frontiers | Research Topics

Perspective taking in language

Edited by
Carla Contemori, Petra Hendriks, Franziska Köder, Emar Maier, Jorrig Vogels and Sonja Zeman

Published in
Frontiers in Communication
Frontiers in Psychology



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 978-2-8325-2204-2
DOI 10.3389/978-2-8325-2204-2

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

Perspective taking in language

Topic editors

Carla Contemori — The University of Texas at El Paso, United States
Petra Hendriks — University of Groningen, Netherlands
Franziska Köder — University of Oslo, Norway
Emar Maier — University of Groningen, Netherlands
Jorrig Vogels — University of Groningen, Netherlands
Sonja Zeman — University of Augsburg, Germany

Citation

Contemori, C., Hendriks, P., Köder, F., Maier, E., Vogels, J., Zeman, S., eds. (2023). *Perspective taking in language*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-2204-2

Table of contents

05 Editorial: Perspective taking in language
Jorrig Vogels, Sonja Zeman, Carla Contemori, Petra Hendriks, Franziska Köder and Emar Maier

13 Processing Speaker-Specific Information in Two Stages During the Interpretation of Referential Precedents
Edmundo Kronmüller and Ernesto Guerra

30 What's New to You? Preschoolers' Partner-Specific Online Processing of Disfluency
Si On Yoon, Kyong-sun Jin, Sarah Brown-Schmidt and Cynthia L. Fisher

47 Free Indirect Discourse in Non-fiction
Andreas Stokke

59 Reading Fictional Narratives to Improve Social and Moral Cognition: The Influence of Narrative Perspective, Transportation, and Identification
Lena Wimmer, Stacie Friend, Gregory Currie and Heather J. Ferguson

78 Storytelling on Oral Grounds: Viewpoint Alignment and Perspective Taking in Narrative Discourse
Kobie van Krieken and José Sanders

92 Extended Perspective Shift and Discourse Economy in Language Processing
Jesse A. Harris

111 Mental State Verb Production as a Measure of Perspective Taking in Narrations of Individuals With Down Syndrome
Isabel Neitzel and Martina Penke

124 Egophoricity and Perspective: A View From Spoken Swedish
Henrik Bergqvist

137 The Benefits and Obstacles to Perspective Getting
Debby Damen, Monique M. H. Pollmann and Teri-Louise Grassow

152 Irony and Perspective-Taking in Children: The Roles of Norm Violations and Tone of Voice
Franziska Köder and Ingrid Lossius Falkum

165 Children's Pronoun Interpretation Problems Are Related to Theory of Mind and Inhibition, But Not Working Memory
Sanne J. M. Kuijper, Catharina A. Hartman and Petra Hendriks

181 On the Interaction of Gestural and Linguistic Perspective Taking
Stefan Hinterwimmer, Umesh Patil and Cornelia Ebert

196 Fictive Deixis, Direct Discourse, and Viewpoint Networks
Barbara Dancygier

212 Consequences of Sensory Modality for Perspective-Taking: Comparing Visual, Olfactory and Gustatory Perception
Elsi Kaiser

232 Epistemic Perspectives and Communicative Acts
Anton Benz

252 In a Manner of Speaking: How Reported Speech May Have Shaped Grammar
Stef Spronck and Daniela Casartelli

274 Reading Minds, Reading Stories: Social-Cognitive Abilities Affect the Linguistic Processing of Narrative Viewpoint
Lynn S. Eekhof, Kobie van Krieken, José Sanders and Roel M. Willems

frontiers | Frontiers in Communication

TYPE Editorial
PUBLISHED 04 April 2023
DOI 10.3389/fcomm.2023.1160727

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Si On Yoon,
The University of Iowa, United States

*CORRESPONDENCE
Jorrig Vogels
✉ jorrigvogels@gmail.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

RECEIVED 07 February 2023
ACCEPTED 20 March 2023
PUBLISHED 04 April 2023

CITATION
Vogels J, Zeman S, Contemori C, Hendriks P, Köder F and Maier E (2023) Editorial: Perspective taking in language. *Front. Commun.* 8:1160727. doi: 10.3389/fcomm.2023.1160727

COPYRIGHT
© 2023 Vogels, Zeman, Contemori, Hendriks, Köder and Maier. 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.

Editorial: Perspective taking in language

Jorrig Vogels[1*†], Sonja Zeman[2†], Carla Contemori[3], Petra Hendriks[1], Franziska Köder[4] and Emar Maier[1,5]

[1]Center for Language and Cognition Groningen, University of Groningen, Groningen, Netherlands, [2]Chair of German Linguistics, University of Augsburg, Augsburg, Germany, [3]Languages and Linguistics, University of Texas at El Paso, El Paso, TX, United States, [4]Center for Multilingualism in Society across the Lifespan, University of Oslo, Oslo, Norway, [5]Faculty of Philosophy, University of Groningen, Groningen, Netherlands

KEYWORDS

perspective taking, language, cognition, theory of mind, audience design, narrative, viewpoint alignment, empathy

Editorial on the Research Topic
Perspective taking in language

1. Introduction

Each language user brings their own unique set of perceptions, knowledge and experiences to the table, which may or may not be aligned with that of other people. In communication with others, or when talking or writing about a third person, these perspectives need to be somehow coordinated. For example, when referring to an object in the world, speakers take into account whether the object and its visual context are shared between speaker and hearer in choosing a particular referring expression, such as "the big duck" in the context of a bigger and a smaller duck (Heller et al., 2008). Using their own knowledge of the situation, hearers can also understand ironic utterances such as "Great weather!" when it is pouring, and quickly draw the inference that the speaker intends to convey something considerably different from the literal meaning. Furthermore, languages offer various constructions for reporting the perspectives of others, including quotation (e.g., "Laura said: 'The weather is great!'"), attitude reports (e.g., "Laura believes that it's raining"), and more subtle stylistic means such as free indirect discourse (e.g., "Yay! She would go jump in all the puddles right now!").

In producing or interpreting these various linguistic forms, language users (speakers, hearers, writers, and readers) need to consider a perspective that is different from their own, for instance by inferring what someone else knows, believes, or feels. Research has shown that the ability to infer and reason about other people's mental states starts to develop in early childhood (early forms of perspective taking have been observed for infants at around 13–15 months of age; Onishi and Baillargeon, 2005). At the same time, even adults sometimes fail to use this ability to take into account the perspective of others. The exact preconditions for when and to what degree language users shift perspective are as yet unknown, although they should probably be sought in a combination of linguistic, cognitive, and social factors.

