. doi: 10.1177/0267658309337641
- Schmid, M. S., and Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism* 17, 729–748. doi: 10.1017/S1366728913000771
- Schmid, M. S., and Köpke, B. (2007). “Bilingualism and attrition,” in *Language Attrition. Theoretical Perspectives*, eds B. Köpke, M. S. Schmid, M. Keijzer, and S. Dostert (Amsterdam: John Benjamins Publishing), 1–7. doi: 10.1075/sibil.33.02sch
- Schmid, M. S., and Köpke, B. (2017a). The relevance of first language attrition to theories of bilingual development. *Linguist. Approaches Biling.* 7, 637–667. doi: 10.1075/lab.17058.sch
- Schmid, M. S., and Köpke, B. (2017b). When is a bilingual an attriter? Response to the commentaries. *Linguist. Approaches Biling.* 7, 763–770. doi: 10.1075/lab.17059.sch
- Schmid, M. S., and Yilmaz, G. (2018). Predictors of language dominance: an integrated analysis of first language attrition and second language acquisition in late bilinguals. *Front. Psychol.* 9:1306. doi: 10.3389/fpsyg.2018.01306
- Smith, M. S. (2019). *Language Attrition as a Special Case of Processing Change: A Wider Cognitive Perspective*, 1st Edn, eds M. Schmid and B. Köpke (Oxford: Oxford University Press). doi: 10.1093/oxfordhb/9780198793595.013.7
- Sulpizio, S., Del Maschio, N., Del Mauro, G., Fedeli, D., and Abutalebi, J. (2020). Bilingualism as a gradient measure modulates functional connectivity of language and control networks. *Neuroimage* 205:116306. doi: 10.1016/j.neuroimage.2019.116306
- United Nations (2018). *International Migration Report 2017-Highlights*. New York, NY: UN.
- Ventureyra, V. A., Pallier, C., and Yoo, H. Y. (2004). The loss of first language phonetic perception in adopted Koreans. *J. Neurolinguist.* 17, 79–91. doi: 10.1016/S0911-6044(03)00053-8
- Yilmaz, G., and Schmid, M. S. (2012). L1 accessibility among Turkish-Dutch bilinguals. *Ment. Lex.* 7, 249–274. doi: 10.1075/ml.7.3.01yil

Conflict of Interest: 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 © 2021 Gallo, Ramanujan, Shtyrov and Myachykov. 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.



Integration of Social Context vs. Linguistic Reference During Situated Language Processing

Katja Maquate^{1*} and Pia Knoeferle^{1,2,3}

¹ Psycholinguistics, Institute for German Language and Linguistics, Humboldt-Universität zu Berlin, Berlin, Germany, ² Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany, ³ Einstein Center for Neurosciences Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Juhani Järviö,
University of Alberta, Canada

Reviewed by:

Andriy Myachkov,
Northumbria University,
United Kingdom
Elsi Kaiser,
University of Southern California,
United States
Arnout Koornneef,
Leiden University, Netherlands

*Correspondence:

Katja Maquate
katja.maquate@hu-berlin.de

Specialty section:

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

Received: 31 March 2020

Accepted: 16 June 2021

Published: 02 August 2021

Citation:

Maquate K and Knoeferle P (2021)
Integration of Social Context vs.
Linguistic Reference During Situated
Language Processing.
Front. Psychol. 12:547360.
doi: 10.3389/fpsyg.2021.547360

Research findings on language comprehension suggest that many kinds of non-linguistic cues can rapidly affect language processing. Extant processing accounts of situated language comprehension model these rapid effects and are only beginning to accommodate the role of non-linguistic emotional cues. To begin with a detailed characterization of distinct cues and their relative effects, three visual-world eye-tracking experiments assessed the relative importance of two cue types (action depictions vs. emotional facial expressions) as well as the effects of the degree of naturalness of social (facial) cues (smileys vs. natural faces). We predicted to replicate previously reported rapid effects of referentially mediated actions. In addition, we assessed distinct world-language relations. If how a cue is conveyed matters for its effect, then a verb referencing an action depiction should elicit a stronger immediate effect on visual attention and language comprehension than a speaker's emotional facial expression. The latter is mediated non-referentially via the emotional connotations of an adverb. The results replicated a pronounced facilitatory effect of action depiction (relative to no action depiction). By contrast, the facilitatory effect of a preceding speaker's emotional face was less pronounced. How the facial emotion was rendered mattered in that the emotional face effect was present with natural faces (Experiment 2) but not with smileys (Experiment 1). Experiment 3 suggests that contrast, i.e., strongly opposing emotional valence information vs. non-opposing valence information, might matter for the directionality of this effect. These results are the first step toward a more principled account of how distinct visual (social) cues modulate language processing, whereby the visual cues that are referenced by language (the depicted action), copresent (the depicted action), and more natural (the natural emotional prime face) tend to exert more pronounced effects.

Keywords: real-time situated language processing, eye-tracking, emotional priming, action depiction, natural facial expressions, schematic faces, social context

INTRODUCTION

Monitoring people's gaze behavior in a visual context provides a unique opportunity for examining the incremental integration of visual and linguistic information (Tanenhaus et al., 1995). During sentence comprehension, non-linguistic visual information can rapidly guide visual attention in adults (e.g., Sedivy et al., 1999; Spivey et al., 2002; Chambers et al., 2004; Knoeferle et al., 2005).

Crucially, non-linguistic information can also facilitate real-time language processing of canonical and non-canonical grammatical sentences (e.g., Knoeferle et al., 2005; Carminati and Knoeferle, 2013). Social cues, such as, for example, a speaker's emotional facial expression (Carminati and Knoeferle, 2013), a speaker's gaze shift (Kreysa et al., 2018), or the speaker's voice information (Van Berkum et al., 2008), can elicit expectations on the part of the listener (just like other non-linguistic cues), and these can in turn influence the processing of upcoming linguistic information. However, existing research has focused mostly on assessing how object- and action-related visual information influences spoken language comprehension. By contrast, little is known about the effect of social (visual) cues (e.g., an emotional facial expression) on real-time sentence comprehension and which degree of naturalness (and corresponding degree of detail) is needed for comprehenders to exploit them. Additionally, we do not know how far and to which extent (schematic vs. natural) facial emotions and action events relative to one another modulate visual attention and language comprehension. Examining these open issues can further clarify our understanding of the integration of distinct visual cues into language processing. This clarification can in turn help us to refine models of real-time language processing, taking the visual and social context into account.

In most serial language comprehension accounts (e.g., Frazier and Fodor, 1979; Friederici, 2002), contextual representations are depleted and come into play very late during language processing. Parallel interactive theories, by contrast, do not restrict the interaction of information (see also Anderson et al., 2011) and emphasize a rapid interplay between syntactic and non-syntactic representations (e.g., MacDonald et al., 1994; Trueswell and Tanenhaus, 1994). Yet, these constraint-based approaches neither feature interpretation building processes nor non-linguistic representations (see also Novick et al., 2008).

Real-time language processing accounts have largely focused on the integration of visual cues, such as depicted actions and objects, that are referenced by the linguistic input. For instance, the coordinated interplay account (CIA; Knoeferle and Crocker, 2006, 2007) comprises three processing steps. These steps are temporally dependent and can overlap or occur in parallel. In the first step, the linguistic input is interpreted incrementally on the basis of existing knowledge. In the second step, expectations and representations are built and guide attention to relevant information in working memory or a visual scene, e.g., depicting actions and objects. In the final step, previously built interpretations and expectations are reconciled with the scene/working memory representations.

By contrast, we know less about the extent to which further visual cues—that are non-referentially linked to language—impact comprehension [but see, e.g., Guerra and Knoeferle (2014, 2017, 2018) on enriching the CIA with non-referential conceptual co-indexing mechanisms; Altmann and Kamide (1999), Huettig and Altmann (2005), and Altmann and Trafton (2002) for non-referential world knowledge effects on language processing]. Consider, for instance, a speaker's smile that a listener might (or might not) relate to the valence of words in a sentence. Indeed, recent work (Münster and Knoeferle, 2018)

has started to extend situated language processing accounts with the biological and experiential properties of the comprehender, as well as with social contextual (visual) information that is non-referential [the (social) CIA (sCIA), see Münster and Knoeferle (2018) also for a more detailed review on how/whether different processing accounts deal with non-linguistic social representations; see also Van Berkum (2018, 2019)]. However, to more fully accommodate how distinct cues contribute toward human language processing, empirical research and these accounts must consider not only the effects of individual cues but also how the effects of (distinct) visual cues measure up against one another. The present research compares the effects of action event depiction with those of (natural vs. schematic) emotional facial expressions (for which the link between the visual and linguistic input is not referential and arguably subtler).

World-Language Relations in Sentence Comprehension: Referential vs. Non-referential Cues

Objects and Actions as Referential Cues

Adults can rapidly use information about objects and depicted action events for disambiguating structurally ambiguous sentences when these cues are referenced and made available by (words in) the utterance. For instance, *kick* refers to a kicking event and makes available the kicking action and knowledge of plausible associated agents such as soccer players. *On the towel* can refer to a location and make available the referent situated at that location. In a real-world study, adults inspected an apple, an apple on a towel, an empty towel, and a box and listened to sentences like *Put the apple on the towel in the box*. Before listeners heard *in the box*, they preferred to interpret *on the towel* as a destination for the apple. However, in a context with two apples, the need to distinguish between them guided participants toward resolving *on the towel* as a modifier of *the apple*, interpreting it as its location: participants quickly looked at the apple on the towel during *on the towel* (location), and not at the empty towel (as a destination) to which the apple could be moved (Tanenhaus et al., 1995; Spivey et al., 2002).

In addition to noun-object relations, adults can use other referential cues, such as verb-mediated depicted action events, to facilitate role assignments, and thus the processing of canonical subject-verb-object (SVO) and noncanonical object-verb-subject (OVS) German sentences [both word orders are grammatical in German but OVS is non-canonical (Knoeferle et al., 2005)]. In a visual-world eye-tracking study, participants inspected clipart scenes depicting a princess as washing a pirate and as being painted by a fencer. The spoken sentence played during scene inspection was initially ambiguous and either related to the princess-washes-pirate event (in SVO order) or the princess-is-painted-by-fencer event (in OVS word order). Shortly after the verb had modulated one of the two depicted actions, participants either visually anticipated the associated pirate (if they had heard *washes*) or the fencer (if they had heard *paints*). From the anticipation of the action's patient (the pirate in SVO sentences) or agent (the fencer in OVS sentences), the authors deduced that listeners had assigned a thematic role to the initially

role ambiguous noun phrase *the princess*. Thus, comprehenders can rapidly exploit referential cues (*on the towel* identifying a location; *paints* referencing a painting action and mediating its associated agent) for language processing and the assignment of thematic roles.

Non-referential (Visual) Social Cues: Facial Emotions

Would comprehenders also benefit from non-referential visual cues? By “non-referential” (visual) cues, we mean the (visual) information that listeners might associate with language but that is not referentially mediated. The listener has to infer and interpret the relationship between the non-linguistic (visual) cue and the linguistic input in a non-referential way (i.e., hearing *Nice to meet you!* and seeing someone smile), rather than identifying the referential link between a visual cue and a mediating linguistic expression (i.e., hearing *kick* and seeing someone kicking something). A non-referential (visual) cue, such as an emotional facial expression, provides additional non-linguistic information, which could be exploited in order to facilitate linguistic processing and interpretation although it is an interesting issue whether comprehenders can exploit it to the same extent given the non-referential link with language.

Human faces, despite sharing general features, differ greatly in their detailed features (Grelotti et al., 2002). Yet, most people can effortlessly discriminate faces based on those detailed features, making us experts in face recognition (Diamond and Carey, 1986). Becoming an expert in the recognition and processing of faces allows us to interact and communicate with each other (Grelotti et al., 2002). Building this expertise already starts in the earliest moments of life: even newborns, only minutes after birth already attend to faces more than to non-face-like stimuli (Johnson et al., 1991; Mondloch et al., 1999). During communication, we use our face to (consciously or unconsciously) convey a nonverbal message alongside our verbal message. In turn, the listener interprets our facial expression and tries to integrate it into the unfolding interpretation to correctly understand and interpret it or even to facilitate sentence processing (Carminati and Knoeferle, 2013). Emotional priming studies, for instance, show that the valenced positive and negative primes can facilitate and/or speed up the processing and recognition of emotionally congruent subsequent targets [see, e.g., Hermans et al. (1994) and Lamy et al. (2008)].

For example, in a reaction time experiment, Aguado et al. (2007) used faces as primes and words as targets. Participants first saw a positive or a negative prime face followed by either a positive or negative target word or a question mark. If the target word appeared, participants had to judge the valence of the word. If the question mark appeared, the task was to detect the gender of the previously seen positive or negative prime face. Participants did not know in advance whether they had to detect the gender of the face or judge the valence of the word, rendering the task unpredictable. The results were in line with classic priming effects: reaction times were shorter for valence-congruent (vs. incongruent) face-word trials.

Crucially, social visual cues, such as emotional faces, can also affect sentence interpretation. In one study (Münster et al., 2014; Carminati and Knoeferle, 2016), participants inspected a video of

a human emotional facial expression. After this speaker's prime face, a new scene appeared showing two event photographs and participants heard a (positively or negatively) valenced sentence related to one of these photographs. The issue was whether a match (vs. mismatch) in the valence of a preceding speaker's face and the valence of the ensuing sentence would boost participants' visual attention to the valence-matching photograph and thus facilitate their sentence comprehension. To experience facilitation, participants had to link the (e.g., positive) valence of the preceding face to the (positive) valence of the ensuing spoken sentence, resulting in a boost of visual attention to a related event photograph. Thus, both links between language and the facial expression were non-referential and the temporal contiguity of the visual cue was less (preceding the target utterance) than for the visual cues examined in a few previous studies [e.g., the action depictions were copresent as comprehenders listened to the utterance in Knoeferle et al. (2005); see Spivey and Geng (2001) and Altmann (2004) on effects in the blank screen paradigm, in which a stimulus sentence is heard after a visual scene had been inspected and removed from the screen; eye movements in the blank screen were measured in response to the sentence].

In spite of these more tenuous world-language links, having seen a smiling/sad speaker face facilitated participants' visual attention and processing of emotionally valenced (positive/negative) canonical SVO sentences (Münster et al., 2014; Carminati and Knoeferle, 2016). The emotional facial expressions were integrated incrementally with the linguistic input and modulated its processing online, again in the absence of referential links. Interestingly, similar effects emerged for static emotional facial expressions (Carminati and Knoeferle, 2013). These findings suggested for the first time that (static and dynamic) facial expressions—like actions—can incrementally modulate adults' processing of emotional sentences. The emotion effects emerged despite the differences in how language conveyed these cues (valence associations vs. verb-action reference) and despite the fact that the speaker's face was not present during comprehension (and thus arguably less accessible). To which extent these findings extend to the processing of other, in particular difficult-to-process non-canonical sentences (e.g., German object-initial sentences) is, however, an open issue.

It is also unclear to which extent the portrayal of emotions (as dynamic human faces or as schematic smileys) matters for emotion effects on language processing. Considering emotional facial expressions, most of the time we interact with other human beings and easily attribute mental states, beliefs, and feelings to our interaction partners, based on our own mental states (“Theory of Mind;” Premack and Woodruff, 1978). We are thus experienced in our interaction with natural human emotional faces. However, research on emotional face recognition has also used computer-generated schematic faces and the evidence suggests that the latter are recognized as well as natural faces (e.g., Öhman et al., 2001; Chang, 2006; Ruffman et al., 2009). ERP (Event-Related Potential) and behavioral research (Schindler et al., 2017; Kendall, 2019; Zhao et al., 2019) suggests that (emotional) cartoon faces are recognized faster and might be analyzed more on a structural level compared with natural human faces (as indexed by shorter RT (Reaction Times), briefer

N170 ERP latencies, and larger N170 amplitudes). By contrast, natural (vs. cartoon) faces are processed more holistically and require more attentional resources during later processing stages [as indexed by larger late positive potential (LPP) amplitudes].

Whether schematic (vs. natural human) facial emotions would yield comparable effects also for real-time visual attention and language processing is, by contrast, an open issue. Is a schematic expression sufficient (e.g., as in smileys, where emotion is stripped down to its bare essential, perhaps rendering valence salient), or do emotional priming effects on online sentence comprehension emerge only following more realistic, detailed, and natural emotional faces?

Toward Differentiating World-Language Relations

The present research compared the effects of two distinct cues (referentially mediated actions and their associated agents with non-referential facial expressions) within a single study, and in addition manipulated the degree of naturalness of the facial expression on incremental sentence processing within a single study, and in addition manipulated the degree of naturalness of the facial expression.

Extant research has begun to compare the influence of referential (object depiction) and non-referential (speaker gaze shift) cues on sentence processing (Kreysa et al., 2014, 2018). In a visual-world eye-tracking study, participants inspected the videos of a speaker uttering German sentences about two virtual characters (translated, e.g., *The waiter congratulates the millionaire in the afternoon* with a Second Life display showing a saxophonist, waiter, and millionaire). The action was (vs. was not) depicted and the speaker either shifted gaze between the characters referred to in the sentence or was obscured, yielding four conditions (neither gaze nor action was present; only either gaze or the action was present; and both of these cues were present). Both cues appeared simultaneously, just after the onset of the verb (the speaker shifted gaze to the millionaire and the action tool appeared between the waiter and the millionaire). Listeners used both cues to anticipate the upcoming patient of the sentence (the millionaire) before its mention. The speaker gaze cue enabled anticipation reliably earlier than the action cue but only when the action was used non-deictically (Kreysa et al., 2018; Experiment 2). When both action depiction and gaze were used deictically, their effects on visual attention and comprehension were comparable. Two cues did not seem to be more helpful than one when the action was used non-deictically.

Due to the diverse nature of the different kinds of cues, we do not yet know if distinct language-world relations ease utterance interpretation to the same extent and in a similar fashion. It could be that the speaker gaze is so effective because it is dynamic and present during comprehension, and the dynamic motion captures and guides listeners' attention. But emotional facial expressions might be as effective as gaze, permitting rapid anticipation: seeing, for example, our interlocutor smile likewise sets up expectations as to what might come next. These might be expectations about a matching emotionally positive surrounding situation. Moreover, it likely fosters the expectation that the upcoming utterance is also positive in emotional valence. Both speaker gaze and a speaker's emotional facial expression raise

expectations; we can link these cues to linguistic material in an utterance matching these expectations and could direct attention to relevant parts in a visual scene.

Carminati and Knoeferle (2013) provide some evidence for rapid effects of a preceding emotional speaker's face for the subsequent processing of at least German subject-initial sentences (and this despite the fact that the speaker's face was not co-present during comprehension). Seeing someone smile and hearing an emotionally positive linguistic expression, such as *happy*, does not foster a referential link (as between an action verb and a perceived action): the hearer first has to recognize and interpret the emotional facial expression, likely resulting in the activation of a representation of the concept of happiness. This might set up expectations regarding the emotionality of the situation between a speaker and a hearer. When the emotionally positive adverb *happily* is encountered, this concept has to be linked in a non-referential way to the encountered linguistic expression. Then, attention can be directed to the visual input in a scene, e.g., seeing another person, such as the agent of an action, smile. Hence, even though there is a link between a speaker's smile and an associated linguistic expression, this link is not referential and arguably more complex than a referential link. Had the linguistic expression been *cheerful* or *friendly* instead of *happily*, a very similar or even identical link to the link between a smile and the word *happily* could have been established.

Hearing an action verb (e.g., *kick*), on the other hand, directs the hearer's attention *via* a referential link to a depiction of the heard action verb (e.g., a man standing on a field who is stretching out one leg in a kicking action). The world-language link is referential because no intermediate processing steps, such as forming non-referential conceptual representations, interpreting these representations in the present situation, and relating them to the perceived action, have to be performed.

As a few previous studies investigating referential and non-referential world-language relations suggest (see Sections Objects and Actions as Referential Cues—Toward Differentiating World-Language Relations), expectations are set up and attention is directed *via* these links to relevant parts in a visual scene when both referential and non-referential links can be established. Perhaps then we will see no difference in the effects of emotional facial cues and actions? Alternatively, the actions are referential (and present during comprehension), and could hence elicit stronger effects than facial cues that are (non-referentially) related to the emotional valence of, for instance, sentential adverbs.

Please note that our aim was neither to test verb cues vs. emotional cues nor to generalize across all referential vs. non-referential cues. Instead, the goal of the present research was to determine how emotional facial expressions (as one specific example of non-referential cues) and depicted action events (as one specific example of referential cues) affect online sentence comprehension. To what extent and in what way do these cues interact with each other, and to what extent does the naturalness of the cue (the emotional facial expression) matter? We acknowledge that other referential and non-referential cues might differ from emotional facial expressions and depicted actions in the way and the degree in which they link to

language. Yet, as there is no prior research in this domain (that we are aware of), we chose emotional facial expressions and depicted actions as cues to maximize the difference between referential and non-referential relations and because these cues have already been shown to affect language processing on their own (cf., Knoeferle et al., 2005; Carminati and Knoeferle, 2013). Examining these issues will provide further insights into the relative effect of distinct kinds of cues on language processing. Models of language processing, such as the (social) CIA (CIA; Knoeferle and Crocker, 2007; sCIA: Münster and Knoeferle, 2018), are underspecified regarding the relative integration of different kinds of extralinguistic cues into language processing since empirical evidence is lacking. They are also underspecified regarding effect-differences for natural compared with schematic cues (e.g., facial expressions). The results of the present research can thus inform their extension.

Three visual-world eye-tracking studies compared the effects of action event cues with those of a speaker's emotional facial expression as a prime (Experiments 1 and 2) and manipulated naturalness of the facial expressions across experiments (Experiments 1–3). To further examine whether facilitative effects of the emotion cues (Carminati and Knoeferle, 2013) extend to other sentence structures and processes of the assignment of thematic roles, we employed noncanonical (but grammatical) German OVS sentences. We know that the actions facilitate OVS sentence comprehension and examine whether the effects of emotional facial stimuli previously only attested for SVO sentences would generalize and be comparable in their effects. Experiment 1 examined the effects of schematic facial expressions and Experiment 2 of natural facial expressions on the assignment of thematic roles during the comprehension of spoken German OVS sentences. Experiment 3 further investigated the effect of natural facial expressions in the absence of depicted action events and set a stronger focus on language processing situated in a more salient emotionally valenced environment.

EXPERIMENT 1

Methods

Participants

40 students of University of Bielefeld between 18 and 30 years (14 male, mean age: 24, SD age: 3.09), all native speakers of German, took part in the experiment. Participants were tested in the eye-tracking laboratory of Bielefeld University. Sample size was set to 40 to ensure comparability with a related study with children as participants. Each participant received 4 Euro for participation and gave written informed consent. The university's ethics board approved the study (Vote 2013-007). The experimental session took about 30 min.

Materials

The design crossed *emotional prime* (prime valence congruous vs. incongruous with the sentence) with *action* (present vs. absent). We realized the first factor of the design *via* prime images (a yellow smiley vs. a red star) constructed by using commercially available software (see **Table 1**). The smiley changed dynamically




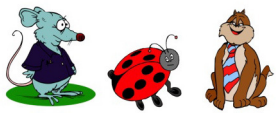

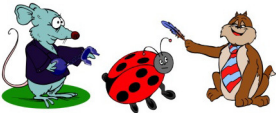

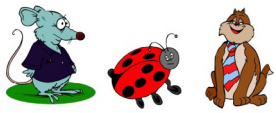
from a light and subtle to a broad smile. The red star was static and had no facial features. The smiley matched the target sentences in valence (see **Table 1**) while the red star was incongruous in that it conveyed no emotional valence *via* facial features. We avoided negative emotional primes for consistency with a planned child study.

The second independent factor was realized *via* the target scenes ($N = 16$). These were created by using Adobe Illustrator and commercially available clipart. Most of the clipart characters were animals, some humans (i.e., three human target agents, one human patient/middle character). Each scene consisted of three clipart characters and either depicted actions (see **Table 1**, A/C) or not (**Table 1**, B/D). The middle character was always the patient of the action performed by the outer characters. Only one of the actions performed by the outer characters was mentioned in the target sentence (see below) and its agent was the target agent; the other outer character performed an action, which is not mentioned in the sentence (competitor). The target agent (only) portrayed a happy facial expression (matching in valence with the prime smiley). The patient had a neutral and the competitor had a slightly negative facial expression. To counterbalance the position of the agent and competitor, we created a mirrored version of each experimental target scene. In one version of a target scene, the agent was thus on the right-hand side of the picture and in the other it was on the left.

For the experimental target sentences, we constructed 16 unambiguous noncanonical OVS sentences in German [e.g., *Den Marienkäfer*_{NP1} [masculine accusative case, patient] *kitzelt*_{action verb} *vergnügt*_{adverb} [positive emotional valence] *der Kater*_{NP2} [nominative case, target agent]], transl.: “The ladybug (acc. obj., patient) tickles happily the cat (nom. subj., agent)”, see **Supplementary Material** and Appendix in Münster (2016) for materials]. A female speaker (PK) recorded the sentences with neutral intonation and at a slow but natural sounding speed. Word region onsets and offsets were marked for later analyses. In addition to the experimental target sentences, the same speaker recorded comprehension questions in the active or passive voice, asking either for the agent or the patient of the sentence (e.g., “Who is doing [previously named action] to [previously named patient]” and passive questions in the fashion “Who is being [previously named action]?”).

In addition to the experimental items, we also constructed filler items ($N = 28$). These comprised filler sentences in either an unambiguously case-marked SVO ($N = 24$) sentence structure or an unambiguously case-marked OVS ($N = 4$) structure, recorded by PK. Some filler sentences had neutral verbs and adverbs ($N = 12$, thereof the four OVS sentences) and some were positively valenced (16 SVO sentences). The corresponding 28 filler pictures consisted of clipart animals and humans. Some always depicted three ($N = 12$) and others two characters ($N = 16$). The filler characters were positioned such that the interacting characters faced each other or looked away from one another; such that the agent faced the competitor character; or such that they faced the participant. This was done to prevent participants from developing a strategy as to who will be interacting with whom. Characters had a positive facial expression when the sentence was positive ($N = 16$). When the sentence was neutral

TABLE 1 | Experimental conditions for Experiment 1.

Condition	Prime	Action	Sentence
A	Happy yellow smiley, congruent prime 	Action depicted 	<i>Den Marienkäfer kitzelt vergnügt der Kater</i> [“The ladybug (accusative object, patient) tickles happily the cat (nominative subject, agent)”]
B	Happy yellow smiley, congruent prime 	No action depicted 	
C	Red star, incongruent prime 	Action depicted 	
D	Red star, incongruent prime 	No action depicted 	

($N = 12$), their facial expressions were also neutral or slightly negative. Half of the filler scenes depicted the action mentioned in the sentence ($N = 14$) while the other half depicted no actions ($N = 14$).

Pretests

We pretested the characters and actions to ensure that participants can recognize them. Moreover, we tested the valence of the emotional adverbs. Since we planned to conduct future child language studies using the same materials, we pretested the stimuli with a sample of 4–5-year-old children ($N = 20$, mean age: 4.8). About 10 children were asked in German to point to the agent and patient characters and the actions of the experimental scenes when an experimenter named them (transl.: *Who is the cat? Who is tickling the ladybug here?*). Character naming trials (presented in the no-action condition) and action naming trials (presented in the depicted action condition) were blocked. In this way, the characters were named before participants identified the character performing the action. The children identified the characters (96.9%) and the actions (88.5%) accurately. Ten additional children were asked to identify the happily acting (target) agent (transl.: e.g., *Who tickles happily the girl?*). For this second test, the target agent and the competitor character performed the same actions (i.e., unlike in our experimental pictures) but only the target agent smiled. Experimental scenes were mixed with filler pictures and sentences, which conveyed a negative or neutral valence. In 89.38% of the cases, children

reliably identified the happy agent and thus successfully linked the positive adverb to the happy target agent.

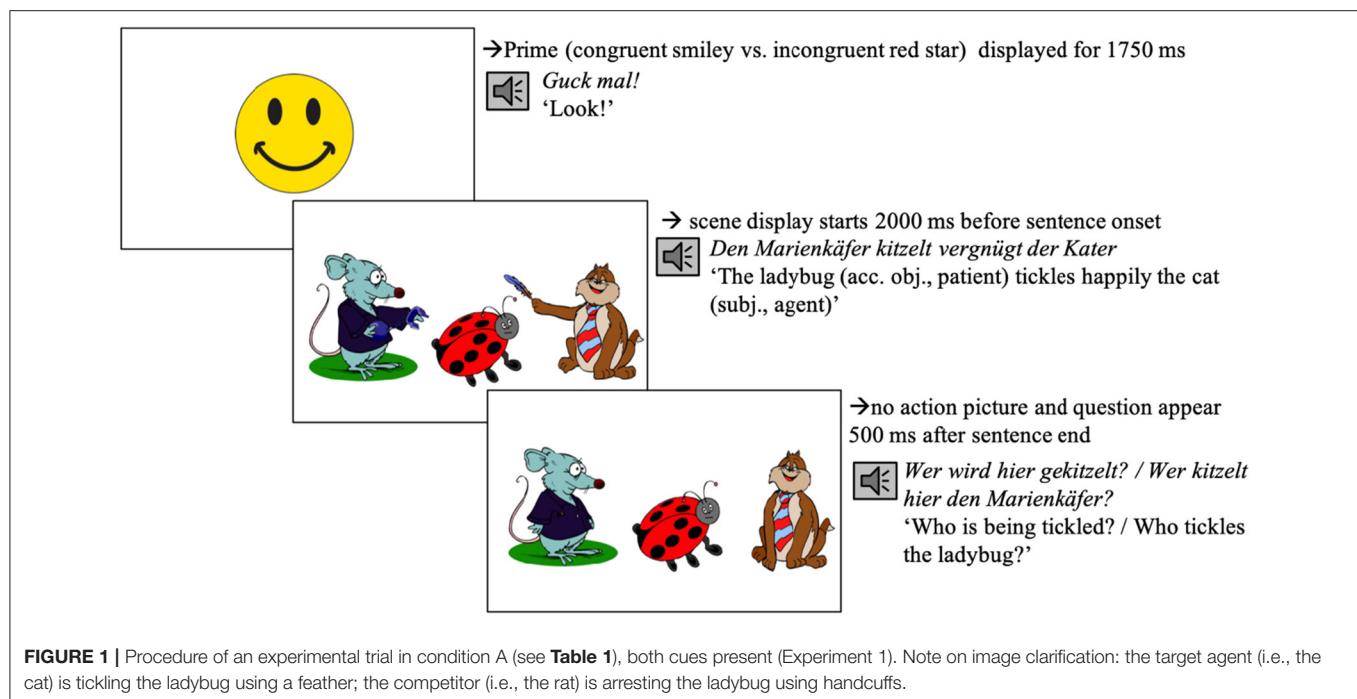
In summary, a 2 (congruent smiley prime vs. incongruent red star prime) \times 2 (action depiction vs. no action) design yielded four conditions (Table 1). The depiction of the prime and the action described by the sentence varied across conditions while the sentence was identical. We created eight lists such that each participant encountered all conditions but each sentence in only one of the four conditions (Table 1). These four lists were doubled to accommodate the mirrored character scenes (see Section Materials) yielding eight lists. Moreover, in each list, half of all comprehension questions were asked in the active and half in the passive voice, and each experimental item was followed equally often by active and passive questions. Each list contained all of the filler trials and was pseudorandomized for each participant. Two critical items never followed another.

Hypotheses
Accuracy

Adults can use case marking to reliably identify OVS word order (e.g., Kamide et al., 2003a, Experiment 3; Kamide et al., 2003b, Knoeferle et al., 2008). Thus, at sentence end, we did not predict significant effects of the prime and action manipulation on accuracy in the comprehension questions.

Eye Movements

Eye movements, by contrast, provide insights into real-time comprehension in the four conditions. We expected condition



differences to the extent that the different cues elicit distinct effects on visual attention and sentence comprehension. If two cues are better than one, then participants should look more and earlier toward the target agent (vs. the competitor), signaling anticipation of the correct role filler when both depicted action and prime smiley are present (vs. the single cue conditions). As the verb refers to the action, guiding the listeners' attention to the action and its associated agent, we predicted this cue to have a stronger effect than the smiley prime, which provides a non-referential link to the target agent. As a result, participants should look more toward the target agent (than the competitor) when only the action than when only the smiley was available. When the actions and smiley were absent, we predicted no clear fixation differences between the target agent and the competitor.

Procedure

All participants first read a participant information sheet and gave written informed consent. They were seated in front of the Eye tracker (Eye-Link 1000 Eye tracker, SR Research, Ontario, Canada) in the remote setup and asked to read the on-screen instructions. These instructions informed them that they would see a series of scene-sentence pairs. They were asked to concentrate on the scenes and to listen closely to the sentences. They were informed that they would have to answer a question about what they saw and heard after each trial. The experimental session started with a manual five-point calibration and validation procedure. Calibration and validation were repeated when necessary during the experiment. After successful calibration and validation, participants completed four practice trials. The experimenter advanced each trial manually after participants successfully fixated the black dot. The fixation

dot was followed by the presentation of the prime smiley video (duration: 1,750 ms, changing from a slight to a full smile after 250 ms), which was accompanied by the phrase *Guck mal!* (*Look*). That phrase served to focus participants' attention and was inserted with a view of planned developmental studies. After the prime, the target scene was previewed for 2,000 ms (**Figure 1**) after which the sentence started. 500 ms after the end of the sentence, the actions (if depicted) were removed from the scene and participants heard a comprehension question while looking at the no-action scene (**Figure 1**). Participants had no time limit and responded orally. After the participant had responded¹, the experimenter wrote the answer down and started the next trial. At the end of the experiment, participants were debriefed: They were asked to report what they thought the experiment was about; whether they noticed anything odd and/or any regularities; and whether they developed any strategies during the experiment.

Exclusion Criteria

If a participant had guessed the purpose of the experiment (nearly) correctly (i.e., "I think the experiment investigates how depicted actions and emotional facial expressions influence language processing"), the participant's data would have been excluded from the data set. This was, however, not the case for any of the participants. Additionally, the fixation data from all experiments was first manually inspected to see if all participants executed fixations to the prime and target scenes in a natural way. If, for instance, a participant had always fixated the middle of the screen or only always the character on the right side of the

¹Note that although reaction times would have been informative, we did not use this measure in Experiments 1 and 2 due to design consistency and comparability with a child language study.

screen, this participant would have been excluded. Since fixation patterns of all participants seemed to indicate natural fixations of the screens, no participant was excluded.

Analysis

Eye Movements

The eye-movement analyses included the data from all experimental trials (correctly and incorrectly answered), since the accuracy for determining thematic roles was at 96%. We divided the target scenes into a target agent and a competitor character and analyzed the real-time data from the target scene presentation onset until 500 ms after sentence offset. The item sentences were divided into individual analysis regions (see **Table 1**): NP1 (i.e., the patient is named), verb, adverb, a combined verb-adverb region to capture spillover effects, and an NP2 (i.e., target agent is named) region. Additionally, we computed a “long region,” spanning from NP1 onset until sentence offset plus 500 ms. Our critical time regions were the verb and adverb, since in the depicted action condition, the verb denotes the first region in which the agent of the sentence can unambiguously be determined. The adverb is the first region, which explicitly conveys linguistic emotional valence information. In the no-action condition, it denotes the first region in which sentence valence can be integrated with the emotional prime to anticipate the target agent based on its emotionally matching facial expression.

To exclude any prior preference in looks toward the agent vs. the competitor, we also analyzed the fixations during the NP1 region. To capture the effects of prime and action during the naming of the target agent, we also analyzed the NP2 + 500 ms region. Finally, extended effects across the sentence were analyzed by using the long region.

Fixations were measured by using the natural logarithm (based on the constant e) of the ratio of the probability of looking at the target agent over the probability of looking at the competitor character [$\ln(p(\text{agent})/p(\text{competitor}))$]. The log ratio is symmetrical around zero. This means, a positive value indicates a preference to look at the target agent over the competitor. A negative log ratio indicates a preference to look at the competitor over the target agent. A value of zero indicates no preference for either of the two characters. Since the log of zero is undefined, we added a constant of 0.1 to account for missing data points regarding fixations to both the agent and the competitor. Hence, this log probability ratio expresses the strength of the visual bias toward the target agent relative to the competitor character. Additionally, it has the advantage that it does not violate the assumptions of independence and homogeneity of variance (Arai et al., 2007).

For visual presentation, we plot time course graphs as a function of prime and action depiction (**Figures 2, 5**) and as a function of prime (**Figure 10**) using the mean log gaze probability ratios calculated on successive 20-ms time slots. For the inferential analyses, the log ratios were subjected to linear mixed-effects models [using *lmer* of the *lme4* package of R (Bates et al., 2015b)] with action (no action vs. depicted action), prime (congruent vs. incongruent) as fixed factors, and participants and items as random intercepts. All factors were centered (to

avoid collinearity) and sum coded. We included random slopes for action and prime in the participant and item random effect structures and, following (Bates et al., 2015a), we are reporting the results for the best-fitting (most parsimonious) models. The syntax for the best-fitting models for each analysis is reported in footnotes. We obtained the best-fitting models by reducing the random effect structure, starting with the maximal model [$\log_ratio \sim \text{action}*\text{prime} + (1+\text{action}*\text{prime} \mid \text{participant}) + (1+\text{action}*\text{prime} \mid \text{item})$]. The fixed effect structure of the model was not reduced. We calculated the values of p using the *lmerTest* package [Kuznetsova et al. (2017), i.e., Satterthwaite degrees of freedom method (cf., Luke, 2017)].

Accuracy Data

Accuracy was computed on the correct and incorrect comprehension-question responses for experimental trials ($N = 640$). To analyze accuracy, we ran generalized linear mixed-effects models in R [R Core Team, 2021; *glmer* function in the *lme4* package (Bates et al., 2015b)]. All models used emotional prime (congruent vs. incongruent) and action depiction (depicted action vs. no action) as fixed factors and subjects and items as random intercepts. Prime and action were included as random slopes into the subject and item random effect structure. Random effect structure selection followed the same procedure as for the eye-tracking data analyses. Question voice (active vs. passive) was used as an additional fixed factor and was likewise a factor in the random slopes. In all models, “family” was set to “binomial” due to the categorical nature of the accuracy scores.

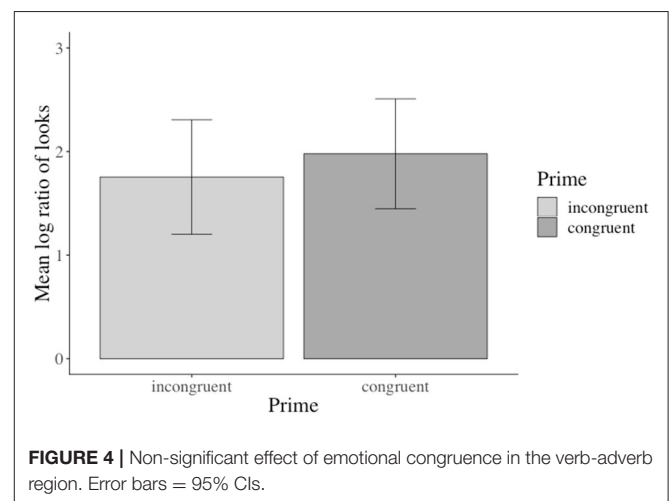
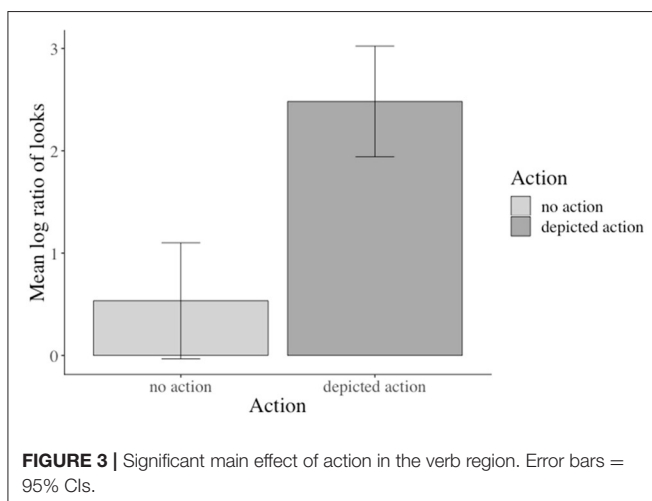
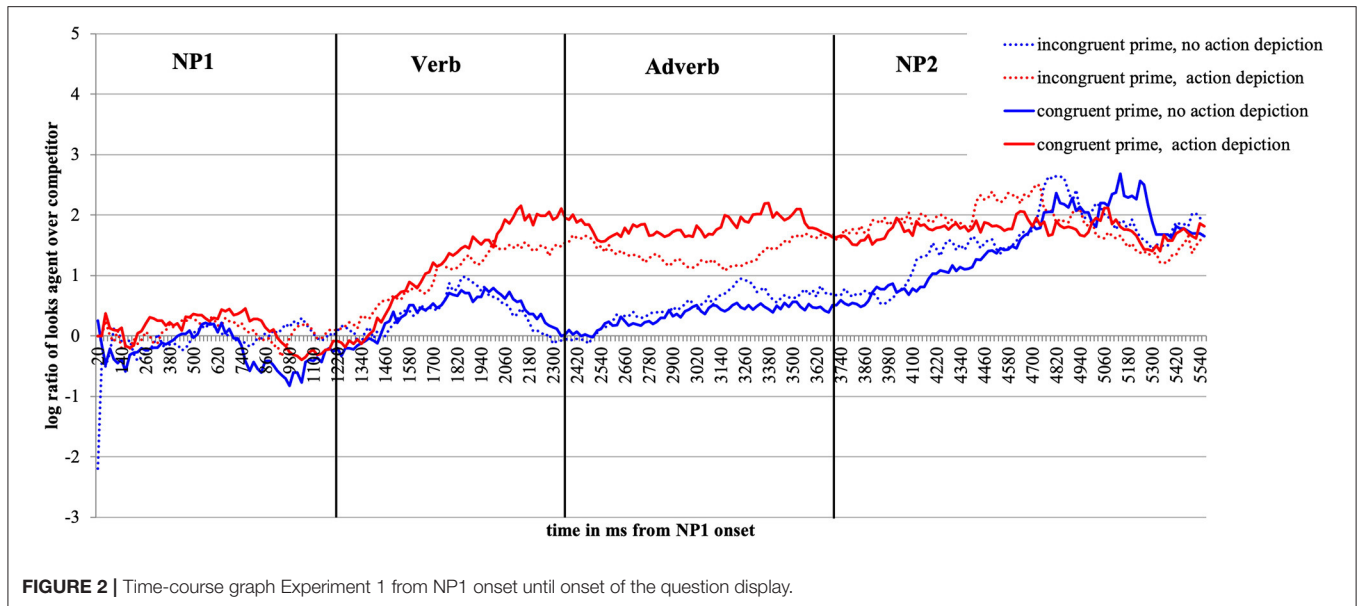
Results Experiment 1

Descriptive Eye-Movement Results

Figure 2 plots the time course of fixations to the target agent relative to the competitor character from the onset of NP1 until the end of the target display in bins of 20 ms. As expected, upon hearing the patient named (NP1), participants show no gaze bias to either the agent or the competitor character. Upon encountering the verb, participants begin to look more at the target agent than competitor, and more so in the action (the red lines) than no action (blue lines) conditions. This effect lasts until the agent is mentioned (middle of NP2).