As already becomes clear from the examples given above, perspective taking plays a role in a variety of functions of language, from irony understanding to narrative writing. There is a rich literature surrounding each of these different types of perspective taking in language. For example, in psycholinguistic studies of conversation, perspective taking may be defined as "the ability to appropriately attend to information that is either shared, or not

shared with one's partner, depending on the context" (Brown-Schmidt and Hanna, 2011, p. 16). Studies of perspective taking in narratives, on the other hand, are concerned with how readers or hearers establish the different viewpoints of the narrator and the characters within the narrative, and take into account what they know, think, or feel (e.g., Sanders, 1994; Salem et al., 2017). Research on narrative perspective taking has also been extended to the domain of visual communication, studying phenomena like role shift in sign languages (e.g., Poulin and Miller, 1995; Lillo-Martin, 2012) and point of view shots in comics and film (e.g., Maier and Steinbach, 2022). In past years, there has been increased attention to perspective taking in both children and adults with Autism Spectrum Disorder (Overweg et al., 2018; Abbot-Smith et al., 2020; Kissine, 2021; Zimmermann et al., 2021), for whom pragmatic impairment is a core deficit, as well as in other clinical populations (e.g., ADHD: Kuijper et al., 2015; Alzheimer's Dementia: Bittner et al., 2022; schizophrenia: Van Schuppen et al., 2019). However, the mutual relationships between the different types of perspective taking are still largely unknown.

In this editorial, we explore the question whether all these different types of perspective taking are manifestations of the same underlying concept, or whether they should be seen as linguistically and/or cognitively different notions. For this, we consider the different perspectives on perspective taking in language presented in the articles in the current Research Topic and the previous literature. However, before we can review the notion of perspective taking, we first need to define what a perspective is.

2. What is a perspective?

"Perspective" (lat. *perspicere* "looking through") is a notion originating in the study of visual perception. In the visual domain, perspective can be defined as a directed relation between a perceiving subject and the perceived aspects of an object in focus, the latter being dependent on the observer's viewpoint: If the observer's or object's position is changed, the perspective changes. In recent years, "perspective" has become a frequent notion in both the linguistic and the more general-cognitive (henceforth "cognitive") literature, in which it tends to refer to various phenomena. This is also reflected in our Research Topic, whose topics range from lexical items (Eekhof et al.), pronouns (Bergqvist; Kuijper et al.), epistemic, evidential, and causal expressions (Bergqvist), and mental state verbs (Neitzel and Penke), to comprehension of irony (Köder and Falkum), communicative acts in conversation (Benz; Damen et al.; Kronmüller and Guerra; Yoon et al.), co-speech gestures (Hinterwimmer et al.), subjective adjectives (Kaiser), represented speech (Dancygier; Spronck and Casartelli; Stokke), and narratives (Harris; Van Krieken and Sanders; Wimmer et al.). The notion of perspective itself has, however, remained rather vague (as has also been stated in the overviews in e.g., Klein and von Stutterheim, 2002; Verhagen, 2007; Zeman, 2017), and the different "perspectival" phenomena seem to be more linked by family resemblance than by reference to a single well-defined concept (Linell, 2002, p. 53).

When applied to language and cognition, "perspective" is a metaphorical concept. Metaphorical mappings are characterized by the fact that some, but not all properties of the source concept are mapped onto the target concept (Lakoff and Johnson, 1980). In order to gain a better understanding of perspectivization as a linguistic notion and to look out for an overarching and more formalized understanding of what it means to "take" a perspective in all these different linguistic domains, we should therefore look for those structural properties of the source concept that can be found within the different studies. In other words, we have to ask which components of perceptual perspective are seen as relevant for linguistic perspectivization, and which properties are not mapped onto the target concept.

Reviewing the literature on cognitive and linguistic perspective, three structural properties of visual perspectivization in particular can be identified that are mapped onto the different "perspectival" phenomena (see Zeman, 2020). First, perspective implies the existence of a set of possible alternatives that allows us to "take" and "shift" a perspective. In real life, an observer is, due to their spatiotemporal position, commonly restricted to one visual viewpoint only. If one stands in front of a statue, one is able to see only its front but not its backside. A choice of perspective is thus linked to the selection of one perspective out of others. Second, perspective is commonly seen as person-bound. In its original sense, "perspective" presupposes an animate subject that is able to perceive and observe. Inanimate entities thus cannot have a perspective, unless they are anthropomorphized as conscious rational agents (like Marjory, the talking Trash Heap in *Fraggle Rock*). This is linked to a third aspect: the perspectival relation is directed at some perceptible object and the result of a cognitive process, prototypically an act of visual perception.

The general idea of these three components of perspective appears in many studies on cognitive or linguistic perspectivization. There are, however, also crucial differences between visual vs. cognitive and linguistic perspectivization, linked to the fact that in language and cognition, viewpoints are not physically located in the real, but in a mental world. As soon as we expand the concept of "perspective" in this way, the structural properties of visual perspective are understood in a more abstract sense when it comes to cognitive and linguistic perspective (see Zeman, 2017, 2020).

(i) Alternative perspectives. The primary prerequisite of cognitive perspectivization is not only the existence of alternative perspectives, but the awareness that such alternatives are available. The requirement for taking another's perspective is thus the cognitive ability to mentally decouple or suppress one's own point of view. Such mental "switches" imply an inherent hierarchy between the different viewpoints, since the original viewpoint is not necessarily canceled but can be maintained (i.e., in hypothetical scenarios like "if I were a millionaire," or when watching a movie, we do not forget our original viewpoint in real life). We are thus accustomed to holding more than one perspective at a time. When applied to language, "perspectivization" is also commonly more than taking one perspective out of a set of mutually exclusive alternatives (Zeman, 2017). Linguistic multiperspectivization has been shown, for example, for phenomena of communicative interaction like irony

(Köder and Falkum) and speech and thought representation (Dancygier; Spronck and Casartelli), but also for grammatical elements like egophoric pronouns and modal particles (Bergqvist). The interaction between multiple viewpoints in discourse has been modeled in terms of viewpoint networks by a.o., Dancygier and Vandelanotte, 2016; Van Duijn and Verhagen, 2019; Dancygier.