Focusing now on the contrast between the valence-congruent (smiley) prime and the incongruent (red star) prime, we see no preference in looks toward the agent (vs. competitor) during the verb and adverb² when no action was depicted (the two blue lines do not diverge). However, if an action was present (red lines), the presence of the congruent smiley (vs. the incongruent red star prime) drew subtly more looks toward the agent during the verb and especially the adverb region (the solid vs. the dotted red line, respectively). During the NP2 region (target agent named), the red and the blue lines begin to converge as the agent was named and could thus be discriminated.

²Recall that these were the critical word regions as the verb relates to the action and the adverb reveals the emotional valence of the sentence.



Inferential Eye-Movement Analyses

The inferential analyses revealed significant main effects of action in the verb³, adverb⁴, verb-adverb⁵, NP2⁶, and the long region⁷ (e.g., verb region: β : -0.973, SE: 0.141, df : 582.000, t : -6.888, p < 0.001), but not in the NP1⁸ region (see the **Supplementary Material** for additional model parameters of all analyses). Participants fixated the agent significantly more

than the competitor when an action was depicted (vs. no action depiction). Moreover, this effect started as soon as the verb information became available (see **Figure 3**). No effects of emotional prime emerged (see **Figure 4**).

Accuracy Results

Participants answered 96% of all trials correctly. The analyses yielded no significant effects of the independent factors regardless of whether question voice (active vs. passing) was an additional factor in the model or not⁹.

Discussion Experiment 1

The eye-tracking pattern corroborated that from the verb onward people made extensive use of the depicted actions for

³Best-fitting model: $\log_ratio \sim \text{prime} * \text{action} + (1 | \text{participant}) + (1 | \text{item})$.

⁴Best-fitting model: $\log_ratio \sim \text{prime} * \text{action} + (1 + \text{action} | \text{participant}) + (1 + \text{action} | \text{item})$.

⁵Best-fitting model: $\log_ratio \sim \text{prime} * \text{action} + (1 + \text{action} | \text{participant}) + (1 + \text{action} | \text{item})$.

⁶Best-fitting model: $\log_ratio \sim \text{prime} * \text{action} + (1 | \text{participant}) + (1 | \text{item})$.

⁷Best-fitting model: $\log_ratio \sim \text{prime} * \text{action} + (1 + \text{action} | \text{participant}) + (1 + \text{action} + \text{prime} : \text{action} | \text{item})$.

⁸Best-fitting model: $\log_ratio \sim \text{prime} * \text{action} + (1 | \text{participant}) + (1 | \text{item})$.

⁹Best-fitting model: $\text{answer} \sim \text{prime} * \text{action} * \text{voice} + (1 | \text{participant}) + (1 | \text{item})$.

discriminating the target agent of the OVS sentence. By contrast, effects of the emotional prime on sentence interpretation were not reliable. One reason for the absence of clear effects of the emotional smiley on the eye-movement pattern could be that the smiley did not evoke a strong impression of positivity, eliciting accordingly a fairly small (and nonsignificant) priming effect on the processing of the emotionally positive sentence.

Although schematic portrayals of emotions, such as the smiley, might highlight the emotional expression and focus on positivity, its schematic nature could also impede strong emotion effects on language processing: emotional facial expressions are reflections of dynamic psychological and physiological states that vary not only from person to person but also in their degree of positivity or negativity and meaning in context. Recognizing emotion from even natural faces out of context is difficult (Barrett et al., 2007). That difficulty might prevent language users from fully and rapidly exploiting facial expressions during language comprehension, especially when the emotional facial expression is presented as the first stimulus (without prior context) and when it is not referenced by language.

At the same time, recognizing our interlocutors' (facial) emotion during an interaction is vital for building and maintaining social relationships; for nonverbal communication, and moreover for interpreting the other person's feelings (Lamy et al., 2008). Perhaps then, a real human speaker face (in contrast to a schematic depiction) is needed for enabling effects of the facial prime during language comprehension. In fact, emotions from a human face are recognized faster and more accurately and elicit enhanced and prolonged cortical responses when they are presented in a more natural manner (dynamic) compared with static [see, e.g., Harwood et al. (1999) for the identification of emotions from moving and static videotaped and photographic displays; Recio et al. (2011) for ERP evidence]. In addition, viewing a smiley prime but then hearing a human voice may have decreased listeners' integration of the prime face with the speech and its associated valence. In Experiment 2, we accordingly hypothesized that using a video of a human speaker's (facial expression) might lead to stronger priming effects than a schematic smiley.

Another explanation for the weak effects of the emotional prime could be a task bias. Since the comprehension questions focused on thematic role relations, participants may have focused on the actions, boosting their effects on comprehension. This explanation receives support from a study by Hajcak et al. (2006) in which an ERP component (the LPP) often observed in emotion processing was significantly smaller for non-affective than affective judgments of emotionally valenced pictures. In Experiment 2, we complemented the questions about who-did-what-to-whom with an emotion-verification task (see below).

In terms of the performance on the comprehension task, participants' scores were at ceiling in all four conditions. Exploratory analyses revealed that participants answered slightly more passive than active comprehensions questions incorrectly (passive: 3.4% vs. active: 0.6%; difference *n.s.*). The at-ceiling accuracy values especially in the active voice might have concealed off-line effects of the emotional prime and action

factors. To explore this possibility, Experiment 2 posed all experiment-trial questions in the passive voice.

EXPERIMENT 2

Methods

Participants

A further 40 students of University of Bielefeld, all native speakers of German with normal or corrected-to-normal vision, between 18 and 30 years (15 male, mean age: 23, SD age: 3.62) took part in the experiment. Participants were tested in the eye-tracking laboratory of Bielefeld University. Sample size was set to 40 to ensure comparability with a related study with children as participants. Participants received 4 Euro for their participation. All gave written informed consent. Ethics approval was given by the university's ethics board (Vote 2013-007).

Materials

The sentences and scenes were the same as in Experiment 1. However, the comprehension questions for the 16 experimental sentences in Experiment 2 were always asked in the passive voice. Across all trials, participants heard equally many active and passive comprehension questions. To foreground emotional valence, the experimenter asked the participant to recall the prime speaker's facial expression after four experimental trials (one per condition). These questions were counterbalanced so that across participants they appeared after each item in each experimental condition. In addition, after 12 of the 28 filler trials participants had to identify emotions ("How are they feeling?"): they saw (posttrial) the facial expression of the speaker (always same valence as the prime) next to the face of one of the characters from the previously seen target scene. The same female speaker who recorded the item sentences recorded these questions. In the remaining 16 filler trials, participants were asked the same active question about the agent/patient of the sentence as in Experiment 1.

The emotional prime in Experiment 2 consisted of face videos. In the positive video, a woman changed her facial expression from neutral into a broad smile (video duration: 5,500 ms; change to positive after 1,300 ms). The negative video was constructed in the same way but the woman's face turned from a neutral into a sad expression (see **Table 2**). We chose the woman's face based on a previous rating study of facial emotion photos [DeCot (2011); unpublished Master's thesis, *N* = 18, mean age: 24.7]. In that study, the woman's happy and sad facial (static) expressions were one of the three most recognizable among 15 faces (large differences between neutral, positive, and negative emotions). We recorded that woman for the videos of the present study¹⁰.

Design and Hypotheses

The design was identical to Experiment 1 (2 × 2, action × emotional prime) with the exception that for the prime face, only the emotional expression varied in valence (**Table 2**). For

¹⁰The woman portraying the facial expressions gave informed consent regarding the use and publication of her facial expressions for research purposes related to our studies.

TABLE 2 | Experimental conditions for Experiment 2.

Condition	Prime	Action	Sentence
A	Happy dynamic natural facial expression, congruent prime	Action depicted	<i>Den Marienkäfer kitzelt vergnügt der Kater</i> ["The ladybug (accusative object, patient) tickles happily the cat (nominative subject, agent)"]
B	Happy dynamic natural facial expression, congruent prime	No action depicted	
C	Sad dynamic natural facial expression, incongruent prime	Action depicted	
D	Sad dynamic natural facial expression, incongruent prime	No action depicted	

the accuracy data, we predicted to replicate the results from Experiment 1. However, to the extent that the passive questions in Experiment 2 minimize the potential ceiling effects (see Experiment 1), we expected to see more correct answers when the actions were present (vs. absent), and the emotional prime face was congruent (vs. incongruent). Participants should further correctly answer the recall questions about the emotional valence of the prime face in the experimental trials. For the eye-movement behavior, we also expected to replicate the fixation patterns from Experiment 1. However, given the increased task focus on emotional valence, we predicted stronger effects in Experiment 2 than 1. If both the naturalness of the prime face and the increased task focus on emotions boost agent anticipation, fixations toward the agent (vs. competitor) should be more pronounced in Experiment 2 than 1. We predicted no between experiment differences for the action effects.

Analyses

The analyses for the eye-movement data were the same as in Experiment 1. Accuracy scores for the face recall task were not analyzed inferentially since they only yielded four data points per participant. Accuracy scores for the questions about how the characters are feeling were also not analyzed as these questions were only asked on filler trials. Analyses of the comprehension

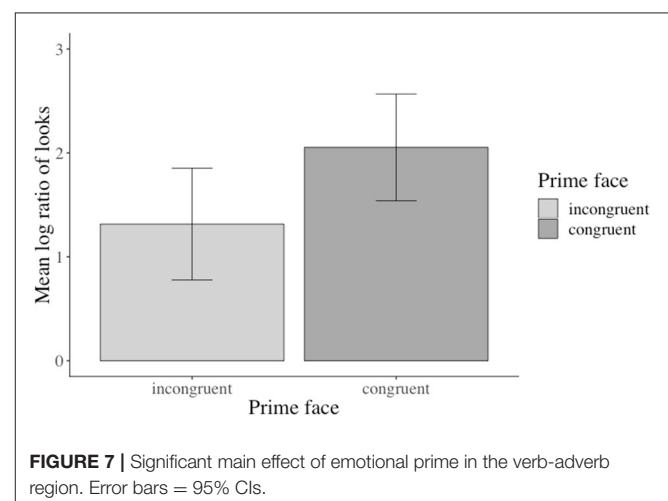
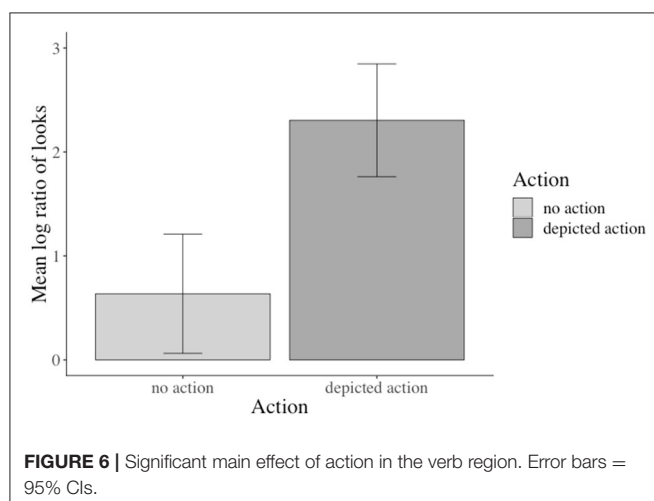
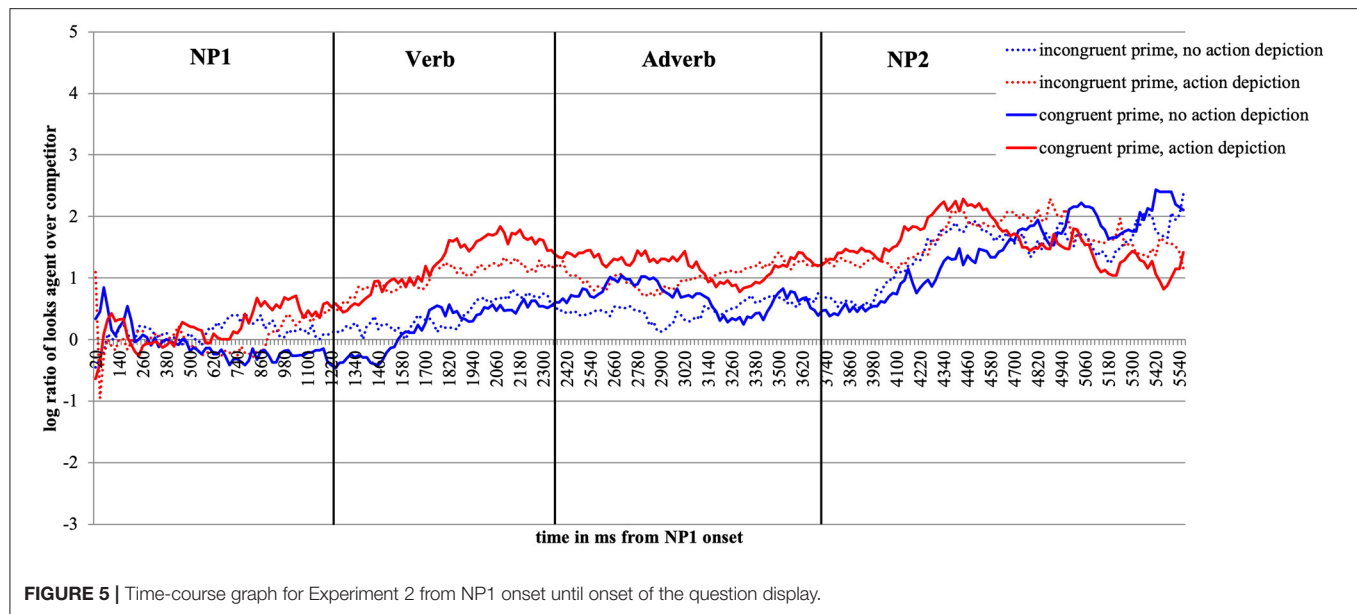
question scores were performed similar to Experiment 1 using prime and action as fixed factors. However, since all experimental comprehension questions were in the passive voice, voice was not included in the models.

Results Experiment 2
Descriptive Eye-Movement Results

Figure 5 plots the time course of fixations to the target agent relative to the competitor character from the onset of NP1 until the end of the target display. The log ratios are plotted as a function of emotional prime and action depiction.

Figure 5 shows that upon hearing NP1, participants, as expected, look equally often at the agent and the competitor character. Upon encountering the verb, which makes available valence and action information, participants begin to look more at the agent than at the competitor as in Experiment 1. Moreover, under the conditions in which an action is depicted (red lines), the listeners' gaze clearly deviates from gaze in the no-action conditions (blue lines). This effect lasts from the verb until the target agent is mentioned (middle of the NP2).

When contrasting the incongruent (dotted lines) and congruent face prime conditions (solid lines), we note that they deviate. When no action is depicted, this happens from the end of the verb and into the adverb region (the solid is above the



dotted blue line). When an action is present (red lines), the solid is above the dotted line from the middle of the verb onward until the middle of the adverb region. The presence of the emotionally congruent (vs. incongruent) prime thus draws more looks toward the target agent during the verb and the adverb region (recall that these are the critical word regions as they give away the action and the valence of the sentence). During the NP2 region, the red and the blue lines converge again as the agent is named and can thus be discriminated.

Inferential Eye-Movement Analyses

Inferential statistics indicated a main effect of action: participants fixated the target agent significantly more than the competitor during all analyzed word regions when an action was (vs. was

not) depicted, except for the NP1 region¹¹ (i.e., verb region¹²: $\beta: -0.826$, SE: 0.1409, $df: 577.300$, $t: -5.867$, $p < 0.01$, see **Figure 6**). Moreover, a main effect of emotional prime emerged in the verb-adverb¹³ ($\beta: -0.364$, SE: 0.129, $df: 522.3$, $t: -2.810$, $p < 0.01$, see **Figure 7**) and the long region¹⁴ ($\beta: -0.224$, SE: 0.106, $df: 572.7$, $t: -2.100$, $p < 0.05$). Participants fixated the target agent significantly more than the competitor when the speaker's prime face was positive (vs. negative) upon hearing the verb-adverb and across the long region. The interaction

¹¹Best-fitting model: $\log_ratio \sim \text{prime} * \text{action} + (1 + \text{action} | \text{participant}) + (1 | \text{item})$.

¹²Best-fitting model: $\log_ratio \sim \text{prime} * \text{action} + (1 + \text{prime} | \text{participant}) + (1 + \text{prime} | \text{item})$.

¹³Best-fitting model: $\log_ratio \sim \text{prime} * \text{action} + (1 + \text{action} | \text{participant}) + (1 + \text{action} | \text{item})$.

¹⁴Best-fitting model: $\log_ratio \sim \text{prime} * \text{action} + (1 | \text{participant}) + (1 | \text{item})$.



FIGURE 8 | Example item from Experiment 3: emotional facial expressions of the competitor character (i.e., the rat) and the target agent (i.e., the cat) are increased in valence compared to items from Experiments 1 and 2 (cf., **Figure 1**).

between action and emotional prime was not significant (see the **Supplementary Material** for additional model parameters of all analyses).

Accuracy Results

Participants answered 96% of the comprehension question and 99.4% of the face recall question correctly. The analysis of the answers for the “who-does-what” comprehension questions¹⁵ yielded no significant effects of prime face or action.

Discussion Experiment 2

The eye-tracking results reinforce the insight from Experiment 1—viz. that people use the depicted action toward discriminating the agent of the OVS sentence starting from the verb. In contrast to Experiment 1, we also observed a significant effect of the emotional prime in the verb-adverb and long region. As before, however, the emotional cue seems to be used to a lesser degree for sentence interpretation than the action depiction as the effects emerged only in the verb-adverb and long region and were not as pervasive as the action effect. Participants’ accuracy in answering both the comprehension and face recall questions was high (no significant effects of the independent factors on accuracies were observed).

Experiments 1 and 2 replicated the effects of depicted actions on the assignment of thematic roles in non-canonical German OVS sentences. Participants reliably used depicted actions as contextual cues to anticipate the target agent, facilitating sentence processing in real time. Yet, whereas the schematic smiley did not affect real-time sentence processing, the natural emotional speaker’s face led to a significant effect in the verb-adverb and the long region, i.e., across the whole sentence. Nevertheless, the emotional prime face still had a less pervasive effect than the depicted action. Our final study addressed this issue to examine the emotional prime effect in more detail. We will briefly address and motivate each change for Experiment 3 with regard to the results of Experiments 1 and 2. Note that Experiment 3 was conducted to maximize finding an effect of the emotional facial

expression, resulting in more changes than for an incremental replication experiment.

Accordingly, we changed the design, materials, presentation, and task in Experiment 3 such that the focus was on the speaker’s facial expression. We omitted the action factor and scenes depicted no actions. Additionally, we removed the middle character; target scenes only contained the happy-looking target agent and the grumpy-looking competitor character. Further, the characters’ facial expressions were rendered more salient (see **Figure 8**). In line with this, the emotional facial prime was introduced in an explicit manner as the speaker of the sentences: Participants were told that it is a mother who be reading short sentences about the actions of cartoon characters to her child. With the prime face now being explicitly introduced as the speaker, we hoped to increase participants’ emotional expectations about the following scene: A happy-looking speaker is more likely going to utter an emotionally congruent (vs. incongruent) event, which more likely features a happy (vs. grumpy) looking agent. The visual scene was reduced in complexity, giving participants the chance to focus on the linking of emotional valence between the visual and linguistic input. Moreover, this linking was made more explicit to participants, highlighting its ecological validity, by framing a real-world interactive setting between a mother and her child.

We hypothesized that the less pervasive effect of the emotional prime effect might be due to the different ways of cue presentation. While the depicted actions were present during real-time sentence processing, the emotional prime face was presented prior to encountering the target scene and sentence. In Experiment 3, we presented the speaker’s facial expression as a prime because presenting it simultaneously together with the target scene while participants listen to the OVS sentence would have drawn attention away from the characters [unpublished pilot data from another study (2011) supports this view]. However, we decreased the scene preview time from 2,000 to 500 ms. This shortens the time for which the emotional prime must be retained in working memory, perhaps facilitating its effects on the processing of the target sentence.