(ii) Person-boundedness. Perspectivization in language and cognition is often seen as "person-bound," that is, as the "introduction of a subjective point of view that restricts the validity of the presented information to a particular subject (person) in the discourse" (Sanders and Redeker, 1996, p. 293). This is most obvious in perspective taking in communicative acts (Benz; Damen et al.; Yoon et al.) and irony understanding (Köder and Falkum). In narratives, as well, perspectives are commonly ascribed to thinking and perceiving characters within a story (Van Krieken et al., 2017). Other linguistic phenomena that are prototypically described as "perspectival" are egophoric pronouns (Bergqvist) and evaluating adjectives (Kaiser), which establish a relation between the "speaker" and the denoted situation and thus presuppose a first-person agent's point of view. Perspectivization has therefore also been described as covering the "non-objective facts of language" (Verhagen, 2007). However, under such a definition, the "perspective holder" does not have to be a real person of flesh and blood. The perspectival center can also be a hypothetical or a fictional character (Stokke). Fictional narratives, for example, offer many examples of "unnatural" viewpoints of (anthropomorphized) animals and objects (Richardson, 2006; Trompenaars et al., 2018), and the status of the narrator as a "person" has been an issue of controversial debate (for an overview see Zeman, 2020). Also, viewpoints can be quite abstract, as seen in studies on perspectivization in grammar (Bergqvist; Spronck and Casartelli). In language, the concept of person-boundedness thus seems to be a scalar property that ranges from actual human beings in the real world over anthropomorphized characters to more abstract conceptions of perspectival centers.

(iii) Object in focus of the cognitive process. Linked to the observation in (ii) that the origin of perspective is not necessarily an animate person but can also be a rather abstract functional instance within the text, cognitive and linguistic perspectivization do not necessarily have to be the result of a perceptual process. Rather, the notion of perspective is extended to all kinds of mental states, such as knowledge, desires, beliefs, or emotions [for an overview see e.g., Taylor and Edwards (2021) who distinguish between visual perspective taking (VPT) vs. mental state attribution, typically referred to as theory of mind (ToM)]. As a result, cognitive and linguistic perspectivization are also not necessarily directed at a perceptible object that one is taking a perspective on. In cases of, for instance, temporal perspective, we can even ask whether perspectivization has to be directed at an object at all.

All three structural properties of perspective are shared by both the source (visual perspective) and the target (cognitive and linguistic perspective) of the perspective metaphor. As such, they appear in various approaches on linguistic perspectivization and also in several papers of our Research Topic. We therefore assume that all three structural properties are considered representative in many accounts of perspective and perspectivization in language. In the next section, we analyze what it means to take on a perspective, both cognitively and linguistically, when two or more viewpoints are considered simultaneously. That is, going beyond the mere selection of a particular perspective from a set of alternatives, which we call *perspective holding*, we focus on situations in which multiple perspectives are considered at the same time (see also Zeman, 2017). For ease of reference, we continue to refer to this more complex type of perspectivization as *perspective taking*. We take the definitions of perspective taking in the papers of the Research Topic as a starting point. However, given the emerging consensus on the concept of "perspective," we argue that our analysis in the following section determines more than just the common denominator of the articles in our volume. Rather, it seems to capture crucial structural aspects of *perspective taking* when applied to language and cognition.

3. A formalized definition of perspective taking

To systematically investigate the notion of perspective taking, we collected the various definitions of perspective taking that we found throughout the articles in this Research Topic, and broke them down into their parts. Next, we determined which parts were shared across multiple definitions. In this way, we arrived at the formulaic description in Equation (1).

$$\textit{Perspective taking} \;=\; \text{a } C \text{ of } M\text{-ing the } I \text{ of } P \qquad (1)$$

Where C is some kind of cognitive process, M is some kind of mental operation, I is some kind of information, and P is a person. In other words, the authors in this Research Topic seem to agree that, at a minimum, perspective taking should be considered a cognitive process, which involves mentally operating on some information that belongs to a person. Most definitions in this Research Topic can be written as a variant of this formula, although not all definitions include all elements. In the following, we discuss each of the elements in the formula in Equation (1) in more depth, determining where researchers diverge in the exact nature of the element. The most prevalent element of perspective taking in the definitions seems to be *P*: virtually all papers in this volume involve taking the perspective of a person. We will therefore start our discussion with this element.

3.1. *P*(erson)

Although we have seen above that the source of a perspective can be rather abstract, the types of perspective taking discussed in this Research Topic generally involve at least one other person or rational agent (aside from the one doing the perspective taking). Two separate ways in which the relationship between the original perspective of the perspective taker and the newly taken perspective

takes form in language can be clearly distinguished: (i) some linguistic choices depend on reasoning about the perspective of other speech act participants, and (ii) language can help us try on other perspectives outside of the current communicative exchange; that is, beyond the here and now. We will call the first type of perspective taking *communicative perspective reasoning* and the second type *perspective shifting*.

3.1.1. Communicative perspective reasoning

Communicative perspective reasoning applies when speakers take their hearer's knowledge or perception into account to choose the right expression (e.g., "here" vs. "there," "left" vs. "right," "the car" vs. "the big car," "the dog" vs. "a dog," etc.). Conversely, hearers also take the speaker's knowledge or perception into account to choose the right interpretation. For example, as Kronmüller and Guerra show, hearers make use of specific information about the speaker in their interpretation of referentially ambiguous terms (e.g., when the word *bat* can refer to either a baseball bat or a flying mammal). In addition, Bergqvist investigates how speakers make use of the egophoric pronouns "I," "you," and "we," modal particles, and mental verbs such as "think" to explicitly signal their own perspective as well as to take into account the perspective of their speech partner. More implicitly, the use of irony and other pragmatic implicatures also require making inferences about what the other speech act participant knows. Note that we do not commit ourselves to the position that such inferences in communicative perspective taking must be made consciously; they may also occur unconsciously. That such reasoning is acquired relatively early becomes clear in the work of Köder and Falkum, who show that 3-year-old children are already sensitive to certain features of irony. Similarly, Yoon et al. find that young children are able to infer from disfluent expressions that their speech partner must be referring to something that is unfamiliar from the speech partner's perspective.

Crucially, the above examples are all about the perspectives of the participants in a specific communicative exchange between an actual speaker and an actual hearer. However, an important question is whether communicative perspective reasoning has to involve a specific speech partner, or whether the same process is also applied in more generic perspective-taking situations (see e.g., Dell and Brown, 1991). According to Kuijper et al., taking the perspective of a (hypothetical) speech partner is also a necessary step in certain aspects of grammar, such as the correct interpretation of object pronouns as non-reflexive, and is thus independent of a specific communicative situation.