A further change concerned participants’ task. In Experiments 1 and 2, participants answered comprehension questions in the active or passive voice for who-is-doing-what-to-whom and/or

¹⁵Best-fitting model: $\text{answer} \sim \text{prime} * \text{action} + (1 + \text{action} | \text{participant}) + (1 | \text{item})$.

questions about the emotional expression of the prime face and the feelings of the depicted characters. No time limit had been imposed and accuracy had been at ceiling. To assess whether participants kept the prime face in memory during the trials, in Experiment 3 we presented them with a screenshot of either the positive or the negative speaker face after *each* trial. Participants were instructed to verify *via* button press as quickly and accurately as possible whether the screenshot of the facial expression matched the speaker's expression in the prime video. We further assessed participants' reaction time for the assignment of thematic roles. On each trial, following the face verification, the target sentence was repeated in SVO structure. Participants were asked to verify as quickly and accurately as possible whether the OVS target sentence and the SVO sentence described the same event or not. A correct answer indicated participants had understood the thematic role relations in the OVS sentence. Finally, as we had 16 experimental items and relatively low power, we increased our sample from 40 to 64 participants.

With the implemented changes, the resulting simplified design and a stronger focus on the interactive and situated setting in the instructions, we expected to find a more pervasive effect of a speaker's emotional facial expression on the real-time processing and thematic role assignment of non-canonical OVS sentences, if, as Experiment 2 suggested, emotional facial expressions can indeed facilitate the assignment of thematic roles.

EXPERIMENT 3

Method and Design

Participants

A further 64 students of Humboldt-Universität zu Berlin, all native speakers of German with normal or corrected-to-normal vision, between 18 and 30 years (32 male, mean age: 23, SD age: 3.39) took part in the experiment. Participants were tested in the eye-tracking laboratory of Humboldt-Universität zu Berlin. One participant had to be excluded due to a technical problem. Participants received 8.50 Euro for their participation. All gave written informed consent. Ethics approval was given by the ethics board of the German association for linguistics (DGfS, Laboratory ethics approval, valid from September 17, 2016 to September 16, 2022).

Design

We are only reporting the changes compared to Experiment 2 here. We removed the depicted actions, and retained the factor emotional prime with two levels: congruent prime vs. incongruent prime. We additionally changed the face recall questions to prime face verification: participants are now asked to decide as quickly and accurately as possible if the speaker's prime face matches the posttrial picture of the speaker's face. Additionally, we changed the who-does-what question: participants heard a sentence either matching or mismatching the target OVS sentence in content, sentence structure, or in both. They decided as quickly and accurately as possible whether the content in the two sentences was identical or not. Participants answered with their left and right index fingers using yes/no buttons on a button box. The position of the yes/no buttons was

counterbalanced across participants. One experimental session took about 50 min.

Materials

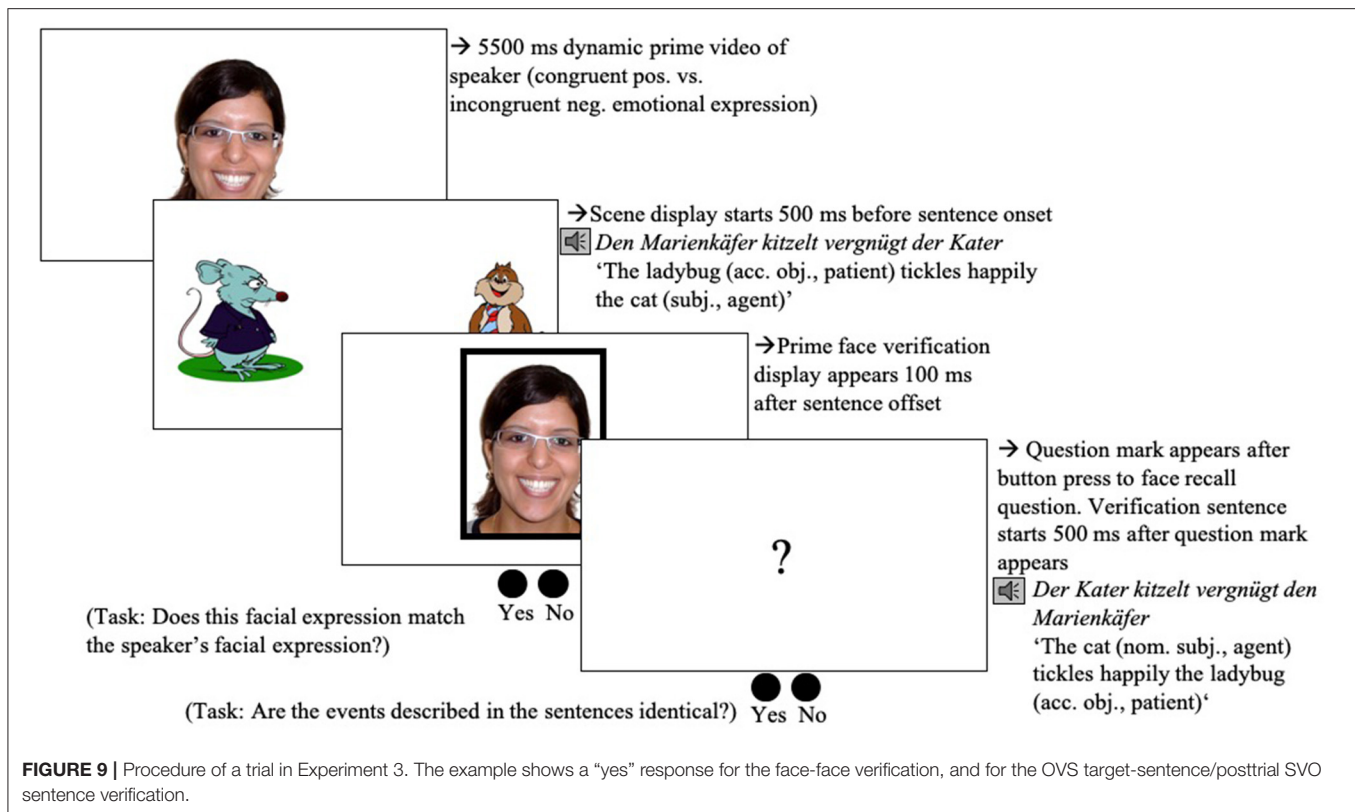
We removed the middle character (the patient) from the scenes, such that participants' attention is only divided between the target agent and the competitor character. Moreover, the target and competitor character's emotional facial expressions were improved such that the competitor character's face was very negative and the target agent's face very positive (see **Figure 8**).

For the prime face verification questions, a screenshot of the final frame of the speaker's emotional prime face video was taken and used in each condition: For all critical items in a list, the picture of the test face was positive in half of the trials and negative in the other half. Moreover, the test face matched the prime face in the video in half of the trials and mismatched in the other half of the trials. The filler trials were also followed by either a (mis) matching positive or (mis) matching negative emotional prime face picture. Across all trials, the valence and emotional match of the pictures used for prime face verification were counterbalanced within and across lists and participants.

We further created 16 SVO versions of the critical OVS item sentences for the sentence verification questions. All critical to-be-verified sentences matched in content, i.e., also in role relations and only mismatched in sentence structure, thus requiring a yes answer. Additionally, we created a verification sentence for each of the 24 filler sentences: The filler sentences either mismatched in sentence structure ($N = 4$), in content ($N = 22$), or matched in structure ($N = 24$) or in content ($N = 6$) with the to-be-verified sentence. Across all 44 critical item and filler trials, 24 of the verification sentences matched in structure and 20 mismatched. Also, half of the 44 verification sentences matched in content with the to-be-verified sentences and the other half mismatched in content. This resulted in half of the trials requiring a yes and half of them requiring a no answer for each participant. The same speaker (PK) that recorded the critical and filler sentences also recorded the sentences for the sentence verification question.

Hypotheses

If we find more looks to the target agent (vs. the competitor) during the verb-adverb region in the positive congruent prime condition (vs. the negative incongruent prime condition), we could conclude that participants have (a) correctly assigned thematic roles even in the absence of depicted action cues and (b) have done so in a facilitative fashion with the help of the positive prime face. The reaction-time data should corroborate that participants are faster in verifying OVS/SVO event identity when the prime is congruent (vs. incongruent). Conclusions regarding the assignment of thematic roles are less clear if, however, we only find effects of the positive prime face in the eye tracking but not the reaction-time data. People would still have established an on-the-fly link between the positive face, the target agent's happy facial expression and the positive verb-adverb. But—in the absence of an emotional prime effect on the response times—we could not be sure that the emotional prime improved the post-trial comprehension of the role relations. By contrast, if we find an effect of the positive prime face in the response times



but not in the eye-movement record, then we could conclude that the prime face helps participants to correctly assign thematic roles but that this effect takes time and participants cannot use it “on the fly” to ease the assignment of thematic roles.

Procedure

Each trial started with a fixation dot (not depicted in **Figure 9**), followed by the positive (vs. negative) emotional prime face video of the speaker of the following sentence. After the video (5,500 ms), participants encountered the scene showing the happy-looking target agent (the cat in **Figure 8**) and the grumpy-looking competitor (the rat in **Figure 8**) for 500 ms before the OVS sentence started to play. Participants looked at the scene while listening to the sentence. Following the sentence, the prime face verification display showed either a positive or a negative static frame from the prime face video. Participants indicated as quickly and accurately as they could if this facial expression matched the speaker's facial expression in the prime video. Following the button press, a question mark appeared on the screen. About 500 ms after the onset of this question mark, participants listened to the to-be-verified SVO sentence. They were asked to judge sentence identity as quickly and accurately as possible.

Analyses

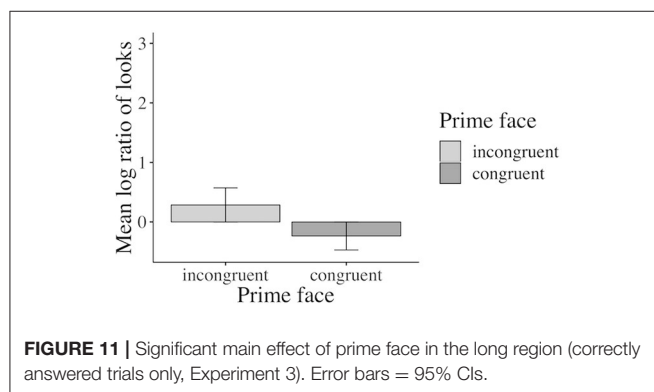
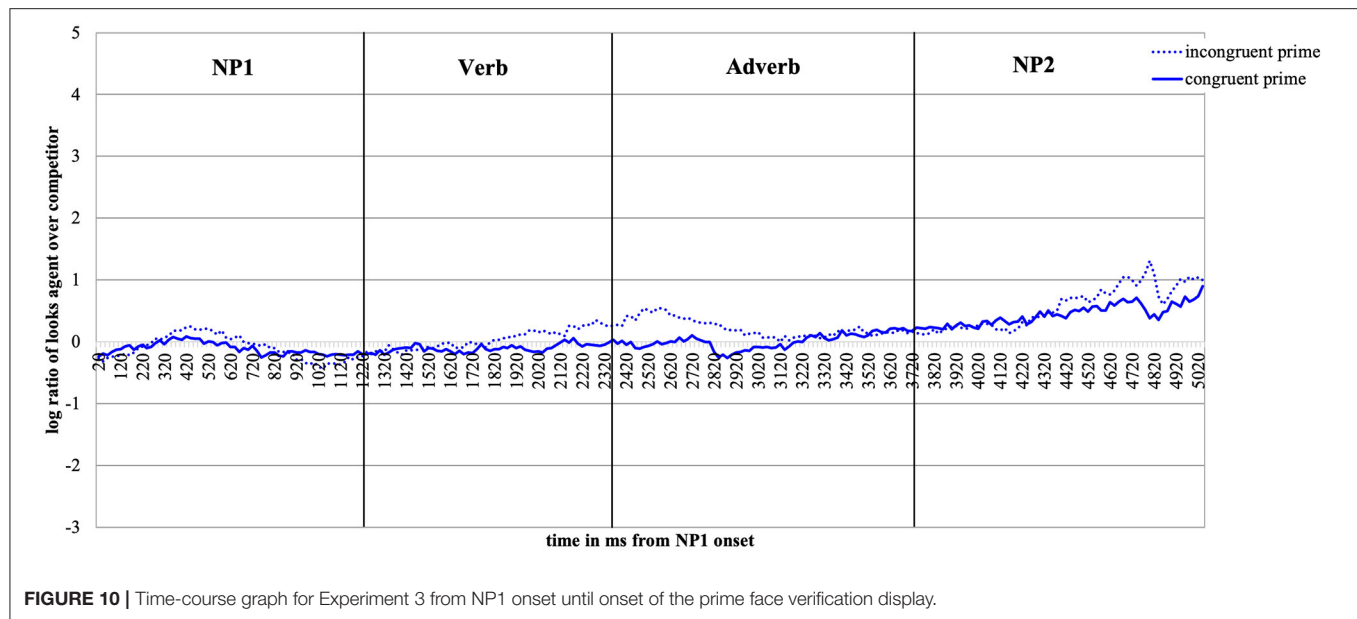
The analyses of the eye-tracking, accuracy, and response-time data followed the same procedure as for Experiments 1 and 2 (see Sections Eye Movements and Analyses) but the eye-tracking

data analysis was performed on correctly answered trials only. The fixed factor for the eye-tracking analyses was emotional prime (congruent vs. incongruent). The fixed factors for the prime face verification accuracy and RT analysis were emotional prime (congruent vs. negative) and match (prime face—test face match vs. prime face—test face mismatch). The fixed factor for the sentence verification accuracy analyses was emotional prime (congruent vs. incongruent). The response-times for the prime face verification and the sentence verification questions (for the correctly answered trials only) were analyzed by mixed-effects models using log transformed reaction times as a dependent variable and prime and match (match was not a fixed factor in the sentence verification question because all critical items were followed by a matching SVO sentence) as fixed factors. Reaction times ± 2.5 SD* condition mean were treated as outliers and excluded from the analysis.

Results Experiment 3

Descriptive Eye-Movement Results

Figure 10 shows the time-course graph for Experiment 3. Whereas participants did not seem to have a preference to either look at the competitor or at the target agent during the first noun phrase and until the middle of the verb region, a preference to look more at the target agent emerged from the middle of the verb until the middle of the adverb region in the incongruent prime condition. Interestingly, this deviation between the solid and the dotted blue line shows that participants fixated the happy-looking target agent (vs. the grumpy-looking competitor) more after



having been primed with a negative and sentence valence-incongruent (vs. positive valence-congruent) speaker face. At the end of the adverb region and until the end of the sentence, participants started to look at the target agent in both conditions when it was named in the NP2 region but descriptively more in the incongruent (vs. congruent) prime condition.

Inferential Eye-Movement Analyses

Accuracy in the prime face and sentence verification questions was lower (92 and 77% correct answers, respectively) than for the who-does-what questions (96% correct answers in both Experiments 1 and 2). Given the lower accuracies, we excluded incorrectly answered prime face verification and sentence verification trials from the analysis. The analyses revealed a significant main effect of emotional prime in the verb-adverb¹⁶ (β : 0.2749, SE: 0.1192, 16.0189, t : 2.307, p < 0.05) and in the

long region¹⁷ (β : 0.25010, SE: 0.09261, df : 17.88568, t : 2.700 p < 0.05, see the **Supplementary Material** for additional model parameters of all analyses): participants in the verb-adverb region and across the sentence fixated the competitor character more after having seen a positive congruent (vs. negative incongruent) prime face and vice versa (see **Figure 11**)¹⁸.

Accuracy Results

Across conditions, participants answered 92% of the prime face verification questions and 77% of the sentence verification questions correctly. The analysis revealed a marginal (β : -0.3842, SE: 0.2020, z : -1.902, p = 0.0571) main effect of prime face for the face verification questions¹⁹: participants answered more verification questions correctly when the prime face was congruent (47% vs. incongruent 43%, see **Figure 12**). Accuracies for match (vs. mismatch) conditions did not differ significantly. The accuracy analysis for the sentence verification question did not yield any significant effects of the manipulated factor²⁰.

Reaction Time Results

The reaction time analysis for the prime face verification question revealed a significant main effect of emotional prime (β : 0.030808, SE: 0.008537, df : 814.045317, t : 3.609, p < 0.001), a significant main effect of match (β : 0.058065, SE: 0.008648, df : 818.756839, t : 6.714, p < 0.001), and a significant emotional prime \times match interaction (β : -0.028995, SE: 0.008649, df : 817.432855, t : -3.352, p < 0.001): participants responded faster when the prime face was congruent (vs. incongruent,

¹⁶Best-fitting model: $\log_ratio \sim \text{prime} + (1 + \text{prime} | \text{participant}) + (1 + \text{prime} | \text{item})$.

¹⁷Best-fitting model: $\log_ratio \sim \text{prime} + (1 + \text{prime} | \text{participant}) + (1 + \text{prime} | \text{item})$.

¹⁸Including all trials (correctly and incorrectly answered) in the analysis yielded a significant main effect in the long region (p < 0.001), the effect in the verb-adverb region did not reach significance, but the result pattern stays identical.

¹⁹Best-fitting model: $\text{answer} \sim \text{prime} + (1 + \text{prime} | \text{participant}) + (1 | \text{item})$.

²⁰Best-fitting model: $\text{answer} \sim \text{prime} + (1 + \text{prime} | \text{participant}) + (1 | \text{item})$.

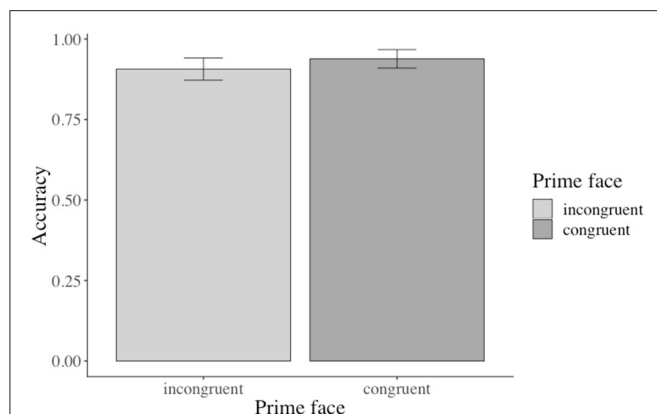


FIGURE 12 | Marginal main effect of the prime face for the accuracy data in Experiment 3. Error bars = 95% CIs.

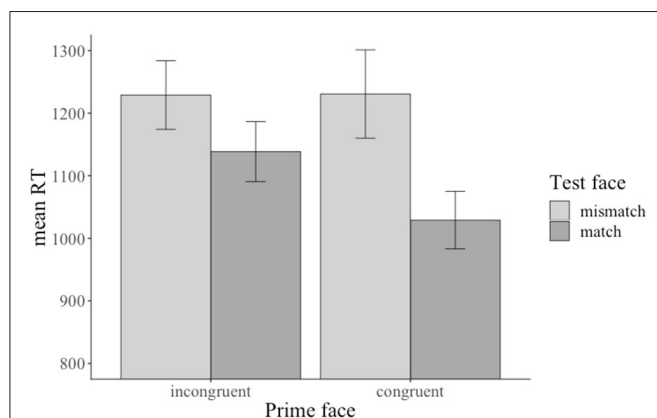


FIGURE 13 | Significant main effects and interaction for the reaction time data in Experiment 3. The prime was congruent (vs. incongruent) with sentence valence. The test face either matched or mismatched with the prime face. Error bars = 95% CIs.

see **Figure 13**) and when the prime face valence and the test face valence matched (vs. mismatched, see **Figure 13**) with each other. Moreover, they responded faster when the prime face was congruent and matched (vs. incongruent and mismatched) with the test face (see **Figure 13**)²¹. The reaction-time analysis for the sentence verification question did not reveal any significant effects of the manipulated factors.

Discussion Experiment 3

In Experiment 3, we set the focus on the emotional prime face as a cue for language processing and highlighted the emotional valence and its relevance in the experimental materials, the task, presentation, and the design. We predicted that the increased focus on the emotional valence information would bring out the effect of the emotional prime face that we observed in Experiment 2 even more, i.e., participants were expected to anticipate the happy-looking target agent during the positively

valenced verb-adverb and long region more when they had been primed with a congruent (happy vs. incongruent, sad) speaker facial expression. The emotional prime face did affect the looks to the characters in the scene, and this effect emerged in the same regions as in Experiment 2 (the verb-adverb and the long region). However, the directionality of the effect goes into the opposite-than-expected direction. Participants preferred to look at the grumpy-looking competitor character more after being primed with a sentence-congruent (positive) prime face and preferred to look at the happy target agent more after being primed with a sentence-incongruent (negative) prime face (**Figures 10, 11**). The reaction time and (to some extent) the accuracy of the face verification data, on the other hand, show that participants verified the positive test face picture faster and more accurately after being primed with a congruent (positive) speaker prime face (vs. incongruent negative). We discuss this finding in the general discussion (see Section The Reversed Priming Effect).

GENERAL DISCUSSION

In three visual-world eye-tracking experiments, we assessed whether participants' incremental comprehension of non-canonical OVS sentences is modulated to the same extent by referential cues (depicted actions) and non-referential cues, i.e., visual (social) cues (positive facial emotions that are arguably non-referentially linked to the sentence meaning).

The Action Effect

In Experiments 1 and 2, participants made extensive use of the depicted action (vs. no action) for online sentence processing. They were more likely to anticipate the correct target agent (vs. a competitor) during OVS sentence processing in real time when the agent depicted in the scene was performing (vs. was not performing) the action mentioned in the sentence. We hence replicated existing findings demonstrating that a supportive visual context, which is referenced by the linguistic input, can rapidly facilitate OVS sentence processing and the assignment of thematic roles (cf., Knoeferle et al., 2005, 2008; Zhang and Knoeferle, 2012).

The Positive Prime Effect

The positive prime face, by contrast, had only a significant effect on sentence comprehension when the prime face was human (Experiments 2 and 3), but not when it was a smiley (Experiment 1). Moreover, the effects of the emotional facial expression were less pervasive compared to the effects of the referenced depicted action.

The latter finding could be caused by the way the listeners link aspects of the scene to language. While the depicted action is mediated by its referring linguistic expression (the verb), matching the emotional valence of a preceding prime face to the target character's facial expression and to the valence of the adverb is arguably more complex. In our view, the greater complexity arises because, first of all, the listener must infer the valence of the emotional prime face, since no emotion labels are provided. Keeping the inferred emotion in memory, s/he then inspects the target scene and starts to interpret it in relation to the

²¹Best-fitting model: $\log_RT \sim \text{prime} * \text{match} + (1 | \text{participant}) + (1 | \text{item})$.

unfolding OVS sentence. Only when the valence of the emotional sentence becomes clear (during the verb and adverb regions), could the listener reactivate the previously seen emotional face and link the valence of the face to the emotional adverb and the happy smile of the target agent. In Experiment 2, this prime face-language connection enabled anticipatory looks to the target agent. We propose that effects of the non-referential cue arguably involve more processing steps than the effects of the referential cue on real-time assignment of thematic roles (the verb mediates the action and the associated agent that can fill a thematic role slot).

Alternatively, the effects of the emotional primes were less pervasive than of the actions because of differences in presentation. While the depicted action was present during sentence presentation, the emotional facial expression was shown and then removed before the onset of the target scene and sentence. Thus, whereas the listener had access to the depicted action throughout comprehension, the emotional prime face had to be held in working memory from the end of the prime presentation until the end of the trial. The effects of the emotional prime might be less pervasive because relating visual and linguistic input is arguably easier when both are presented together than when they are presented separately. This assumption is indirectly supported by Glenberg and Robertson's (1999) indexical hypothesis. It assumes that the referencing (or indexing) of language to the visual context is easier when both are co-present compared to when they are presented separately. That said, other studies have reported emotional face effects in a serial prime-target paradigm. The prime and target presentation in Carminati and Knoeferle (2013) were—just like in our studies—serial in nature.

Another reason for why the effects of the non-referential cue were less pervasive could be the different sentence structures investigated. Recall that Carminati and Knoeferle (2013) and Münster et al. (2014) investigated emotional priming effects on the processing of SVO sentences. Processing structurally challenging non-canonical OVS sentences and assigning thematic roles, as was the case in our studies, is arguably a cognitively more demanding task than reconciling a prime face with the semantic meaning and valence of a canonical SVO sentence. Linking a speaker's emotional facial expression to a corresponding emotionally valenced adverb and a character's emotional facial expression while assigning thematic roles in a non-canonical OVS sentence might account for the subtler effects compared to the study by Carminati and Knoeferle (2013) and Münster et al. (2014).

However, while being less pervasive than the effects of the depicted action in Experiment 2, the emotional prime effects provide some support for the view that the valence of a prime face can also facilitate the real-time assignment of thematic roles in non-canonical German OVS sentences.

The Naturalness of the Emotional Prime Face

The analyses also revealed significant effects of the dynamic *natural* (but not the smiley) emotional facial expressions. One possible reason for that could lie in the changes in emotional prime presentation between Experiments 1 and 2. While the

happy utterance-congruent smiley (Experiment 1) contrasted with a (nonemotional utterance-unrelated) static red star, the valence congruence manipulation in Experiment 2 was achieved by contrasting a woman's happy (utterance-congruent) and sad (incongruent) facial expressions. Hence, we contrasted an emotional with a non-emotional prime in Experiment 1 but one woman's facial expression of contrasting valences in Experiment 2. In order to strengthen our claim that only the naturalness of the emotional prime face elicited an effect in Experiment 2 but not 1, the happy smiley prime in Experiment 1 would have to be contrasted directly with a sad smiley prime instead of a red star in a follow-up study. Additionally, using valence-contrasting faces in Experiment 2 may have increased the overall awareness of emotions in the study and boosted prime effects compared to Experiment 1. This might additionally be the case because in Experiment 2, the focus on emotions was further increased by occasionally asking participants to comment on the characters' feelings and to recall the valence of the prime face.

However, it could also be argued that the longer exposure duration of the natural facial prime (5,500 vs. 1,750 ms for the schematic smiley) caused its better integration into sentence processing. A longer exposure to the emotional expression might have led to more in-depth processing of the emotional content although the studies that have varied stimulus duration (from as short as 50 ms to 10 s) of emotional facial expressions found no effect of the duration manipulation on emotion perception and recognition for happy faces (Calvo and Lundqvist, 2008; Codispoti et al., 2009). Hence, the stronger effects of the natural dynamic facial expression are likely due to its higher ecological validity and not the longer stimulus exposure.

Ecological validity might have contributed to a better integration of the human (compared to the smiley) prime face into the real-time processing of the OVS sentence. The target sentence was spoken by a female voice, and the gender match with the face prime might have encouraged participants to more readily perceive the happy-looking woman as the speaker than a smiley. That this is likely the case is also supported by "Theory of Mind" assumptions (e.g., Premack and Woodruff, 1978). Using the facial expressions of our interlocutors, we attribute our own mental states to others and expect the other person to act like we would have acted with a similar facial expression. Additionally, the use of more naturalistic real-world stimuli in experiments, especially those on emotion and language, is recommended (e.g., Adolphs, 2006). Many studies on human (social) cognition often use morphed, synthetic, or static faces. The social link that two humans establish during communication may not emerge in the same way when one partner is not human (e.g., a synthetic face or a smiley). This conclusion is also supported by recent behavioral and ERP evidence showing that schematic/cartoon faces are not processed as deeply as a natural facial expression and that more attentional resources are allocated to natural faces (e.g., Schindler et al., 2017; Kendall, 2019; Zhao et al., 2019; see Section Non-Referential (Visual) Social Cues: Facial Emotions). In our Experiment 2, the natural facial prime may have set up the expectation of a human speaker more than the smiley in Experiment 1, eliciting a better integration with the target agent's facial expression and the positive sentence.

The Reversed Priming Effect

However, when the emotional facial expressions of the competitor character and target agent were more salient, the anticipatory effect of the congruent prime face reversed. Participants were more likely to fixate the grumpy-looking competitor character (vs. happy-looking target agent) during the verb-adverb and long region when having been primed with a positive (vs. negative) speaker face. We did not predict this reversal but identified possible reasons *post-hoc*.

To recap, in Experiment 2, the visual scene that participants saw while listening to the positively valenced sentence contained the happy-looking target agent alongside a neutral-looking patient and a slightly grumpy-looking competitor character. In Experiment 3, we omitted the neutral-looking patient and increased the valence of both the target agent's and the competitor character's emotional facial expressions. This means that the slightly grumpy-looking competitor character was given a really grumpy face (e.g., by lowering the corners of the mouth and lowering and tilting the eyebrows) while the happy-looking target agent's face was made even happier looking (e.g., increasing the size of the eyes, opening the smile, and raising the eyebrows). That is, in Experiment 3, the emotional valence of the target scene was foregrounded and the valence highlighted.

We assume that in both experiments, the congruent happy speaker prime face raised expectations toward an upcoming happy event. In Experiment 2, following this congruent prime face, and with the expectations toward a positive event in mind, participants faced a scene in which the emotionally most salient character is the happy-looking target agent. The slightly grumpy-looking character might not have been as emotionally salient and might hence not have violated the primed positive expectations. Hearing the verb and adverb describing the happy event in the scene, participants quickly used their expectations and the congruent positively valenced linguistic input to direct their fixations toward the also happy-looking target agent. When that agent was named, their expectation that this is indeed the character carrying out the positive action is confirmed.

In contrast, in Experiment 3, following the congruent prime face, participants encounter a scene featuring a very grumpy-looking and a very happy-looking character. Both characters' facial expressions might be salient, but it could be that the really grumpy-looking character drew more attention than the really happy-looking character. This idea is supported by evidence suggesting that negative emotional facial expressions and events draw more attention and are better memorized than positive faces and events [at least for younger adults (e.g., Grühn and Scheibe, 2008; Lamy et al., 2008; Finn et al., 2012; Bach et al., 2014)].

The very grumpy-looking character might also violate the positive expectations more than it did in Experiment 2 (in which the face was only slightly grumpy). Rothermund et al. (2011), for example, showed incongruency effects in affective processing in emotionally mismatching situations. Further findings show that attention is directed to materials opposite in emotional valence to the perceiver's current focus of attention (Rothermund, 2003; Rothermund et al., 2008). This attentional state moreover can

but does not have to be explicitly induced by, for instance, imagining personal positive or negative events or watching emotionally valenced movies (Schwager and Rothermund, 2013). In Schwager and Rothermund (2013), participants first imagined a personal emotional situation (Experiment 1) or watched positive or negative emotional movie clips (Experiment 2). Following the imagination or movie, they performed a visual search task (Experiment 1) or an emotional Stroop task (Experiment 2) featuring positive and negative target words (Experiment 1) or pictures (Experiment 2). The results showed that participants were more accurate in both experiments for opposite-valence targets. Moreover, in both experiments, the reaction times suggested that participants were also faster in detecting opposite-valence targets.

A similar interference effect has recently been reported for real-time language processing (Guerra and Knoeferle, 2017, 2018). In an eye-tracking reading study, Guerra and Knoeferle (2017) investigated how visually perceived spatial distance influences the interpretation of social relationships between agents and patients in sentence processing. Visually perceived spatial distance was established *via* pairs of playing cards, which either moved closer together or farther apart. The nouns referring to the agent and patient in the following German target sentence were printed each on one of the two cards. The card movements primed the written target sentence either expressing a friendly (i.e., close) or an unfriendly (i.e., far) relationship between the agent and patient, such as *Sandra met her aunt cheerfully/grumpily at the health center* (translation). In Experiment 1, in these sentences, the adverb expressing the social relation between the protagonists (i.e., *Sandra* and *her aunt*) appeared after the second noun (i.e., *aunt*). The results revealed shorter reading times in the adverb region when the cards moved close together and the sentence expressed a friendly relationship between protagonists compared to when prime card movement and target sentence did not match in social meaning. Hence, participants experienced a facilitation effect in sentence reading. However, in Experiment 2, Guerra and Knoeferle (2017) decreased the temporal distance between the card presentation and the (mis)matching adverb by moving the adverb in the sentence from after to in front of the second noun phrase (i.e., *Sandra met cheerfully/grumpily her aunt*). The results revealed longer reading times for matching vs. mismatching social meaning. Thus, participants experienced an interference effect. The authors argue that the shift from facilitation to interference is due to an increased competition between two strongly activated representations of the same conceptual representation: When the previously seen playing cards prime the activation of the concept of un/friendliness and the adverb is encountered late after the agent–patient relationship has been established, sentence processing is facilitated. However, when the adverb is encountered before the second noun phrase, the concept of un/friendliness is activated earlier and competes in activation with the same concept activated by the matching playing cards. In the latter case, sentence processing took longer.

This reversal might be similar to the reversed priming effect in the present Experiments 2 and 3: whereas we see increased fixations to the happy-looking target agent in the

adverb region in Experiment 2, we see increased fixations to the grumpy-looking competitor agent in the adverb region in Experiment 3. In Experiment 3, in contrast to Experiment 2, we not only increased the salience of the target agent and the competitor agent's opposing emotionally valenced facial expressions, but also we decreased the preview time of the target scene from 2,000 to 500 ms. Just like in Guerra and Knoeferle (2017), two conceptual representations might be competing for activation because they were activated in temporal proximity, thus reversing the effect. Crucially, in their final experiment, when Guerra and Knoeferle (2017) used the German sentences from their first experiment and inserted an additional neutral word between the second noun and the adverb (transl.: e.g., *Sandra met her aunt unequivocally cheerfully at the health center*), they again found facilitation effects. Hence, the activation time between two similar conceptual representations seems to be crucial for the directionality of the effects regarding real-time language comprehension.

Taken together, the reversal of the result pattern that we see in Experiment 3 could (a) be due to the competitor character's highlighted negative facial expression, which violated the primed positive expectations and the positively valenced sentence. This explanation is also in line with studies showing that semantically inconsistent objects in scenes attract more attention (i.e., more fixations) than consistent objects based on the viewer's expectations and world knowledge about what a particular scene usually contains [e.g., a cocktail glass vs. a microscope in a kitchen scene (Henderson et al., 1999)]. However, following Guerra and Knoeferle (2017, see also 2018), the reversal of the result pattern that we see in Experiment 3 could (b) be also due to the timing of stimulus presentation in our experiments. Since we implemented both changes (i.e., the increase in emotional salience and the reduced target scene preview time), further investigations need to tease apart which of the two factors is responsible for the reversal of the effect in Experiment 3 compared to Experiment 2.

Real-Time Assignment of Thematic Roles via Emotional Facial Expressions

The results from Experiment 2 suggest that participants use the speaker's emotional facial expression to anticipate the correct target agent when they can link the facial expression to the verb and adverb of the sentence, thus filling the agent role slot in real-time processing before the target agent is mentioned. Accuracy scores in Experiment 2 were at ceiling. These ceiling effects might have overshadowed potential off-line effects of the emotional prime face.

Experiment 3, however, seems to paint a slightly different picture compared to our initial interpretation of the results. In Experiment 3, in order to verify whether the content of the SVO and OVS sentences is the same, participants need to have correctly assigned the thematic roles in the sentence. Although this task was more difficult for participants than the passive who-does-what questions from Experiment 2 (77 vs. 96% correct answers), no effects of the prime face on sentence verification

accuracy emerged. Experiment 3's null effect in the sentence verification accuracy plus the reversed result pattern in the eye-tracking data calls our interpretation that the positive prime face facilitates real-time assignment of thematic roles (Experiment 2) into question.

What the eye-tracking data does tell us though is that the linguistic input together with the emotional facial expression modulates attention to the characters in the scene, since the first effects emerged after the NP1 region, triggered by the verb and the adverb. Whether these effects are, however, specific to the assignment of thematic roles is less clear. The reversal of the result pattern in Experiment 3 could simply be due to the enhanced negative facial expression of the competitor character in the scene (see Section The Reversed Priming Effect). It could mean that the assignment of thematic roles can be facilitated by using the expectations based on the face in real time, if the visual context does not portray strongly opposing valence information (i.e., an only slightly grumpy-looking competitor, a neutral patient, and a happy-looking target agent) and/or if similar conceptual representations do not compete for activation in temporal proximity. The expectations to what should come next might be violated by seeing a valence-incongruent character on the screen next to a valence-congruent character. That contrast might have drawn listeners' attention to the valence mismatching (vs. matching) character when the positive verb-adverb region confirmed the expectations derived from the positive prime face (cf., Henderson et al., 1999; Schwager and Rothermund, 2013).

Alternatively, the effects we find in Experiment 2 are due to linking the positive facial expression to the positive sentence, which in turn directs attention to the character with the valence-matching facial expression without any assignment of thematic roles involved. Since the off-line effects did not support the real-time emotional prime effects, it might be argued that participants did not assign thematic roles based on the match between the emotional prime face, the positive sentence, and the target agent's facial expression, but merely performed a "valence match" between the different positively valenced information types. However, note that this conclusion cannot be drawn for the depicted action effect. Even though our data does not reveal significant off-line effects, previous research indicates that depicted actions can clearly be used to facilitate the assignment of thematic roles in real time and posttrial (e.g., Knoeferle et al., 2005, 2008). Our results suggest that utterance-congruent emotional facial expressions of a speaker can modulate attention during real-time language processing. Further research needs to investigate whether emotional face primes also facilitate the assignment of thematic roles.

Finally, the diverging results between Experiments 2 and 3 have crucial implications for psycholinguistic visual-world research: Even slight changes in the materials and timing could change the pattern of results in unexpected ways (e.g., Guerra and Knoeferle, 2014; Knoeferle et al., 2014). Future research may want to replicate these timing- and stimulus-related changes in the directionality of the observed effects and further assess timing and stimulus variability.

CONCLUSIONS

Our results provide some support for the view that a natural facial expression of emotions elicited a priming effect on the interpretation of emotionally valenced OVS sentences (in contrast to a schematic smiley). Crucially, our studies show that different kinds of information in a visual context and distinct language-world relations yield distinct visual context effects on language processing. As one reviewer pointed out, it is, of course, also possible to establish different degrees of referentiality by using different linguistic cues, i.e., the emotion verb *like* is arguably non-referential in contrast to the action verb *kick*. Even though this would be an interesting issue to explore and likewise foster our insights on different degrees of referentiality, limiting referentiality vs. non-referentiality to linguistic expressions alone does not allow us to pit different extralinguistic (social) cues, i.e., emotional facial expressions and depicted actions, against each other. Interestingly, Experiment 3 showed that the exact nature of the situation in which emotional facial cues are embedded seems to influence the directionality of the effects. Clearly, more research is needed, which explores the effects of referential and non-referential (e.g., social) cues and their relation to each other in real-time language processing. Our studies represent a first step in going beyond the integration of referential cues in language comprehension and suggest that subtler and crucially also social cues, such as the emotional facial expression of a speaker, can affect real-time language comprehension. The reported findings pave the way for extant real-time language processing accounts to include the effects of non-referential (social) visual cues, their modulation by the naturalness of facial expressions, and their relative effect differences compared with referential (action) cues.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article can be made available upon request to the first author.

REFERENCES

- Adolphs, R. (2006). How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain Res.* 1079, 25–35. doi: 10.1016/j.brainres.2005.12.127
- Aguado, L., Garcia-Gutierrez, A., Castañeda, E., and Saugar, C. (2007). Effects of prime task on affective priming by facial expressions of emotion. *Spanish J. Psychol.* 10, 209–217. doi: 10.1017/S1138741600006478
- Altmann, E. M., and Trafton, J. G. (2002). Memory for goals: an activation-based model. *Cogn. Sci.* 26, 39–83. doi: 10.1207/s15516709cog2601_2
- Altmann, G. T. (2004). Language-mediated eye movements in the absence of a visual world: the “blank screen paradigm.” *Cognition* 93, B79–B87. doi: 10.1016/j.cognition.2004.02.005
- Altmann, G. T., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1
- Anderson, S. E., Chiu, E., Huette, S., and Spivey, M. J. (2011). On the temporal dynamics of language mediated vision and vision-mediated language. *Acta Psychol.* 137, 181–189. doi: 10.1016/j.actpsy.2010.09.008
- Arai, M., Van Gompel, R. P. G., and Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cogn. Psychol.* 54, 218–250. doi: 10.1016/j.cogpsych.2006.07.001
- Bach, D. R., Schmidt-Daffy, M., and Dolan, R. J. (2014). Facial expression influences face identity recognition during the attentional blink. *Emotion* 14:1007. doi: 10.1037/a0037945
- Barrett, L. F., Lindquist, K. A., and Gendron, M. (2007). Language as context for the perception of emotion. *Trends Cogn. Sci.* 11, 327–332. doi: 10.1016/j.tics.2007.06.003
- Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Calvo, M. G., and Lundqvist, D. (2008). Facial expressions of emotion (KDEF): identification under different display-duration conditions. *Behav. Res. Method.* 40, 109–115. doi: 10.3758/BRM.40.1.109
- Carminati, M. N., and Knoeferle, P. (2013). Effects of speaker emotional facial expression and listener age on incremental sentence processing. *PLoS ONE* 8:e72559. doi: 10.1371/journal.pone.0072559

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Bielefeld ethics board (Vote 2013-007) and ethics board of the German association for linguistics (DGfS, valid from 17.09.2016 to 16.09.2022). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

PK and KM designed the experiments and wrote the manuscript. KM conducted the studies and analyzed the data. The data are available upon request. Analyses of the data from Experiments 1 and 2 have previously been published in Münster (2016) and in parts in Münster et al. (2015). All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Excellence Center 277 Cognitive Interaction Technology (German Research Foundation, DFG).