3.1.2. Perspective shifting

Perspective shifting applies in contexts beyond the speaker-hearer relationship in the here and now. Here, a language user takes a "third party" perspective, such as that of a fictional character, into account. This type of perspective taking can be cued with a grammatical construction that semantically forces the shift, but it can also be a more global pragmatic inference process. A common reporting construction that forces a perspective shift is direct discourse. As Van Duijn and Verhagen (2019, p. 213) note, direct discourse "suggests a viewpoint shift in its fullest form" (e.g., "'I am tired', said Alice"), whereas indirect discourse (e.g., "Alice said that she was tired") presents the event from a third person's point of view. Another perspective-shifting construction, often found in literary texts, is free indirect discourse, which mixes elements of direct and indirect discourse. More specifically, both tenses and pronouns are taken from the narrator's perspective (as in indirect discourse), whereas everything else comes from the character's perspective (as in direct discourse; e.g., "Ellen made a decision. Yes! She would tell him later today"; Stokke, p. 3). In his article, Stokke shows that this construction also occurs in non-fictional texts, and there serves a special function of taking the (fictionalized) perspective of an actual (historical) person, which may have cognitive benefits in taking in historical facts.

Thus, both direct discourse and free indirect discourse shift the interpretation of indexicals (e.g., "then" > "now") and other perspective-sensitive expressions away from the current communicative situation, whereas indirect discourse constructions do not affect indexicals but do provide information about someone's mental state and hence their perspective in a broader sense. The perspectival source can also remain abstract, however. Spronck and Casartelli point out that speech or attitude report constructions may be extended to include other constructions that have a "say," "think," "want," or even "cause" meaning but do not literally report someone's speech or thoughts. For example, in the West-African language Wan, a construction that would literally translate as "The water said: let me boil!" can be used to express the non-reporting meaning "The water was about to boil" (Spronck and Casartelli, p. 2). Spronck and Casartelli's typological inventory suggests that perspective shifting may be much more pervasive in grammar than is apparent at first glance.

Beyond speech and attitude report constructions, various linguistic elements may "provide access to the inner world of characters" (Eekhof et al., p. 1). These may be lexical elements such as "happy," but also elements of a more general narrative style, such as the voice in which the narrative is told (first vs. third-person narrator) or whether the narrator has access to the character's thoughts and feelings (internal vs. external focalization; Wimmer et al.).

Perspective shifts may even be established in the absence of explicit markers, in a more pragmatic fashion. Harris notes that "speakers and hearers are finely attuned to perspectives and viewpoints that are not their own, even though perspectival information is not encoded directly in the morphosyntax of languages like English" (p. 1). When we read or listen to a story, we can get transported away from the here and now, to the story world (or, in the case of non-fiction, a different time and place in the actual world), identifying or empathizing with a protagonist and hence considering or taking on their perspective in some sense. Harris explores the contextual cues that can invoke or retain a certain perspective during language processing in the absence of explicit reporting constructions. Kaiser undertakes a similar investigation, focusing on the processing of subjective adjectives. That contextual cues for perspective shifting are not limited to the spoken or written modality is shown by Hinterwimmer et al., whose experimental results suggest that perspective information can also be independently encoded in co-speech gestures.

As noted earlier, another aspect of perspective taking that is less explored in this Research Topic but is worth mentioning here is that language users can also take *their own* perspective in a situation other than the here and now. For example, speakers can talk about a past event (temporal displacement) or a hypothetical or imaginary situation ("if I were a millionaire"), representing their own thoughts or feelings in that situation (cf., e.g., Overweg et al., 2018; Zeman, 2020). These examples show that the presence of another person or mind is actually not a necessary condition for perspective taking (but perhaps only a different mind*set*).

3.1.3. Unifying perspective taking types

Several authors in this Research Topic attempt to unify the different types of perspective taking outlined above in a single model. Dancygier connects the grammatically induced perspective shift in direct discourse constructions to the pragmatic concept of deixis. Similarly, Van Krieken and Sanders make a connection between narrative perspective shifting on the one hand and communicative perspective reasoning on the other. In their framework of narrative perspective taking, they propose that both written and oral narratives revolve around the dynamic alignment of different viewpoints: those of the speaker/narrator, of the hearer/reader, and of the narrative characters. A similar model is put forward by Van Duijn and Verhagen (2019), who propose that perspective taking can take place along several axes: speakers/narrators coordinate with their addressees about some shared object of conceptualization, and together narrator and addressee may also be coordinating with a third party, not included in their communicative situation. This third party itself can also have a stance toward the object of conceptualization and further parties, thus allowing for a recursive series of perspective taking.

3.2. *I*(nformation)

According to the definition in (1), perspective taking not only requires another person or rational agent (*P*), but also certain information (*I*) that this person or agent has access to. Perspective taking seldom involves a complete identification with, or transportation into, the mind of this other person. In most cases, the perspective is restricted to one type of information. Based on the articles in this volume, we may distinguish at least three types of information that can serve as a basis for perspective taking: (i) factual knowledge; (ii) subjective attitudes; and (iii) goals and intentions.

3.2.1. Factual knowledge

Language use has been characterized as a joint activity, in which speaker and hearer work together to get to the intended meaning of a speech act (Clark, 1996). In this view, speaker and hearer take into account the knowledge that they both share, the common ground (e.g., shared knowledge about the world, about the current situation, or about each other), to work out the optimal form or meaning, respectively. Building on this background, both Benz and Bergqvist provide a theoretical analysis of how such epistemic perspective taking by speech-act partners takes form in language. Using an experimental paradigm, Kronmüller and Guerra provide evidence about when and how this type of perspective taking takes place during real-time language processing in adults. Furthermore, Yoon et al. provide evidence that children as young as 4 years old can already distinguish between different partners' knowledge states, and use this information in language comprehension.

3.2.2. Subjective attitudes

Besides assessing what factual knowledge another person has, perspective taking may also involve ascertaining another person's more subjective attitudes. People might not be inclined to do this overtly: Damen et al. asked participants to estimate a conversation partner's emotion toward or opinion about various matters, and found that they were unlikely to ask the other person, even when this was explicitly presented as a recommended option. However, perspective taking may be a necessity when it comes to comprehending subjectively colored linguistic expressions, such as "tasty." After all, as Kaiser shows, one needs to know *for whom* something is tasty to fully grasp the meaning of the adjective. Following the results of Kaiser's study on the modality-specificity of such attributions, it appears that perspective taking has a top-down, context-sensitive impact on semantics.

3.2.3. Goals and intentions

Rather than merely estimating what another person knows, believes or feels, perspective taking often has the additional aim of understanding the other person's intentions. In conversation, for example, the hearer needs to work out what the speaker intended to say from what was actually said. This fact forms the basis of the field of Gricean pragmatics (e.g., Grice, 1975; Horn, 1984; Levinson, 2000). That the speaker's intentions may not always coincide with what was literally said is especially clear in the case of irony, where speakers may occasionally even say the opposite of what they intend to say. The cognitive underpinnings of working out the speaker's intention are explored by Köder and Falkum and Kuijper et al.