ACKNOWLEDGMENTS

We acknowledge support by the German Research Foundation (DFG) and the Open Access Publication Fund of Humboldt-Universität zu Berlin. We thank M. N. Carminati for her advice on the study designs of Experiments 1 and 2. Experiments 1 and 2 were run at Bielefeld University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.547360/full#supplementary-material>

- Carminati, M. N., and Knoeferle, P. (2016). Priming younger and older adults' sentence comprehension: insights from dynamic emotional facial expressions and pupil size measures. *Open Psychol. J.* 9:10129. doi: 10.2174/1874350101609010129
- Chambers, C. G., Tanenhaus, M. K., and Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *J. Exp. Psychol. Learn. Memory Cogn.* 30, 687–696. doi: 10.1037/0278-7393.30.3.687
- Chang, Y. H. (2006). *Comparing Affective Perception in Cartoon, Schematic and Real Faces*. Perception [Abstract], ECVF Abstract Supplement, 35. doi: 10.1177/03010066060350S101
- Codispoti, M., Mazzetti, M., and Bradley, M. M. (2009). Unmasking emotion: exposure duration and emotional engagement. *Psychophysiology* 46, 731–738. doi: 10.1111/j.1469-8986.2009.00804.x
- DeCot, B. M. (2011). *The Role of Speaker Facial Expression in Processing Emotional Sentences*. (Unpublished Master's Thesis), Bielefeld University, Bielefeld, Germany.
- Diamond, R., and Carey, S. (1986). Why faces are and are not special: an effect of expertise. *J. Exp. Psychol. General* 115, 107–117. doi: 10.1037/0096-3445.115.2.107
- Finn, B., Roediger, H. L., and Rosenzweig, E. (2012). Reconsolidation from negative emotional pictures: is successful retrieval required?. *Mem. Cogn.* 40, 1031–1045. doi: 10.3758/s13421-012-0203-7
- Frazier, L., and Fodor, J. (1979). The sausage machine: a new two-stage parsing model. *Cognition* 6, 291–325. doi: 10.1016/0010-0277(78)90002-1
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8
- Glenberg, A. M., and Robertson, D. A. (1999). Indexical understanding of instructions. *Discour. Processes* 28, 1–26. doi: 10.1080/01638539909545067
- Grelotti, D. J., Gauthier, I., and Schultz, R. T. (2002). Social interest and the development of cortical face specialization: what autism teaches us about face processing. *Dev. Psychobiol.* 40, 213–225. doi: 10.1002/dev.10028
- Grühn, D., and Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): do ratings become more extreme with age?. *Behav. Res. Methods* 40, 512–521. doi: 10.3758/BRM.40.2.512
- Guerra, E., and Knoeferle, P. (2014). Spatial distance effects on incremental semantic interpretation of abstract sentences: evidence from eye tracking. *Cognition* 133, 535–552. doi: 10.1016/j.cognition.2014.07.007
- Guerra, E., and Knoeferle, P. (2017). Visually perceived spatial distance affects the interpretation of linguistically mediated social meaning during online language comprehension: an eye tracking reading study. *J. Memory Lang.* 92, 43–56. doi: 10.1016/j.jml.2016.05.004
- Guerra, E., and Knoeferle, P. (2018). Semantic interference and facilitation: understanding the integration of spatial distance and conceptual similarity during sentence reading. *Front. Psychol.* 9:718. doi: 10.3389/fpsyg.2018.01417
- Hajcak, G., Moser, J. S., and Simons, R. F. (2006). Attending to affect: appraisal strategies modulate the electrocortical response to arousing pictures. *Emotion* 6:517. doi: 10.1037/1528-3542.6.3.517
- Harwood, N. K., Hall, L. J., and Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental Retardation. *Am. J. Mental Retard.* 104, 270–278. doi: 10.1352/0895-8017(1999)104<0270:ROFEEF>2.0.CO;2
- Henderson, J. M., Weeks, P. A., and Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 210–228. doi: 10.1037/0096-1523.25.1.210
- Hermans, D., Houwer, J. D., and Eelen, P. (1994). The affective priming effect: automatic activation of evaluative information in memory. *Cogn. Emot.* 8, 515–533. doi: 10.1080/02699399408408957
- Huetig, F., and Altmann, G. T. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition* 96, B23–B32. doi: 10.1016/j.cognition.2004.10.003
- Johnson, M. H., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40, 1–19. doi: 10.1016/0010-0277(91)90045-6
- Kamide, Y., Altmann, G. T. M., and Haywood, S. L. (2003a). The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J. Memory Lang.* 49, 133–156. doi: 10.1016/S0749-596X(03)00023-8
- Kamide, Y., Scheepers, C., and Altmann, G. T. M. (2003b). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *J. Psycholinguist. Res.* 32, 37–55. doi: 10.1023/A:1021933015362
- Kendall, L. (2019). *The processing of cartoony and photorealistic faces* (Doctoral dissertation). University of British Columbia.
- Knoeferle, P., and Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cogn. Sci.* 30, 481–529. doi: 10.1207/s15516709cog0000_65
- Knoeferle, P., and Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: evidence from eye movements. *J. Memory Lang.* 57, 519–543. doi: 10.1016/j.jml.2007.01.003
- Knoeferle, P., Crocker, M. W., Scheepers, C., and Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition* 95, 95–127. doi: 10.1016/j.cognition.2004.03.002
- Knoeferle, P., Habets, B., Crocker, M. W., and Münte, T. F. (2008). Visual scenes trigger immediate syntactic reanalysis: evidence from ERPs during situated spoken comprehension. *Cereb. Cortex* 18, 789–795. doi: 10.1093/cercor/bhm121
- Knoeferle, P., Urbach, T. P., and Kutas, M. (2014). Different mechanisms for role relations versus verb-action congruence effects: evidence from ERPs in picture-sentence verification. *Acta Psychol.* 152, 133–148. doi: 10.1016/j.actpsy.2014.08.004
- Kreysa, H., Knoeferle, P., and Nunnemann, E. (2014). “Effects of speaker gaze versus depicted actions on visual attention during sentence comprehension,” in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, eds P. Bello, M. Guarini, M. McShane, and B. Scassellati (Austin, TX: Cognitive Science Society), 2513–2518.
- Kreysa, H., Nunnemann, E. M., and Knoeferle, P. (2018). Distinct effects of different visual cues on sentence comprehension and later recall: the case of speaker gaze vs. depicted actions. *Acta Psychol.* 188, 220–229. doi: 10.1016/j.actpsy.2018.05.001
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Statist. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Lamy, D., Amunts, L., and Bar-Haim, Y. (2008). Emotional priming of pop-out in visual search. *Emotion* 8, 151–161. doi: 10.1037/1528-3542.8.2.151
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* 49, 1494–1502. doi: 10.3758/s13428-016-0809-y
- MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* 101, 676–703. doi: 10.1037/0033-295X.101.4.676
- Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., et al. (1999). Face perception during early infancy. *Psychol. Sci.* 10, 419–422. doi: 10.1111/1467-9280.00179
- Münster, K. (2016). Effects of emotional facial expressions and depicted actions on situated language processing across the lifespan (Dissertation). Bielefeld University, Bielefeld, Germany.
- Münster, K., Carminati, M. N., and Knoeferle, P. (2014). “How do static and dynamic emotional faces prime incremental semantic interpretation? comparing older and younger adults,” in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. Québec City, Canada.
- Münster, K., Carminati, M. N., and Knoeferle, P. (2015). “The effect of facial emotion and action depiction on situated language processing,” in *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (Austin, TX: Cognitive Science Society), 1673–1678.
- Münster, K., and Knoeferle, P. (2018). Extending situated language comprehension (accounts) with speaker and comprehender characteristics: toward socially situated interpretation. *Front. Psychol.* 8:2267. doi: 10.3389/fpsyg.2017.02267
- Novick, J. M., Thompson-Schill, S. L., and Trueswell, J. C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition* 107, 850–903. doi: 10.1016/j.cognition.2007.12.011
- Öhman, A., Lundqvist, D., and Esteves, F. (2001). The face in the crowd revisited: a threat advantage with schematic stimuli. *J. Personal. Soc. Psychol.* 80, 381–396. doi: 10.1037/0022-3514.80.3.381

- Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Sci.* 4, 515–526. doi: 10.1017/S0140525X00076512
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/>
- Recio, G., Sommer, W., and Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Res.* 1376, 66–75. doi: 10.1016/j.brainres.2010.12.041
- Rothermund, K. (2003). Motivation and attention: incongruent effects of feedback on the processing of valence. *Emotion* 3, 223–238.
- Rothermund, K., Gast, A., and Wentura, D. (2011). Incongruity effects in affective processing: automatic motivational counter-regulation or mismatch induced salience? *Cogn. Emot.* 25, 413–425.
- Rothermund, K., Voss, A., and Wentura, D. (2008). Counter-regulation in affective attentional biases: a basic mechanism that warrants flexibility in emotion and motivation. *Emotion* 8, 34–46.
- Ruffman, T., Ng, M., and Jenkin, T. (2009). Older adults respond quickly to angry faces despite labeling difficulty. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 64, 171–179. doi: 10.1093/geronb/gbn035
- Schindler, S., Zell, E., Botsch, M., and Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Sci. Rep.* 7:45003. doi: 10.1038/srep45003
- Schwager, S., and Rothermund, K. (2013). Counter-regulation triggered by emotions: positive/negative affective states elicit opposite valence biases in affective processing. *Cogn. Emot.* 27, 839–855. doi: 10.1080/02699931.2012.750599
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., and Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition* 71, 109–147. doi: 10.1016/S0010-0277(99)00025-6
- Spivey, M. J., and Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychol. Res.* 65, 235–241. doi: 10.1007/s004260100059
- Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., and Sedivy, J. C. (2002). Eye movements and spoken language comprehension: effects of visual context on syntactic ambiguity resolution. *Cogn. Psychol.* 45, 447–481. doi: 10.1016/S0010-0285(02)00503-0
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863
- Trueswell, J., and Tanenhaus, M. (1994). “Toward a lexical framework of constraint-based syntactic ambiguity resolution,” in *Perspectives on Sentence Processing*, eds C. Clifton, L. Frazier, and K. Rayner (Mahwah, NJ: Lawrence Erlbaum Associates), 155–179.
- Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.* 20, 580–591. doi: 10.1162/jocn.2008.20054
- Van Berkum, J. J. A. (2018). “Language comprehension, emotion, and sociality: aren’t we missing something?” in *Oxford Handbook of Psycholinguistics*, eds S. A. Rueschemeyer and G. Gaskell (Oxford: Oxford University Press), 644–669. doi: 10.1093/oxfordhb/9780198786825.013.28
- Van Berkum, J. J. A. (2019). Inclusive affective neurolinguistics. *Lang. Cogn. Neurosci.* 2019:1665191. doi: 10.1080/23273798.2019.1665191
- Zhang, L., and Knoeferle, P. (2012). “Visual context effects on thematic role assignment in children versus adults: evidence from eye tracking in German,” in *Proceedings of the Annual Meeting of the Cognitive Science Society*, eds N. Miyake, D. Peebles, and R. P. Cooper (Boston, MA: The Cognitive Science Society), 2593–2598.
- Zhao, J., Meng, Q., An, L., and Wang, Y. (2019). An event-related potential comparison of facial expression processing between cartoon and real faces. *PLoS ONE* 14:e0198868. doi: 10.1371/journal.pone.0198868

Conflict of Interest: 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.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Maquate and Knoeferle. 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.



Cognitive Processes of ESL Learners in Pragmatic Role-Play Tasks in Academic Settings

Nick Zhiwei Bi*

College of Foreign Languages, University of Shanghai for Science and Technology, Shanghai, China

OPEN ACCESS

Edited by:

Ramesh Kumar Mishra,
University of Hyderabad, India

Reviewed by:

Xuesong Gao,
University of New South
Wales, Australia
Veeky Baths,
Birla Institute of Technology and
Science, India

*Correspondence:

Nick Zhiwei Bi
nick.zhiwei.bi@usst.edu.cn

Specialty section:

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

Received: 23 July 2020

Accepted: 28 July 2021

Published: 30 August 2021

Citation:

Bi NZ (2021) Cognitive Processes of
ESL Learners in Pragmatic Role-Play
Tasks in Academic Settings.
Front. Psychol. 12:586588.
doi: 10.3389/fpsyg.2021.586588

The purpose of the study was to investigate the cognitive processes of English as second language (L2) learners that are involved in their task-based pragmatic performances in academic settings. This study, therefore, examined the cognitive processes of 30 English L2 learners when engaging in various role-play-based pragmatic performances, such as requesting a recommendation letter from a professor and negotiating an agreeable meeting time with classmates. The qualitative analyses of the retrospective verbal reports (RVRs) data of the participants indicated that the learners employed a series of cognitive, metacognitive, and pragmatic strategies when accomplishing various speech acts (e.g., requests and refusals). This study hoped to make two new contributions to the field. First, the study provided empirical evidence to validate the theoretical taxonomy of the strategy use of learners in L2 pragmatics. Additionally, the theoretical foundations of current research on cognitive processes are primarily informed by pragmatic theories. Thus, the study aims to explicate a more comprehensive view of the cognitive processes of L2 learners in pragmatic performances by employing the theories from both pragmatic and learner strategy perspectives.

Keywords: cognitive processes, metacognitive and cognitive strategies, pragmatic strategies, L2 pragmatic competence, task-based language teaching, role-plays

INTRODUCTION

Various communicative language ability models (see Canale and Swain, 1980; Bachman and Palmer, 1996, 2010) have suggested that language ability includes both language knowledge and strategic competence (e.g., the metacognitive and cognitive strategies of L2 learners). These strategies are “conscious or semi-conscious thoughts and actions deployed by learners, often with the intention of enhancing their knowledge of, and facility with an L2” (Ishihara and Cohen, 2010, p. 228). Pragmatic competence has been widely recognized as an essential ability to use language appropriately in a social context (Taguchi and Roever, 2017). Similar to other linguistic competence, pragmatic competence also consists of both the pragmatic knowledge and cognitive processes of L2 learners. With increasing attention on pragmatic performance, in particular, a more comprehensive understanding of the pragmatic competence of L2 learners could be reached if research also focuses on the cognitive processes involved during the pragmatic performance of these learners besides their pragmatic knowledge (e.g., Robinson, 1992; Cohen, 2005; Ren, 2014; Chen, 2015). Such research would yield more information about the reasons underlying language choices and productions related to pragmatic competence (Gass and Mackey, 2000; Li and Gao, 2017).

Currently, one particular focus in L2 pragmatics is understanding pragmatic learning needs in an English for academic purposes (EAP) setting. Studies have investigated the cognitive processes of L2 students that are involved in spoken and written communication in various EAP contexts (e.g., Chen, 2015) and various speech acts, such as apologies, complaints, and requests (Woodfield, 2010, 2012). With the rise of a task-based approach to L2 pragmatic competence (Taguchi and Kim, 2018), pragmatic tasks, such as role-plays, with concrete communicative goals have been considered as viable pragmatic research instruments as they can tap into language use in real-life contexts. Accordingly, the scope of L2 pragmatic competence is well-represented by this model. For this reason, This study investigated the cognitive processes of L2 learners that are involved in role-play-based pragmatic performances in order to reveal these cognitive processes more accurately.

The cognitive processing of an L2 learner is extremely complex and multidimensional (Bi, 2015, 2017). To this end, the study argues that the majority of previous research on the cognitive processes of L2 learners in their speech acts are informed by pragmatic theories rather than the theories of learner strategy research. In order to fully understand the nature of the cognitive processes underlying L2 pragmatic performance, more studies connecting theories in diverse disciplines (e.g., language learning, psychology, and metacognition theories) to pragmatic cognitive processes are necessary.

LITERATURE REVIEW

Cognitive Processes From a Pragmatic Perspective

To date, the majority of studies on the pragmatic cognitive processes of learners have primarily relied on L2 pragmatic theories. In addition, most pragmatic researchers (e.g., Robinson, 1992; Cohen and Olshtain, 1993; Ren, 2014; Chen, 2015) used retrospective verbal reports (RVRs) to examine the cognitive processes in pragmatic production. This group of researchers has revealed that a number of mental activities occur in the minds of L2 learners during different speech acts. Notably, the researchers paid close attention to the cognitive processes of L2 learners regarding their sociopragmatic knowledge and pragmalinguistic knowledge production, which characterize two key theoretical dimensions of pragmatic competence (Leech, 1983). For instance, a longitudinal study by Ren (2014) on the variations in the cognitive processes of L2 learners reported that one effective strategy was the development of the control learners have over-attention to pragmatic knowledge through the application of additional cognitive processes that can control and regulate other processes. However, these studies did not present specific findings related to pragmatic cognitive processes. For instance, there were no detailed explanations about what sociopragmatic or pragmalinguistic cognitive processes were employed by learners in order to successfully complete real-life pragmatic tasks in academic contexts.

Furthermore, a closer look at the most recent investigations of cognitive processes in L2 pragmatic competence shows that

there seems to be a lack of representation in the L2 pragmatic construct. For instance, Chen (2015) used the RVRs approach to examine the cognitive processes of L2 learners in an email task involving different requests to the lecturers in their university. Fifteen Chinese EFL learners reported a number of politeness, planning, and evaluation processes to compose their emails. The contributions of these studies to the current understanding of cognitive processes in pragmatic performance are evident. However, the instruments used in these previous studies do not fully represent L2 pragmatic competence. L2 pragmatic competence entails complex and multi-faceted components, including abilities to take and organize turns effectively when speakers accomplish pragmatic actions in conversation (Kasper, 2006; Taguchi and Roever, 2017). Accordingly, the multi-faceted dimensions of L2 pragmatic competence need to be reflected in instruments if researchers aim to uncover more accurate and comprehensive cognitive processes involved in L2 pragmatic performance.

A study by Li and Gao (2017) was one of the few that attempted to use interactive data such as simulated role-play tasks to explore the metapragmatic awareness of L2 learners in pragmatic performances. The findings revealed that the learners self-monitored their pragmalinguistic and sociopragmatic knowledge, which led to their metapragmatic awareness. Meanwhile, the self-evaluational behaviors of learners also played a role in managing task demand and intentional linguistic choice to respond to a particular communication setting. The study was one of the few studies that examined both performance data and processing data to illustrate what learners said and why in given situations. Nonetheless, the study only focused on the self-regulation dimension of cognitive processes. Thus, further investigations on both self-regulation, as an umbrella notion, and strategies, as concrete mental processes, are needed.

Another insightful recommendation from Li and Gao (2017) is that more in-depth investigations looking at the cognitive processes of learners before and during their pragmatic performances are needed. This view is consistent with Oxford (2017), who suggested that the cognitive processes of learners in language performance would go through the "forethought, performance, and self-reflection phase" (p. 72). However, little empirical research identified how these cognitive processes occurred differently during each phase of the L2 performances.

Overall, although researchers have started bridging the gap between pragmatic performance and cognitive processes, few human cognition theories were applied in the previous studies to understand the cognitive processes underlying pragmatic performance. Hence, aside from L2 pragmatic theories, cognitive processes in the pragmatic competence of L2 learners should be cross-referenced to other research traditions, such as learner strategy research, metacognition, and human information processing theories.

Cognitive Processes From a Learner Strategy Perspective

Since 1990, numerous studies have started documenting the cognitive processes employed by L2 learners. In the literature,

there are many different terms for the cognitive processes of learners, such as learning/learner strategies, cognitive processes, metacognition, and self-regulation (for a review, see Rose et al., 2018). To avoid confusion, the current study used cognitive processes as the macro category for mental processes. However, the cover term, strategies, included the subset of cognitive processes, for instance, the commonly known language learner strategies (e.g., metacognitive and cognitive strategies) and language-related strategies (e.g., pragmatic strategies). This classification also conformed to how pragmatic researchers referred to mental activities as cognitive processes in their research.

Compared to other language skills, pragmatic strategies have drawn relatively less attention (Oxford, 2017; Cohen, 2020). Nonetheless, researchers (e.g., Cohen, 2005) have already developed the taxonomies of pragmatic strategies. In the taxonomy, Cohen proposed the following types of strategies: cognitive, metacognitive, social or affective strategies, communication, and cover strategies learners use in their speech acts. The taxonomy provided us with a useful guideline to investigate cognitive processes in speech acts. However, the taxonomies were based on general language-use situations rather than specific language-use contexts, such as an academic setting. Therefore, more empirical research is needed in order to validate the taxonomies and develop a more fine-tuned measure for cognitive processes in L2 pragmatics (Cohen, 2005, 2020).

Regarding cognitive processes in specific situations, in the past decade, the pragmatic competence of L2 learners in EAP context has begun to draw attention among some strategy researchers (e.g., Cohen and Sykes, 2013; Youn and Bi, 2019) due to the increasing numbers of international students pursuing a University degree in English-speaking countries. The research reported that many international students may not be adequately prepared for cultural, societal, and interpersonal communications. Cohen and Sykes (2013) recommended that explicit attention to language learner strategies in the field of intercultural pragmatics and intercultural education could enable students to deal with complex real-life situations in their academic life. In their study on the effectiveness of strategy-based instruction, the results suggested that familiarizing learners with strategies would make a difference in their pragmatic performance. The study by Youn and Bi (2019) was one of the few empirical studies that quantitatively investigated how L2 learners at varying levels of pragmatic performance used metacognitive, cognitive, and pragmatic strategies to complete a range of pragmatic tasks in an academic setting.

Despite the scarcity of research, these studies have provided us with some empirical evidence of cognitive processes in L2 pragmatic performances from the learner strategy perspective. In order to have a more accurate understanding of the cognitive processes of L2 learners, empirically validated pragmatic tasks should be adopted as research instruments (Cohen and Sykes, 2013). The current study argue that a pragmatic task, such as a simulated role-play, can be used to examine the cognitive processes of learners during pragmatic interactions.