Goals and intentions may also be important for narrative comprehension: to be able to transport yourself into the story world and identify yourself with the narrative characters, it helps to understand what moves and motivates these characters. However, based on the results of their experiments, Wimmer et al. place doubts on a central role for perspective taking in identification and transportation in narratives.

3.3. *M*(ental operation)

Having established different types of information that may be associated with someone else's perspective, we now enter the more cognitive side of perspective taking, and ask what kind of mental operation (*M*) must be executed on the perspectival information to be able to call it perspective *taking*. At the minimum, the perspectival information needs to be represented somewhere in the mind of the speaker, hearer, or reader. For example, the first step in the acquisition of a theory of mind (ToM) is that the child needs to learn to represent other people's beliefs as distinct from their own beliefs (e.g., Perner, 1991; see also Yoon et al.). However,

there is discussion about whether representing another person's beliefs (knowledge, attitudes) is sufficient for perspective taking (see Deschrijver and Palmer, 2020 for an alternative proposal in terms of relations rather than representations). In their model of narrative perspective taking, Van Krieken and Sanders (see also Dancygier) propose that taking someone's perspective involves a *mental alignment* of viewpoints. This suggests that in addition to representing another person's perspective, perspectives also need to be aligned; that is, one's own perspective needs to be adjusted to match the other. If this theory is true, it raises the question how far this alignment needs to go: to what extent do you really need to step into someone else's shoes to take their perspective? Does the original perspective get lost in the process, or is it still available? For example, does perspective taking require that you empathize with another person (recognize their mindset or feelings; Van Krieken et al., 2017; Neitzel and Penke), or do you also need to identify with them (*adopt* their mindset or feelings; Van Krieken et al., 2017; Wimmer et al.)?

3.4. *C*(ognitive process)

While some authors in this Research Topic treat perspective taking purely as a linguistic property, triggered by certain linguistic elements, and part of the semantic denotation or pragmatic understanding of a sentence, ultimately perspective taking needs to take place in the mind of the language user, and hence be a cognitive process (*C*). It is as yet unclear what kind of cognitive process perspective taking in language entails, and it has been framed, for instance, as an important part of social cognition (Eekhof et al.) as well as a fully grammaticalized process in specific cases (Kuijper et al.). Over the past couple of decades, there has been a fierce debate over the question whether perspective taking in conversation should be considered an early automatic process or a late high-level reasoning or monitoring process (see, e.g., Keysar et al., 2000; Hanna et al., 2003; Brown-Schmidt and Hanna, 2011). Kronmüller and Guerra unite these two views by showing that it could be both: they argue that perspective taking involves an automatic cue-driven memory process as well as a higher-level inferential mechanism. Similarly, Kuijper et al. suggest that perspective taking can be an automatized grammatical process, but also an effortful pragmatic process, depending on the situational variability. An important question is whether this division can also be applied to other types of perspective taking, such as narrative perspective shifts.

4. Discussion and outlook

In this Editorial, we started out with the question whether the different types of perspective taking are fundamentally different processes, or whether they have a common base. We subsequently inventoried some definitions of perspective taking in language, and came to an overarching characterization, repeated as Equation (2), consisting of four critical elements: a person (or rational agent; *P*), information associated with that person (*I*), and a mental operation (*M*), which is part of a more general cognitive process (*C*).

$$\textit{Perspective taking} \;=\; \text{a } C \text{ of } M\text{-ing the } I \text{ of } P. \qquad (2)$$

This characterization may help us to break down boundaries between disciplines focusing on a specific type of perspective taking, from perspective taking in conversation to narrative perspective shifting, and from visual to cognitive and emotional perspective taking. In turn, this crossing of boundaries may shed light on the question whether and how the various types of perspective taking are related. Future research should clarify the exact nature of the four elements outlined above, thereby preferably bridging the different types of perspective taking. Whether this will result in a common base of perspective taking remains to be seen, but at the very least, the plurality of uses and definitions of the term "perspective" available in linguistics, literary studies, cognitive science, and psychology shows that there is a real need for researchers to be more consistent in how they use the term.

We would like to open the discussion by positing a number of questions arising from our inventory that require further investigation:

(i) How specific does the person or rational agent whose perspective is taken need to be? Does it have to be an actual person (addressee, narrative character), or can it be more abstract (hypothetical, grammaticalized)?
(ii) In taking someone's perspective, is it sufficient to merely represent the other person's beliefs, knowledge, or attitudes? Or do you also need to identify with someone to be able to take their perspective? If so, what happens to the original perspective?
(iii) What are the cognitive mechanisms (e.g., executive function, theory of mind) underlying different types of perspective taking in language? For example, to what degree do different types of perspective taking involve automatic cue-driven memory processes and to what degree are they higher-level inferential mechanisms?
(iv) How can we find suitable ways to operationalize perspective taking in experimental research, for instance via measures of visual attention (eye gaze), cognitive effort (e.g., pupil dilation, reaction time, processing speed) or behavioral responses?
(v) To what extent can perspective taking be considered an integral part of semantics or pragmatics, and to what extent should we consider it as a more general socio-cognitive ability that influences language production and comprehension, but is not part of language itself?

To be able to get a firm grasp of the notion of perspective taking, future research on perspective taking should try to find an answer to these questions. To aid in this endeavor we hope here to have clarified the relevant theoretical distinctions and terminology.

Author contributions

JV and SZ contributed equally to the writing of this editorial. CC, PH, FK, and EM provided detailed feedback to a first draft of the text and are listed in alphabetical order. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Netherlands Organisation for Scientific Research (NWO), under Grant 275-89-0360 to JV, and 276-80-004 (FICTION) to EM.

Acknowledgments

We thank all authors of the articles in this Research Topic for their contributions. We are also grateful to all participants of the workshop on Perspective Taking in Language (PeTaL), held on June 23–24, 2021 at the CLCG, University of Groningen, where most of the articles of this Research Topic have been presented and discussed.

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.

References

Abbot-Smith, K., Williams, D., and Matthews, D. (2020). 'What's new for you?': interlocutor-specific perspective-taking and language interpretation in autistic and neuro-typical children. *Res. Autism Spectr. Disord.* 70, 101465. doi: 10.1016/j.rasd.2019.101465

Bittner, D., Frankenberg, C., and Schröder, J. (2022). Changes in pronoun use a decade before clinical diagnosis of Alzheimer's dementia—linguistic contexts suggest problems in perspective-taking. *Brain Sci.* 12, 121. doi: 10.3390/brainsci12010121

Brown-Schmidt, S., and Hanna, J. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue Discourse* 2, 11–33. doi: 10.5087/dad.2011.102

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620539

Dancygier, B., and Vandelanotte, L. (2016). "Discourse viewpoint as network," in *Viewpoint and the Fabric of Meaning: Form and Use of Viewpoint Tools Across Languages and Modalities*, eds B. Dancygier, W. Lu, and A. Verhagen (Berlin: de Gruyter), 13–40. doi: 10.1515/9783110365467-003

Dell, G. S., and Brown, P. M. (1991). "Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener"," in *Bridges Between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman*, eds D. Napoli, and J. A. Kegl (Hillsdale, NJ: Lawrence Erlbaum), 105–129.