Investigating L2 Cognitive Processes Using a Task-Based Approach

In terms of research methods in L2 pragmatics, a discourse completion task (DCT), one of the most popular data elicitation methods, has been predominant. A typical DCT involves a brief situational description and asks participants to respond with a speech act utterance (e.g., apology, refusal) either in a spoken or written format. Since participants are asked to provide a single response within a planning time, they do not directly interact with interlocutors while completing DCTs. Thus, DCTs are practical to administer as researchers can systemically fluctuate contextual variables in the scenarios. For example, Ren (2013) used computer-based DCTs to examine the refusal strategies and cognitive processing of learners. Ren designed eight DCT scenarios with photos and audio conversations to increase the authenticity of the situations. However, the single-turn refusal responses of learners in DCTs do not involve the online pragmatic performance in spoken interaction. Despite the common use, DCTs are not suitable for eliciting a wide range of L2 pragmatic abilities (Golato, 2003; Kasper, 2008; Cohen, 2020).

Alternatively, L2 pragmatic researchers have started paying attention to a task-based approach to designing research instruments that reflect real-life communicative situations and enable learners to negotiate for meaning to elicit a multi-dimensional scope of pragmatic competence (e.g., Taguchi and Kim, 2018). Despite such strengths, pragmatic tasks need to be validated. This means that identified pragmatic tasks need to be meaningful for learners to elicit a full scope of L2 pragmatic competence. To this end, explicit attention to gathering valid evidence of pragmatic tasks used in previous research on pragmatic strategy is still lacking. For example, in her study on the pragmatic cognitive processes of Chinese learners, Chen (2015) used an email request task to a professor, which is known to be meaningful for learners in an academic context. Nonetheless, it is only the request being in a specific written genre (i.e., email) that limits the generalizability of the findings. Ways in which learners manage task-specific communicative demands while interacting with interlocutors to accomplish pragmatic actions remain unknown as well. It is highly possible that learners rely on distinct pragmatic strategies to cope with the cognitively demanding nature of online task-based pragmatic performances.

In order to address these research gaps, the study relied on validated role-play tasks involving several speech acts (e.g., request, refusal, agreement, and disagreement) (Youn, 2015) in this study. Youn gathered a range of validity evidence for role-play tasks, which were designed based on the task-based pragmatic learning needs of EAP stakeholders (e.g., students and teachers) (Youn, 2018). Quantitative and qualitative evidence suggested that the role-play tasks elicited a wide scope of pragmatic competence, ranging from pragmalinguistic knowledge to interactional resources. Taken together, this study argues that the role-play tasks used in this study adequately measured L2 pragmatic competence and are accordingly suitable for the empirical validation of task-based cognitive processes.

Since previous studies have not systematically researched the pragmatic performances and cognitive processes of L2 learners in real-life situations, studies using empirically validated role-play

tasks would reveal more meaningful and accurate cognitive processes. Furthermore, no matter how learners apply pragmatic knowledge in real-life situations, this knowledge contains the complex and distinct cognitive processes that depend on when they are used, such as before and during a performance (Li and Gao, 2017). Therefore, cognitive processes need to be examined in terms of the different stages of a pragmatic performance (e.g., cognitive processes before and during the pragmatic performance). To this end, this study answered the following research questions.

1. What metacognitive and cognitive strategies did L2 learners employ to deal with task-based pragmatic performance?
2. What pragmatic strategies did L2 learners employ to deal with task-based pragmatic performance?

METHODS

Participants

Prior to the data collection, the Institutional Review Board (IRB) in a University in North America had approved the ethical application of this study. In order to comply with the guidelines set down by the IRB, the researchers first announced the research and called for research participants at an English Intensive Program and other degree-seeking programs. Thirty international students with varying first language (L1) backgrounds voluntarily participated in the study. Upon completion, each participant received monetary compensation (20 US\$). The selection criteria included the participants being L1s, English proficiency, and the length of living in English-speaking countries. Although the University had a large percentage of students from a similar cultural background (e.g., Chinese), this study aimed to explore the pragmatic performances of students within various cultures. Therefore, the dominance of a specific L1 group was minimized. In the study, the L1s included Chinese ($n = 11$, 37%), Russian ($n = 4$, 13%), Arabic ($n = 5$, 17%), Indonesian ($n = 4$, 13%), Romanian ($n = 2$, 7%), Spanish ($n = 1$, 3%), Portuguese ($n = 1$, 3%), Urdu ($n = 1$, 3%), and Hindi ($n = 1$, 3%) participants.

Given that the role-play tasks in this study represented real-life communication in an academic setting, at least a low-intermediate level of English proficiency was required for all participants. According to the alignment of Test of English as a Foreign Language (TOEFL) Internet-based test (iBT) Scores with the Common European Framework of Reference (CEFR) levels, the TOEFL scores of lower-intermediate level learners are between 57 and 86 and advanced learners are above 110. In this study, the TOEFL iBT scores of the learners ranged from 57 to 112. Of the 30 participants, 14 were graduate-level students; 16 were undergraduate students. For the purpose of this study, the pragmatic performance level was determined based on the rating of two trained raters using the rating criteria from previous studies (Youn, 2015; Youn and Bi, 2019). The pragmatic performances of the participants were rated based on content delivery, appropriate language use, sensitivity to the situation, engaging interaction, and turn organization. Therefore, the criteria were different from other general language proficiency tests. The study identified 17 high-level and 13 low-level learners

TABLE 1 | Pragmatic performance levels of participants.

Levels	ID (Gender F/M)
High	ID 1 (F), ID 3 (F), ID 4 (F), ID 5 (F), ID 6 (F), ID 7 (F), ID 8 (F), ID 9 (M), ID 10 (M), ID 12 (F), ID 15 (F), ID 16 (F), ID 20 (F), ID 21 (F), ID 23 (F), ID 24 (M), ID 28 (F)
Low	ID 2 (M), ID 11 (M), ID 13 (F), ID 14 (F), ID 17 (M), ID 18 (M), ID 19 (M), ID 22 (F), ID 25 (M), ID 26 (M), ID 27 (M), ID 29 (F), ID 30 (F)

(see **Table 1**). However, lower-level learners did not represent true beginner learners.

Role-Play Tasks

All participants completed three pragmatic role-play tasks (two professor role-plays and one classmate role-play). The role-play tasks reflected real-life situations in an academic context that require pragmatic competence. The validity of the role-play tasks, in terms of their quantitative function and construct representation, was examined (Youn, 2015). Through the large-scale needs analysis on task-based pragmatic learning in an EAP context (Youn, 2018), the most needed and relevant situations that various stakeholders (students and teachers) wanted to learn were identified. The role-play situations involved two interlocutors, professor, and classmate (see **Appendix A**). For the professor situation, the participants completed two speech acts. First, they requested a recommendation letter with a short due date from a professor in a visit during office hours. Secondly, they refused the request of the professor to change an upcoming class presentation schedule due to a scheduling conflict. As for the classmate role-play, two participants played as classmates who were working on a class project and interacted on how and where they planned to meet for an upcoming project. In order to ensure some degree of authenticity and standardization, role-play cards for each participant were not shared and various contingencies were embedded in the role-play cards. For example, for the refusal role-play with a professor, the participants did not expect a request from a professor. They were given a schedule constriction (i.e., an upcoming final) which most likely led them to refuse the request of the professor.

Data Collection Procedure

Task Completion

Each participant individually met with the professor interlocutor to complete the two role-plays with professors. Two participants were randomly assigned to complete the classmate role-plays, with the proficiency level controlled. The order of role-play tasks was counterbalanced. Each role-play interaction lasted ~2–3 min on average.

Retrospective Method for Eliciting Strategic Processing Data

In order to elicit the strategy use of each participant before and during the completion of the tasks, the current study used stimulated-recall to ask learners to report their thinking processes at various stages. In language learner strategy research, quantitative questionnaire analysis has been a dominant approach (Rose et al., 2018). However, if we

conducted a preliminary investigation into the mental activities of learners, especially to explore under-researched abilities such as pragmatics, a qualitative approach would provide much more in-depth information (see Woodrow, 2005; Tseng et al., 2006; Rose et al., 2018). In this study, the participants first practiced reporting their strategy use after reading the instructions of each role-play situation. Then, each participant completed the first role-play situation. To minimize the memory effect, each participant was asked to report the strategies right after completing each role-play task. A set of questions was prepared, which was consistently used for all participants. However, for some participants who were noticeably answered the minimum in their responses, more guiding questions (e.g., asking to explain more) were asked. The data were all audio-recorded and transcribed for subsequent analysis.

Coding Reported Strategy Use

The purpose of the qualitative analysis of the elicited verbal reports data was to identify distinct types of strategies across the participants. For the analysis of the processing data, various steps with a bottom-up approach were taken to identify the distinct cognitive processes.

As a first step, the researchers examined the data to get a general sense of what was happening in the data. Then, the apparent features and themes that emerged from the data were identified. For example, in a response to the question of the researcher to report any strategy utilized during the role-play request performance, one of the participants reported: *"I really want to give a space for the professor to think or decide, not to push. So I kinda try to give options or maybe like if the professor cannot do it's totally okay too."* The researchers agreed that this response contained an independent thought that represented an explicit strategy. Here, the learner listed explicit social rules (e.g., not to push and giving options for a professor to consider when responding to the request of a student) when talking with a professor. As the researchers went through the rest of the data, the orientation of the participants to appropriate social rules in specific contexts was noticeable, which resulted in a concrete

strategy type, situation-related sociopragmatic strategy (see the Results section for details).

While examining the retrospective data of the learners, a list of strategy uses and cognitive processing literature was consulted in establishing a coding scheme for this study. The study referred to investigations in pragmatic strategy use research (e.g., Cohen, 2005; Ishihara and Cohen, 2010) and cognitive processing research in pragmatics (e.g., Robinson, 1992; Cohen and Olshtain, 1993; Ren, 2014; Chen, 2015; Li and Gao, 2017). These strategy taxonomies and existing coding schemes were considered as a prerequisite for comparability with the current study.

Lastly, the codes for distinct strategies that appeared across participants at different performance levels were identified. After multiple rounds of careful revisions, the final coding scheme (see **Appendix B** for the detailed coding scheme) was developed. The researchers coded half of the data from the verbal reports until accuracy rates over 90% in terms of inter-coder reliability were accomplished. In doing so, a cyclical process of coding, which involved the trial of coding schemes, revision, and recoding the data, was taken to ensure coding accuracy. After finalizing the coding scheme, the researchers independently coded the remaining data. While some undecided coding cases were found, the researchers examined the cases together to reach a final agreement.

RESULTS

This study examined the cognitive processes of learners underlying task-based pragmatic performance. The analyses mostly focused on the retrospective data of higher-level learners as they were more likely to elicit a wide range of strategies to effectively complete the tasks. The results revealed the emergence of different cognitive processes. Based on the existing literature and verbal report data from the students, the following strategies were identified: cognitive, metacognitive, and pragmatic strategies. **Table 2** illustrates the number of participants and

TABLE 2 | Raw frequency of strategy use.

Categories	Strategies	High (<i>n</i> = 17) frequency (% out of total number of strategies)	Low (<i>n</i> = 13) frequency (% out of total number of strategies)	Total (<i>n</i> = 30) frequency
Cognitive strategies	Comprehending	21 (57%)	16 (43%)	37
	Linking to prior knowledge or experiences	20 (77%)	6 (23%)	26
	Recalling appropriate L2 linguistic knowledge	3 (43%)	4 (57%)	7
	Put yourself in the task situation	11 (85%)	2 (15%)	13
Metacognitive strategies	Setting goals	20 (59%)	14 (41%)	34
	Evaluating performance	38 (59%)	26 (41%)	64
	Evaluating the execution of plans	5 (83%)	1 (17%)	6
	Evaluating emotional state/effect	20 (65%)	11 (35%)	31
Pragmatic awareness and strategies	Assessing task-related situations	36 (69%)	16 (31%)	52
	Pragmatic awareness	17 (63%)	10 (37%)	27
	Situational-related pragmalinguistic strategy	41 (84%)	8 (16%)	49
	Situational-related sociopragmatic strategy	49 (58%)	35 (42%)	84
	Situational-related interactional strategy	11 (73%)	4 (27%)	15

their raw frequency of strategy use. Using situation-related sociopragmatic strategies, evaluating performance, and assessing task-related situations were the most frequently reported strategies. It can be noted that high-level learners reported more strategy use in almost all categories. With regard to cognitive strategies, the most common strategy used by higher- and lower-level learners was comprehending, while, unlike, lower-level learners, higher-level learners tended to link their prior knowledge and put themselves in the task situations. In terms of metacognitive strategies, there was a clear trend that higher-level learners were more strategic, since they were more likely to set plans, evaluate their plans, performances, and emotions, and assess task situations. For the pragmatic strategies, both groups employed a number of situation-related sociopragmatic strategies. However, higher-level learners differed from lower-level learners when utilizing substantially more situation-related pragmalinguistic strategies during task completion.

The following section will present transcriptions of the verbal reports illustrating the strategy use and differences in the approach used by higher-level and lower-level learners. These excerpts were chosen for the insight they provide into the cognitive processes of these students in completing various pragmatic tasks.

Reported Cognitive Strategies of Learners

For the purpose of this study, cognitive strategies referred to conscious mental activities when using language and world knowledge to complete pragmatic tasks. The following types of cognitive strategies were identified in the study.

Comprehending

In general, comprehending strategies are commonly used in a wide range of L2 language-use situations, as learners need to identify main ideas and the attitudes of speakers, translate, predict, and make inferences in their language use (Phakiti, 2007). In this study, comprehending strategies often occurred before the learners started the task performance at the pre-task stage, as illustrated below:

I read the situation introduction and the task for us to complete and then I will see the content of the task like, then I will see what suggestions can I provide to him based on my knowledge or experience. (ID21 Pre-task)

ID21 tried to comprehend the role-play tasks. She reported that she understood the task situations and gained a full understanding of the requirements quickly. Although the wording of the task requirements was simple, the higher-level students appeared to start making inferences about the communicative situations in the tasks. The data also suggested that the comprehending of learners usually goes along with their pragmatic awareness (see the section on pragmatic strategies). For instance, ID16 indicated that, after understanding the tasks, she started to think about “*how to be courteous and at the same time be in good relationship with a professor.*”

During the role-play tasks, the learners also reported comprehending strategies frequently. For example, when

interacting with another participant on a classmate role-play, ID16 tried to understand what an interlocutor meant and then used proper “neutral phrase” to reply, as illustrated below:

Sometimes I understood what she said but I thought maybe that's not what she really thinks and I should change somehow the flow of the conversation, and maybe not and that's why I tried to use some neutral phrases and then she replied again and then from those phrases I understood that, okay, that's what she means exactly, and I just followed her ideas. (ID16 Classmate Interaction)

Linking to Prior Knowledge or Experiences

This cognitive strategy was prominent among all learners in terms of its occurrence rate and reported use throughout the different stages of their performances. The learners reported that they dealt with task situations based on their personal knowledge or prior experiences. Since the role-play tasks reflected commonly occurring academic language-use situations, many learners tended to refer to their own experience from their lives in university. At the pre-task stage, ID21 reported:

Yeah because we had some like group projects to do in classes and we had to discuss with our group members about the meeting and yeah I think I had some previous experience like this. (ID21 Pre-task)

When the learners completed the classmate role-play, which involves the negotiation of a meeting time for a group project, ID9 noted the preference of choosing a comfortable schedule, as seen below:

But for me personally if I had a set schedule or something, I was going to stick with that schedule because it's what more comfortable for me. (ID9 Classmate Interaction)

Recalling Appropriate L2 Linguistic Knowledge

The learners used this strategy to invoke L2 linguistic knowledge, such as appropriate vocabulary and grammar. Then, the learners analyzed different linguistic choices for the interactions. For instance, ID17 reported that, to be polite and respectful to the professor, he used particular words and grammatical conventions over others. The higher-level learners tended to retrieve their linguistic knowledge and use the knowledge appropriately:

Yeah to make sure that my sentence is clear that I get vocab and get grammar and I can pronounce the word to make him think it's clear and understand. (ID17 Professor Interaction)

However, for lower-level learners, they mainly focused on only linguistic knowledge. For instance, ID11 mentioned in the task completion: “*[I thought about] grammar and spelling words because I'm not so good at English, as you can see.*” In contrast, the higher-level learners tended to focus on pragmalinguistics and sociopragmatics (see discussion on pragmatic strategies) rather than just linguistic knowledge when performing pragmatic tasks. This finding is consistent with previous research suggesting that upper-intermediate-level learners think about pragmatics more than linguistic planning (Hassall, 2008).

Put Yourself in the Task Situation

This strategy was applied when the learners imagined themselves in a particular situation to solve various real-life communication problems. This strategy appears to be unique to role-play tasks. Compared with other tasks (e.g., DCT), the task-based approach creates a *de facto* real-life situation. Consequently, the learners were more likely to imagine themselves as characters in the performance. The higher-level learners utilized this strategy during the pre-task stage, as illustrated in the following example:

I was just thinking about myself when I'm really in that situation. So what I would say first since I don't know what the role card will be so I was just thinking the opening and maybe the closing part. (ID6 Pre-task)

Meanwhile, the learners also employed this strategy throughout their task performances. For instance, when talking to the professor, ID4 reported to put herself in similar academic conversations, which helped her to respond to the professor. She noted the following.

I'm just understanding the content and then think about the situation, put myself in that kind of situation. If I were in this kind of condition what would I say and again it is easier because you give me your question, I just need to respond. So that's the strategy I guess (ID4 Professor Interaction)

Furthermore, the learners constantly thought about what they would do if similar situations had occurred. The results repeatedly showed that this strategy can effectively help students to be more capable of coping with academic communications properly, if such a strategy is appropriately employed.

Reported Metacognitive Strategies of Learners

Metacognitive strategies refer to conscious goal-directed processes and are considered as an executive cognitive function that regulates the thinking and decision-making processes of a person during a pragmatic performance. In this study, the following metacognitive strategies were reported:

Setting Goals

This strategy reflected the degree to which the students felt they utilized the higher-order capacity of identifying explicit objectives and goals before or during the task-based pragmatic performance. This strategy was mostly employed at the pre-task stage or at the beginning of the performance. The following excerpt shows that the higher-level learners were likely to set a clear goal to follow or to be guided by certain appropriate pragmatic norms. However, this was mainly observed only in the performances of higher-level learners. For instance, ID5, at the pre-task stage, set task completion goals, which included being respectful, figuring out what needs to be accomplished, and identifying the role-play task expectations as reported in the excerpt below:

I think I tried to be like, basic standard is try to be respectful to the partner, like the professor or my classmates, and tried to have the real conclusions, like if we want to discuss about the meeting time, we need to make it short and make it an efficient conversation, not like we talk a lot and we don't know when we finally meet. (ID5 Pre-task)

Evaluating

Evaluating strategies refer to the "past and current actions or performance, such as assessing levels of difficulties, self-questioning, evaluating performance/product accuracy" (Phakiti, 2007, p. 3). In this study, the learners reported to evaluate their performances, execution of plans, and emotional status.

Evaluating Performance

The higher-level learners continuously evaluated their performances for potential areas for correction. Apart from evaluating their speaking performances in general, the learners tended to evaluate whether or not their performances were appropriate by linking their pragmatic processes (see the following section on pragmatic strategies). As the excerpt below shows, ID15 constantly evaluated her performance throughout the tasks. She expressed concern over whether appropriate social norms were applied in her interaction with the professor. At the same time, she was very conscious about the accuracy of her language. The excerpt demonstrated that higher-level learners know what particular social and linguistic knowledge should be implemented in the performance and what should be avoided.

I said more questions whether I'm asking questions correctly, if it is a polite way of addressing the professor, and whether the order it's of as it was for some point in previous tasks in the written emails, can I know where your office is, not where is your office. I'm all the time monitoring for these interactions because it's a little bit different from my first language, Russian, so I have to make sure, and especially articles as well. I have to make sure I have articles everywhere. I'm sure I made many mistakes everywhere. (ID15 Professor Interaction)

Evaluating Execution of Plans

Due to the highly cognitive demanding nature of the role-play tasks, the learners evaluated whether their plans went as planned. For instance, apart from evaluating the role-play performance itself, ID15 reported that she constantly thought about the initial goals to make sure that what she had to say and how she said it was done properly. As seen below, ID15 set the goal of interacting with the professor appropriately. In order to successfully execute this plan, she needed to stay on track to find ways to help the professor to complete the role-play.:

Before I think the goal was achieved as well because professor ask me whether I have time and according to the case I didn't have time. And also trying to be polite kind of politely say no. so it was fine with the prof as well. And um so at the end we were trying to find an option and a chance how to, I was trying to find a way to help the professor to deal with the situation. (ID15 Professor Interaction)

Evaluating Emotional State/Affect

The learners evaluated emotional state/affect during the pragmatic performance. For instance, ID15 consciously noticed herself feeling tense in the performance as shown below:

I was a bit nervous. That's how I feel all the time with talking to the professor, I'm all the time concerned with making a mistake and trying to monitor like output, so and its kind of all information have to be polite, provide exact information when trying to what exactly I want to tell the prof. and it's all this information, maybe that's why I'm a bit nervous. (ID15 Professor Interaction)

This excerpt shows that the L2 learners were engaged in self-evaluating while still performing the pragmatic tasks. Interestingly, some very strategic learners showed the ability to ease nervousness by quickly shifting their focus back to the tasks. For instance, when interacting with classmates, ID15 evaluated the situation and calmed herself down: "I was feeling much more comfortable with my classmate because I didn't feel this kind of barrier between the professor and the student" (Classmate Interaction). After knowing that the role-play involved two close friends, ID5 again evaluated the emotional status at the moment: "So I knew that Phoenix is my best friend and I have nothing to be afraid of and in this way I was really relaxed, I was thinking clearly, I didn't have any fear inside" (Classmate Interaction).

Assessing Task-Related Situations

The last type of metacognitive strategy found in this study was assessing task-related situations, which acts as a higher-order executive function. The excerpt below indicates that the learners often evaluated the complexities involved in refusing a request from a professor.

Rejecting the request from professor was harder. I asked to write... Is it the same scenario? Okay. I asked your help. And then you ask me to do something and I couldn't you know be reciprocal. So I feel like really sorry. (ID7 Professor Interaction)

Since the role-play tasks elicit task-based pragmatic performance with spoken interaction, the learners monitored the responses of the interlocutors and assessed their expectations. As seen in the excerpt above, ID7 assessed the difficulty of rejecting the request of a professor and elaborated why it was not expected to reject it, especially after the professor agreed to write a recommendation letter. It is worth noting that this strategy was not previously identified in any theoretical taxonomies or empirical research. This study suggests that higher-level L2 learners assess task-specific interactional contexts to ensure context-appropriate speech act performance.

Reported Pragmatic Awareness and Strategies of Learners

Pragmatic cognitive processes refer to conscious mental processes involved in task-based pragmatic performances, including the general awareness and online processes of learners. The four types of pragmatic cognitive processes found in this study were closely related to various components of pragmatic knowledge.

Pragmatic Awareness

The learners reported that pragmatic awareness was related to either the target language or the individual culture of the learners, which occurred during the performance rather than at the pre-task stage. For example, ID5 (higher-level) reported the language use involved in his interaction with his professor as "keep it somewhat formal, but informal at the same time" by explaining his current relationship with a professor as a graduate student in the culture of the target language, which is different from his previous relationship in his own culture, as seen below.

So mainly it's more professional kind of environment when you use the first name. And with my professors... and you have open hours when you come in and you talk to the professor most of the time, you really become friends... I tried to (ID5 Professor Interaction)

ID10 displayed pragmatic awareness during the classmate interaction focusing on the different degrees of formality in the US compared his L1 culture.

In my mind, I think at this point after being in the US an English speaking country for more than a year I was not having too much to strategize right now. But I remember when I just came to the US I had similar situation. So the culture back home is like totally different. And people are more informal. That's what I feel. (ID10 Classmate Interaction)

Other learners also reported thought processes related to their own culture and linguistic repertoire in L1, which contrast with L2 English.

In my own language, in Indonesian language there are steps of words. So this word is higher in position than this one although the meaning is the same. So if I want to talk to the professor or something, I use the higher word. (ID9 Professor Interaction)

In Taiwan we don't even call the last name. We just call, "hi teacher, hi professor." And then also how to use this polite form... we use "please." It already sounded like you're polite. But in English it's also you have to use a lot of hedging. (ID7 Classmate Interaction)

ID9 was oriented to the polite words in her own L1, which led her to realize polite words in English. With regard to address terms, ID7 reported a difference between her own culture and the target culture, displaying an explicit awareness of cross-linguistic differences. Although the amount of information provided is different depending on the levels of the learners, the majority of learners, regardless of their levels, displayed pragmatic awareness.

Situation-Related Pragmalinguistic Strategy

The learners were oriented to two important dimensions of pragmatics, pragmalinguistics, and sociopragmatics (Leech, 1983). Pragmalinguistics concerns the linguistic means necessary to accomplish pragmatic meaning and comprehend the utterances of speakers. Similar to pragmatic awareness, the learners reported pragmalinguistic strategies during the performances and not at the pre-task stage. As seen in the excerpts below, the learners actively referred to various linguistic

repertoires, such as formulaic expressions and modal verbs. ID9 noted “trying to be more polite. ... mostly primarily through the intonation when talking to a professor.” At the same time, they shifted the formality when talking with the classmate by using “Direct [language],” such as “Okay what about this? Okay I like that (ID9).” With regard to linguistic resources specific to speech acts, the higher-level learners reported different resources used when refusing the request of a professor. For example, ID7 provided very explicit linguistic resources as seen in the following.

I didn't think I could tell my professor in the face. ... I'd always still say, "I'll check" and indeed I will check and they maybe email my professor and say, "I'm sorry that they said to be on time" and then just kind of "can you please help me." But... I won't say, "I won't do it, I don't think so." (ID7 Professor Interaction)

ID7 specified numerous linguistic means, such as an elaborated account and indirect expressions using modal verbs. These higher-level learners were very strategic in terms of choosing appropriate linguistic resources in given contexts. Although this study primarily focused on higher-level learners, it is worthy to note that there was a noticeable absence of pragmalinguistic strategies among lower-level learners.

Situation-Related Sociopragmatic Strategy

Parallel to pragmalinguistics, the learners were also oriented to the sociopragmatic. Sociopragmatics focuses on the understanding of social rules and contextual variables that influence language use and interpretation (Leech, 1983). In other words, one needs to understand social and contextual variables, such as a relationship between speakers and the context in which situations occur, to be pragmatically appropriate. It should be noted that the sociopragmatic strategies reported were somewhat different from the pragmatic awareness strategies discussed above. The sociopragmatic strategies contain the orientation of the learners to specific contextual variables, such as the degrees of imposition and the different relationships between interlocutors. Interestingly, compared to other pragmatic strategies, the learners were oriented to the sociopragmatic dimension even before completing the tasks. The excerpt below shows that ID6 noticed the different degrees of formality for each role-play and explained the informal nature of classmate role-play because the conversation was between friends.

I think this will come more naturally than the previous one because this one is more formal so this one says informal so I didn't really think much. This is just with my friend. (ID6 Pre-task).

Right after the performance, the learners actively reported the sociopragmatic strategies they utilized during their performances. ID10 expressed an explicit awareness of the fluctuation of the manner of speaking depending on the relationship with an interlocutor.

I think it's a lot different because when talking to the classmate I was kind of assertive and I said "okay, this is what I want" and I might want your opinion on it, but maybe I'm not able to

change my opinion. But with a professor I was choosing which word say that I'm okay with anything that you're doing. (ID10 Professor Interaction)

Unlike the pragmalinguistic resources, the lower-level learners also displayed sociopragmatic strategies. ID13 (low-level) reported “because you are a teacher I sort of keep polite.” Although this reported sentence was noticeably shorter compared to those of higher-level learners, it still included a reference to the relationship between a speaker and a hearer (i.e., teacher), which is an important situational variable in the reason for being polite in given situations.

Situation-Related Interactional Strategy

Finally, situation-related interactional strategies were identified, which primarily characterize the explicit strategy of learners in dealing with interactional demands in conversation. It should be noted that no one mentioned this strategy before the performance or during the pre-task stage. All identified interactional strategies were reported right after the performance. Because the targeted performances involved interacting with various interlocutors, the learners managed the conversational demands consciously. For example, as seen below, some learners were oriented to the appropriate order of sequencing the turns. ID5 listed how he intentionally placed the greeting sequence (“asking how the week was”) at the beginning of the conversation with a professor, which displays his knowledge of turn-taking rules in conversation.

So informal would be I'm using the first name, I'm asking how the week was, was it busy or not. That would give me a better understanding if the professor can accomplish the ask that I want him to do. So I asked at the very beginning. If the professor would tell me he's busy, I would change my way of deliverance, of my inquiry from the very beginning. (ID5 Professor Interaction)

ID7 also noted that she displayed hesitation when refusing the request of a professor using distinct prosodic properties. ID7 displayed an explicit knowledge of turn-taking and noted “if you say, “oh that would really be great if you can do it” and I would say, “okay, I would try, I would try.” So I already feel that that part of me already prepared.” This reported strategy indicates that ID7 possessed an explicit knowledge of providing an appropriate answer to a previous turn (i.e., turn-taking).

DISCUSSION AND LIMITATIONS

A wide range of cognitive processes was identified in the study, including processes, discussed in existing taxonomies and empirical studies, and other processes uniquely reported before and during the task-based performances. Regarding the cognitive processes from the learner strategy perspective, according to Cohen and Sykes (2013), little research on the strategies of learners in L2 pragmatic performances is available. The present study provided an empirical basis for the taxonomy by Cohen (2005) of pragmatic strategy and what L2 pragmatic competence entails. For instance, in the study, the learners employed

cognitive strategies to assist themselves in retrieving linguistic knowledge and experiences to complete the pragmatic tasks. At the same time, they utilized metacognitive strategies to monitor and evaluate their pragmatic knowledge application. The learner strategy literature in other language skills have suggested that metacognitive strategies are higher-level processes that regulate cognitive strategies (e.g., Phakiti, 2007; Bi, 2015, 2017, 2020). The current study also found similar patterns in pragmatic performances. Furthermore, it is interesting to note that the metacognitive and cognitive strategies occurred throughout the task performances, that is, before and during the performance. This provides further evidence that pragmatic competence is not only composed of knowledge, but cognitive processes play a vital role in this competence.

The study has also revealed additional cognitive processes that are rarely discussed in previous research and theories. For instance, cognitively, the learners tended to put themselves into the task situations; metacognitively, they were more likely to assess the task-related situations. These strategies may not be exclusively reported in our study, but they are certainly key strategies for coping with role-play pragmatic tasks. Role-play tasks are more cognitively demanding compared to other pragmatic instruments (e.g., DCTs) and involve resolving real-life communicative situations, which, in turn, evoke a wide range of L2 pragmatic knowledge (Kasper, 2008). Accordingly, it can be argued that the learners are in need of utilizing additional strategies for task-based pragmatic performances.

The pragmatic strategies reported in the study illuminated the cognitive processes underlying pragmatic performance and the nature of L2 pragmatic competence. The pragmatic awareness of learners reflected the recognition of appropriate sociocultural and linguistic norms both in their own culture(s) and target-language culture. Pragmatic awareness stored in the long-term memory of learners potentially contributes to the regulation of online situation-related pragmalinguistic and sociopragmatic strategies in a concerted effort to perform L2 speech acts. Between the pragmalinguistic and sociopragmatic dimensions, the learners, regardless of their levels, utilized sociopragmatic strategies, which were also attended both before and after the pragmatic performance. On the other hand, pragmalinguistic strategies were more commonly reported among the higher-level learners. In addition, with regard to the interactive nature of role-play performance, the findings revealed that the high-level learners actively utilized the interaction-related strategies. This means that the learners were aware of how to utilize interactional repertoire (e.g., how to start the conversation, how to sequence context-relevant information, and how to respond to a question) in a context-fitting manner for a successful pragmatic performance. This finding supports the current expanded view of L2 pragmatics-in-interaction, which includes the ability to jointly accomplish pragmatic actions contingent upon an unfolding course of conversation (Kasper, 2006; Taguchi and Roever, 2017).

Another contribution of the current study is that, through a comprehensive investigation from the perspectives of both learner strategy and pragmatic research, the findings have provided further empirical evidence for the exceptionally complex and interrelated nature of the cognitive processes of

learners. The results highlighted the role of L2 metacognitive strategies for pragmatic awareness. Li and Gao (2017) emphasized the role of self-monitoring and self-evaluative behaviors on the pragmatic awareness of learners. This study further endorsed the view. For instance, we found that the higher-order processes, such as metacognitive and pragmatic awareness, taking a concerted effort would significantly impact the pragmatic choices of L2 learners. Nevertheless, we also argued that learner strategies coexist with pragmatic awareness and situation-related pragmalinguistic, sociopragmatic, and interactional strategies.

The limitations of this research must be acknowledged. First of all, since it was an exploratory study, the results lacked generalizability. Secondly, the reported cognitive processes were based on the retrospective verbal report. Although the retrospective data were collected immediately after the performance data, the reported strategies were still delayed. Consequently, they may not reflect the strategies the students actually used. Future research may use think-aloud methods to ask participants to verbalize their thoughts while performing a pragmatic task. Such methods can provide rich data when investigating the mental processes underlying complex pragmatic task performances. Third, the performance data of the learners themselves were not discussed in relation to the reported cognitive processes due to limited space. Connecting cognitive processes with pragmatic performances would be necessary for further investigations. Follow-up studies that address these issues will advance the research on cognitive processes.

PEDAGOGICAL IMPLICATIONS

Despite the above limitations, the findings of this study can strengthen pedagogical practices regarding the value of strategy instruction in L2 pragmatic learning. Cohen (2019) pointed out that, although pragmatic rules related to the target language culture have been taught in English as a foreign language (EFL)/English as a second language (ESL) classes, students may still not know “when, why, and how to use them” (p. 141). Unlike other language skills, pragmatic instruction and learner strategies for successful pragmatic performance have not been explicitly included in L2 classrooms (Taguchi and Roever, 2017; Youn, 2018). However, L2 learners can benefit greatly from learning different types of strategy (Cohen, 2019).

The current study provided further empirical evidence to support the notion that successful learners tend to employ a variety of metacognitive and cognitive strategies (see **Table 2**). Additionally, higher-level learners utilized more situation-related pragmalinguistic strategies to accurately and appropriately deliver their messages, which were not commonly employed by lower-level learners. This adds much-needed evidence to support the “widely debated relationship between strategy use and language learning success” (Rose et al., 2018, p. 157). Consequently, when teaching L2 pragmatics, teachers may introduce effective metacognitive strategies to learners to regulate and control their performance. Also, learners can be taught to employ cognitive strategies such as referring to past cultural and linguistic experiences to perform L2 pragmatics appropriately.

Our study also found that pragmatic strategies assisted L2 learners in understanding the expected target sociocultural norms and interactional demands to accomplish pragmatic actions in spoken interaction. For instance, in section Reported Pragmatic Awareness and Strategies of Learners, learners reported pragmatic awareness in their own culture and target-language culture and thus chose appropriate linguistic expressions and placed turns appropriately to achieve shared understanding and maintain the continuity of the interaction. Given that the learners only reported the interaction-related strategies after the performance, it is possible that the learners might not be consciously aware of such strategies until they are engaged in the interaction. In addition, lower-level learners rarely reported interactional strategies, which emphasizes the need for explicit strategy instruction. Accordingly, teachers need to help raise and sharpen the L2 pragmatic awareness of learners. Especially, since learners have the awareness of their first languages and target-language cultures, specific strategy instructions are useful to help learners shift from their own culture to the target culture. Students should also be trained to retrieve appropriate sociopragmatic, pragmalinguistic, and interactional strategies for successful pragmatic performances in various academic settings.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Bachman, L. F., and Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bachman, L. F., and Palmer, A. S. (2010). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bi, N. Z. (2015). *A multi-dimensional examination of strategic competence in the lexico-grammar test performance of Chinese EFL university students*. (Unpublished doctoral dissertation). University of Sydney, Australia.
- Bi, N. Z. (2017). "The impact of strategic processing on lexico-grammatical test performance," in *Advancing the Field of Language Assessment. Studies in Language Testing* (Vol. 42), eds M. A. Christison and N. Saville (Cambridge, UK: UCLES/Cambridge University Press), 229–248.
- Bi, N. Z. (2020). "The effectiveness of construct-relevant and -irrelevant cognitive processes on Chinese EFL learners' test performance," in *Chinese-Speaking Learners of English: Research, Theory, and Practice*, eds R. M. Damerow and K. M. Bailey (New York, NY: Routledge), 38–51.
- Canale, M., and Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Appl. Linguist.* 1, 1–47. doi: 10.1093/applin/1.1.1
- Chen, Y. S. (2015). Chinese learners' cognitive processes in writing email requests to faculty. *System* 52, 51–62. doi: 10.1016/j.system.2015.04.020
- Cohen, A. D. (2005). Strategies for learning and performing L2 speech acts. *Intercult. pragmat.* 2, 275–301. doi: 10.1515/iprg.2005.2.3.275
- Cohen, A. D. (2019). "Strategy instruction for learning and performing target language pragmatics," in *Learning Strategy Instruction in the Language Classroom: Issues and Implementation*, eds A. U. Chamot and V. Harris (Bristol: Multilingual Matters), 140–152.
- Cohen, A. D. (2020). Considerations in assessing pragmatic appropriateness in spoken language. *Lang. Teach.* 53, 183–202. doi: 10.1017/S0261444819000156

ETHICS STATEMENT

The study was reviewed and approved by Human Research Protection Program. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported by the Ministry of Education of China Social Science and Humanities Research Grant No. 17YJC740002, and NAU Faculty Grants Program 2014.

ACKNOWLEDGMENTS

NB would like to acknowledge Dr. Soo Jung Youn for her significant input into the research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.586588/full#supplementary-material>

- Cohen, A. D., and Olshtain, E. (1993). The production of speech acts by EFL learners. *TESOL Q.* 27, 33–56. doi: 10.2307/3586950
- Cohen, A. D., and Sykes, J. M. (2013). "Strategy-based learning of pragmatics for intercultural education," in *Linguistics for Intercultural Education*, eds F. Dervin and A. J. Liddicoat (Amsterdam/Philadelphia, PA: John Benjamins), 87–111.
- Gass, S., and Mackey, A. (2000). *Stimulated Recall Methodology in Second Language Research*. Mahwah, NJ: Erlbaum.
- Golato, A. (2003). Studying compliment responses: a comparison of DCTs and recordings of naturally occurring talk. *Appl. Linguist.* 24, 90–121. doi: 10.1093/applin/24.1.90
- Hassall, T. (2008). "Pragmatic performance: what are learners thinking?" in *Investigating Pragmatics in Foreign Language Learning, Teaching and Testing*, eds E. A. Soler and A. Martinez-Flor (Bristol: Multilingual Matters), 72–93.
- Ishihara, N., and Cohen, A. D. (2010). *Teaching and Learning Pragmatics: Where Language and Culture Meet*. Abingdon: Routledge.
- Kasper, G. (2006). "Speech acts in interaction: towards discursive pragmatics," in *Pragmatics and Language Learning. Vol. 11*, eds K. Bardovi-Harlig, C. Félix-Brasdefer, and A. S. Omar (Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai'i), 281–314.
- Kasper, G. (2008). "Data collection in pragmatics research," in *Culturally Speaking: Culture, Communication, and Politeness Theory*, ed H. Spencer-Oatey (New York, NY: Continuum), 279–302.
- Leech, G. (1983). *Principles of Pragmatics*. London: Longman.
- Li, C., and Gao, X. (2017). Bridging 'what I said' and 'why I said it': the role of metapragmatic awareness in L2 request performance. *Lang. Awar.* 26, 170–190. doi: 10.1080/09658416.2017.1387135
- Oxford, R. L. (2017). *Teaching and Researching Language Learning Strategies: Self-Regulation in Context*. Abingdon: Routledge.
- Phakiti, A. (2007). *Strategic Competence and EFL Reading Test Performance: A Structural Equation Modelling Approach*. Frankfurt: Peter Lang.

- Ren, W. (2013). The effect of the study abroad on the pragmatic development in internal modification of refusals. *Pragmatics* 23, 715–741. doi: 10.1075/prag.23.4.06ren
- Ren, W. (2014). A longitudinal investigation into L2 learners' cognitive processes during study abroad. *Appl. Linguist.* 35, 575–594. doi: 10.1093/applin/amt019
- Robinson, M. A. (1992). "Introspective methodology in interlanguage pragmatics research," in *Pragmatics of Japanese as Native and Target Language*, ed G. Kasper (Honolulu: HI: University of Hawaii Press), 27–81.
- Rose, H. L., Briggs, J. G., Boggs, J. A., Sergio, L., and Ivanova-Slavianskaia, N. (2018). A systematic review of language learner strategy research in the face of self-regulation. *System* 72, 151–163. doi: 10.1016/j.system.2017.12.002
- Taguchi, N., and Kim, Y. (eds.). (2018). *Task-Based Approaches to Teaching and Assessing Pragmatics*. Amsterdam/Philadelphia, PA: John Benjamins.
- Taguchi, N., and Roever, C. (2017). *Second Language Pragmatics*. Oxford: Oxford University.
- Tseng, W., Dörnyei, Z., and Schmidt, R. (2006). A new approach to assessing strategic leaning: the case of self-regulation in vocabulary acquisition. *Appl. Linguist.* 23, 1–2. doi: 10.1093/applin/ami046
- Woodfield, H. (2010). What lies beneath? verbal report in interlanguage requests in English. *Multilingua* 29, 1–27. doi: 10.1515/mult.2010.001
- Woodfield, H. (2012). "Pragmatic variation in learner perception: the role of retrospective verbal report in L2 speech act research," in *Pragmatic Variation in First and Second Language Contexts: Methodological Issues*, eds C. Felix-Brasdefer and D. A. Koike (Amsterdam/Philadelphia, PA: John Benjamins), 209–223.
- Woodrow, L. (2005). The challenge of measuring language learning strategies. *Foreign Lang. Ann.* 38, 90–98. doi: 10.1111/j.1944-9720.2005.tb02456.x
- Youn, S., and Bi, N. Z. (2019). Investigating test-takers' strategy use in task-based L2 pragmatic speaking assessment. *Intercult. Pragmat.* 16, 185–218. doi: 10.1515/ip-2019-0009
- Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Lang. Test.* 32, 199–225.
- Youn, S. J. (2018). Task-based needs analysis of L2 pragmatics in an EAP context. *J. English Acad. Purp.* 36, 86–98. doi: 10.1016/j.jeap.2018.10.005

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Bi. 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: frontiersin.org/about/contact



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