Deschrijver, E., and Palmer, C. (2020). Reframing social cognition: relational versus representational mentalizing. *Psychol. Bull.* 146, 941–969. doi: 10.1037/bul0000302

Grice, H. P. (1975). "Logic and conversation," in *Syntax and Semantics. Vol III: Speech Acts*, eds P. Cole, and J. Morgan (Cambridge, MA: Academic Press), 41–58. doi: 10.1163/9789004368811_003

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

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

Horn, L. (1984). "Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature," in *Meaning, Form, and Use in Context: Linguistic Applications*, ed D. Schiffrin (Washington, DC: Georgetown University Press), 11–42.

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

Kissine, M. (2021). Autism, constructionism and nativism. *Language* 97, e139–e169. doi: 10.1353/lan.2021.0055

Klein, W., and von Stutterheim, C. (2002). "Quaestio and L-perspectivation," in *Perspective and Perspectivation in Discourse*, eds C. F. Graumann, and W. Kallmeyer (Amsterdam: Benjamins), 59–88. doi: 10.1075/hcp.9.06stu

Kuijper, S. J., Hartman, C. A., and Hendriks, P. (2015). Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns. *PLoS ONE* 10, e0132408. doi: 10.1371/journal.pone.0132408

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT press. doi: 10.7551/mitpress/5526.001.0001

Lillo-Martin, D. (2012). "Utterance reports and constructed action," in *Sign Language. An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: de Gruyter), 365–387. doi: 10.1515/9783110261325.365

Linell, P. (2002). "Perspectives, implicitness and recontextualization," in *Perspective and Perspectivation in Discourse*, eds C. F. Graumann, and W. Kallmeyer (Amsterdam: Benjamins), 41–57. doi: 10.1075/hcp.9.05lin

Maier, E., and Steinbach, M. (2022). Perspective shift across modalities. *Annu. Rev. Linguist.* 8, 59–76. doi: 10.1146/annurev-linguistics-031120-021042

Onishi, K. H., and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* 308, 255–258. doi: 10.1126/science.1107621

Overweg, J., Hartman, C. A., and Hendriks, P. (2018). Temporarily out of order: temporal perspective taking in language in children with autism spectrum disorder. *Front. Psychol.* 9, 1663. doi: 10.3389/fpsyg.2018.01663

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Poulin, C., and Miller, C. (1995). "On narrative discourse and point of view in Quebec sign language," in *Language, Gesture, and Space*, eds K. Emmorey, and J. Reilly (Hillsdale, NJ: Lawrence Erlbaum), 117–131.

Richardson, B. (2006). *Unnatural Voices. Extreme Narration in Modern and Contemporary Fiction*. Columbus: The Ohio State University Press.

Salem, S., Weskott, T., and Holler, A. (2017). Does narrative perspective influence readers' perspective-taking? An empirical study on free indirect discourse, psycho-narration and first-person narration. *Glossa* 2, 61. doi: 10.5334/gjgl.225

Sanders, J. (1994). *Perspective in Narrative Discourse* [Doctoral dissertation]. Tilburg: Tilburg University.

Sanders, J., and Redeker, G. (1996). "Perspective and the representation of speech and thought in narrative discourse," in *Spaces, Worlds, and Grammar*, eds G. Fauconnier, and E. Sweetser (Chicago, IL: The University of Chicago Press), 290–317.

Taylor, T., and Edwards, T. L. (2021). What can we learn by treating perspective taking as problem solving? *Perspect. Behav. Sci.* 44, 359–387. doi: 10.1007/s40614-021-00307-w

Trompenaars, T., Hogeweg, L., Stoop, W., and de Hoop, H. (2018). The language of an inanimate narrator. *Open Linguist.* 4, 707–721. doi: 10.1515/opli-2018-0034

Van Duijn, M., and Verhagen, A. (2019). Recursive embedding of viewpoints, irregularity, and the role for a flexible framework. *Pragmatics* 29, 198–225. doi: 10.1075/prag.18049.van

Van Krieken, K., Hoeken, H., and Sanders, J. (2017). Evoking and measuring identification with narrative characters – a linguistic cues framework. *Front. Psychol.* 8, 1190. doi: 10.3389/fpsyg.2017.01190

Van Schuppen, L., Van Krieken, K., and Sanders, J. (2019). Deictic navigation network: linguistic viewpoint disturbances in schizophrenia. *Front. Psychol.* 10, 1616. doi: 10.3389/fpsyg.2019.01616

Verhagen, A. (2007). "Construal and perspectivization," in *The Oxford Handbook of Cognitive Linguistics*, eds D. Geeraerts, and H. Cuyckens (Oxford: Oxford University Press), 48–81.

Zeman, S. (2017). Confronting perspectives: modeling perspectival complexity in language and cognition. *Glossa* 2, 6. doi: 10.5334/gjgl.213

Zeman, S. (2020). Parameters of narrative perspectivization: the narrator. *Open Libra. Humanit.* 6, 28. doi: 10.16995/olh.502

Zimmermann, J. T., Meuser, S., Hinterwimmer, S., and Vogeley, K. (2021). Preserved perspective taking in free indirect discourse in autism spectrum disorder. *Front. Psychol.* 12, 2710. doi: 10.3389/fpsyg.2021.675633

frontiers in Psychology

ORIGINAL RESEARCH
published: 23 November 2020
doi: 10.3389/fpsyg.2020.552368

Processing Speaker-Specific Information in Two Stages During the Interpretation of Referential Precedents

Edmundo Kronmüller[1]* *and Ernesto Guerra*[2]

[1] Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago, Chile, [2] Centro de Investigación Avanzada en Educación, Instituto de Educación, Universidad de Chile, Santiago, Chile

To reduce ambiguity across a conversation, interlocutors reach temporary conventions or *referential precedents* on how to refer to an entity. Despite their central role in communication, the cognitive underpinnings of the interpretation of precedents remain unclear, specifically the role and mechanisms by which information related to the speaker is integrated. We contrast predictions of one-stage, original two-stage, and extended two-stage models for the processing of speaker information and provide evidence favoring the latter: we show that both stages are sensitive to speaker-specific information. Using an experimental paradigm based on visual-world eye tracking in the context of a referential communication task, we look at the moment-by-moment interpretation of precedents and focus on the temporal profile of the influence of the speaker and linguistic information when facing ambiguity. We find two clearly identifiable moments where speaker-specific information has its effects on reference resolution. We conclude that these two stages reflect two distinct cognitive mechanisms, with different timings, and rely on different representational formats for encoding and accessing information about the speaker: a cue-driven memory retrieval process that mediates language processing and an inferential mechanism based on perspective-taking abilities.

Keywords: language processing, reference, speaker specificity, perspective taking, pragmatics, eye tracking

OPEN ACCESS

Edited by:
Franziska Köder, University of Oslo, Norway

Reviewed by:
Helene Kreysa, Friedrich Schiller University Jena, Germany
Mariapaola D'Imperio, Rutgers, The State University of New Jersey, United States

***Correspondence:**
Edmundo Kronmüller
ekr@uc.cl

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

Received: *15 April 2020*
Accepted: *20 October 2020*
Published: *23 November 2020*

Citation:
Kronmüller E and Guerra E (2020) Processing Speaker-Specific Information in Two Stages During the Interpretation of Referential Precedents. Front. Psychol. 11:552368. doi: 10.3389/fpsyg.2020.552368

INTRODUCTION

Definite repeated reference is a ubiquitous phenomenon during a conversation. Interlocutors talk about uniquely identifiable entities that are referred to multiple times using the same or a similar expression, such as "the black guitar," "my classmate," or "your boss." Because each entity can be referred to in multiple ways (e.g., "the old guitar," "she," or "Mrs. Smith"), and the same expression can be used to refer to different things (there are many black guitars, as well as classmates and people whose last name is Smith), interlocutors reach temporary conventions on how to refer to each entity. These conventions or referential precedents help in reducing ambiguity (Clark and Wilkes-Gibbs, 1986; Brennan and Clark, 1996). Therefore, speakers tend to use the same expression, and listeners expect that expression to be used to refer to the same entity across a conversation. Despite the central role that this phenomenon plays in communication and the considerable amount of

research done to explain it, the cognitive underpinnings of the interpretation of precedents remain a matter of debate, specifically concerning the mechanisms by which information related to the speaker is integrated.

One proposal is that interpreting precedents is a *one-stage* process resulting from the functioning of a single language processing mechanism that integrates speaker, linguistic, and all other available contextual information, as soon as it is available, and with little or no delay in this availability (Metzing and Brennan, 2003; Brown-Schmidt, 2009). This one-stage process is in line with constraint-based models postulated in the context of sentence processing and definite reference resolution (McRae et al., 1998; Spivey and Tanenhaus, 1998; Hanna et al., 2003).

An alternative proposal postulates two stages. The first one is accomplished by an initial interpretation relying solely on linguistic information, and the second one is achieved by a perspective-taking mechanism, which relies on speaker-specific information, and it is triggered on demand for correcting potential misunderstandings (Barr and Keysar, 2002; Kronmüller and Barr, 2007). This *original two-stage* approach is in line with an inference making mechanism that "anchors" an initial interpretation egocentrically and later "adjusts" based on mutual knowledge and speakers' beliefs (Keysar et al., 1998; Keysar et al., 2000; Epley et al., 2004). The following sections present a discussion of the existing literature that has been previously taken as support for either of these accounts of the online interpretations of precedents and offers an alternative to both proposals: an *extended two-stage* account.

Referential Precedents and Speaker Specificity

What a speaker wants to achieve when uttering a referring expression is to bring a specific entity into joint attention. The listener, in turn, should go beyond the conventional meanings of the words in the referring expression and consider background information, such as the time and place of the interaction, the goals of the conversation, and, critically, the identity and shared history with the speaker (Strawson, 1950).

An initial account on how reference resolution is achieved, under both the time constraints of a conversation and the cognitive demands on the language processing system, postulates an intrinsic context for comprehension. This context does not comprise all possible background information but only information shared among interlocutors and known as shared (Clark and Marshall, 1981; Clark and Carlson, 1981). It is by this mutual knowledge or *common ground* that a definite reference can point toward a unique referent; it narrows down the possible alternatives that otherwise would be many. Indeed, the same expression (e.g., "the black guitar") can be used to refer to many different objects in the world, but only to one—or a few—when the referential domain is restricted to the guitars that are mutually known by the interlocutors.

The common ground view is applicable to explain how referential precedents work; in fact, referential precedents can be considered a paradigmatic case for testing it. Under this view, precedents reflect a *conceptual pact* that is common ground among interlocutors (Brennan and Clark, 1996). Following our example, we can conceptualize a guitar as "black" (instead of "electric," "old," or "mine") and create the precedent "black guitar," such that it reflects that shared conceptualization. The benefit of using precedents on comprehension, in the strong version of this view, should be specific to the partner of the conversation with whom the pact was reached. To test this prediction, Barr and Keysar (2002) conducted an eye-tracking experiment. They showed that the benefit of precedent use—specifically on the speed of resolving reference as measured by the latency to look at the target object among all other possible objects in the referential domain—was not dependent on the precedent being common ground but in the use of the precedent *per se*.

In the experiment, a listener participant played the role of matcher in a referential communication game. The task was to arrange a set of objects in a 4×4 cubbyhole following the instructions from two confederate speakers playing the role of directors. One of the speakers interacted live with the participant, while the other had recorded the instructions previously that were played back through headphones only to the participant. This manipulation generates a situation in which the participant had privileged information regarding the names given to the different objects in the referential domain; in other words, one precedent—the one that was provided by the live speaker—was common ground between the speaker and the listener, but the other was privileged knowledge of the listener. They found that looks to the target object were faster when there was a referential precedent previously established. However, this benefit was independent of whether the precedent was common ground with the speaker or privileged for the listener. Based on this result, they postulated that the processing of referential precedents was *speaker-independent* instead of *speaker-specific*, in the sense that precedents provide a "linguistic index to the representation of the referent in memory" (Barr and Keysar, 2002, p. 392). In their proposal, speaker specificity only appears as the result of a slower adjustment mechanism, specifically perspective taking, that keeps track of common ground and has a role in correcting for potential misunderstandings.

In an influential study, Metzing and Brennan (2003) challenged this conclusion. They postulated that the interpretation of a precedent is guided by memory representations encoding not only the link between a precedent and its referent (as the idea of "linguistic index" suggests) but also a link to contextual features. Among these features is speaker-specific information (e.g., her identity and shared knowledge with the listener). Besides the situations where a precedent was maintained by speakers, as in Barr and Keysar (2002), they added two cases where a precedent was broken either by the original speaker who had established it or a new speaker uninformed of its existence. In one condition, one speaker established a precedent for a strange object (without a conventional name in English) calling it, in a first instance, "the silver pipe," but in a second instance, "the shiny cylinder." In the contrasting condition, one speaker called the strange object in a first instance "the silver pipe," while a second speaker, not informed about the precedent because she was not present at the moment it was established, call it "the shiny cylinder."

In the first case—when the original speaker breaks the precedent—the authors reasoned that interference, which is manifested in a delayed resolution of reference, should be observed because there are no good reasons for the original speaker to change the previously established precedent. In the second case, when a new speaker breaks a precedent, no interference should be expected because this new speaker could call that object in many different ways, inasmuch as she did not know about the existence of the precedent. They replicated Barr and Keysar's (2002) results for the maintained case (i.e., the speaker-independent effect of precedent), but they did find speaker specificity when the precedent was broken: listeners took longer to resolve reference when a new expression was used by the original speaker than when the precedent was broken by the new speaker. In terms of cognitive processing, they interpreted their results as reflecting the functioning of a single cognitive mechanism that integrates many different sources of information, which might have different levels of influence on comprehension based on their relative strength. In the case when a precedent is maintained, the effect of the speaker's identity was overwhelmed by the strength of the linguistic cue, not allowing the former to express. In contrast, when the linguistic cue is not as strong as the speaker identity cue—as in the case when a precedent is broken by presenting a brand new expression—the latter can be expressed, and its influence can be seen on comprehension.

These two studies set the current debate on the cognitive underpinnings of referential interpretation of precedents. It is either a one-stage process, as proposed by Metzing and Brennan, or a two-stage process, as proposed by Barr and Keysar (for an overview of this debate, see Brennan et al., 2010, 2018). In this debate, the main source of disagreement has been whether speaker-specific information influence comprehension at the same moment as linguistic information (Brown-Schmidt, 2009), or whether it plays a secondary role, expressed in a delayed influence compared with the linguistic input (Kronmüller and Barr, 2007). Because timing is at stake, the eye-tracking technique has been crucial since it allows observing the interpretation processes as it unfolds, making possible to determine the moment at which the different sources of information have their influence (Allopenna et al., 1998; Altmann and Kamide, 1999; for a review see Knoeferle and Guerra, 2016).

The Interpretation of Referential Precedents: One-Stage or Two-Stage?

The one-stage alternative postulates that the language processing system integrates linguistic and speaker information, along with other contextually relevant information, in an online fashion, that is, as soon as the information is available. Thus, the linguistic information carried by a precedent, the information about the speaker, the common ground between interlocutors, and other relevant contextual information all have an immediate effect on the interpretation of that precedent. Evidence for this immediate effect of speaker information has been found in cases where a precedent is maintained: there are faster looks to the intended referent when it is the original speaker that maintains a precedent than when a new speaker does (Barr, 2008; Brown-Schmidt, 2009; Horton and Slaten, 2012). Further evidence for single-stage processing of linguistic and speaker information has been found for definite reference (Hanna et al., 2003) and contrastive definite reference (Heller et al., 2008). There is also convergent electrophysiological evidence showing an early integration of speaker and linguistic information (see Van Berkum et al., 2008) and common neural pathways for processing lexical and speaker characteristics (Tesink et al., 2008).

The original two-stage alternative, in contrast, postulates a first stage where an utterance is processed independently of speaker information (for example, based solely on the linguistic input), followed by a second stage where speaker-specific information is integrated, particularly speaker's knowledge and beliefs (Kronmüller and Barr, 2007). Under this view, the influence of speaker-specific information should be clearly identifiable but only after the influence of the linguistic input (Barr and Keysar, 2002; Kronmüller and Barr, 2007; Kronmüller et al., 2017). In other words, the effect of who said what should not be seen before the effect of what is said. Evidence for two stages has been found for the cases when a precedent is broken: listeners avoid mapping a new linguistic expression onto a referent that has been named previously, independently of whether or not the previous referent-expression mapping is common ground (Kronmüller and Barr, 2007). Specifically, listeners look away from an object associated with a precedent, independently of whether that precedent was established by the original speaker or the new one (not aware of the existence of the precedent). Only after this initial interpretation will listeners allow a new referent-expression mapping for the new speaker for whom the original mapping was unknown, which is expressed in faster looks to the target with the new speaker than with the original speaker. This late process was characterized as a delayed *recovery* based on common ground information, and importantly, it was completely impaired by cognitive load, suggesting that it is a different cognitive process than the initial interpretation, which is only mildly affected (Kronmüller and Barr, 2007). Online evidence for two distinct stages has also been found for definite reference (Wu et al., 2013) and negated referential expressions (Kronmüller et al., 2017). Finally, brain imaging research has found convergent evidence for two distinct mechanisms for processing linguistic information and the speaker's beliefs and knowledge (Willems et al., 2010; Bögels et al., 2015).

Considering the wide disagreement between these two alternatives, which is mainly expressed empirically on the moment when speaker-specific information influences the interpretation of precedents (operationalized, as we have described above, as looks to the intended referent or target over time), Kronmüller and Barr (2015) conducted a meta-analysis. In this meta-analysis, they included all the experiments implementing the situations described above: (a) an original (same) speaker maintaining a precedent; (b) a different speaker maintaining a precedent from another speaker; (c) an original speaker breaking his/her own precedent, and (d) a different speaker breaking a precedent established by the original speaker.

Figure 1 presents the main results of Kronmüller and Barr's meta-analysis. The dependent variable was derived from eye-tracking data, specifically the *target advantage score*, which

reflects the preference to the target object compared with the other objects across time. *The blue line (left panel) represents a same speaker advantage for maintained precedents*, and it is computed as the difference of the target advantage scores of the same speaker maintained precedent condition and the different speaker maintained precedent condition. Thus, a positive value indicates an advantage for the same speaker when precedents are maintained. As can be seen, there is an early influence of speaker-specific information at approximately 500 ms. However, it decreases as interpretation continues (approximately 1,100 ms). *The red line (middle panel) represents the advantage for broken precedents*. This time, a positive value represents an advantage for the different speaker against the same speaker, in other words, the effect of speaker information when a precedent is broken. For this case, there is a delayed effect compared with the same speaker advantage (starting at 750 ms), but it only increases across time. Finally, *the green line (right panel) shows the advantage of the existence of a precedent*, independently of the speaker. As can be seen, the previous two are relatively small compared with this speaker-independent *effect of precedent*.

The meta-analysis showed clearly that the integration of speaker-specific information has two different timing profiles depending on whether the precedent is maintained or broken. As we will argue below, the time profiles of the effect of speaker-specific information are at odds with both of the accounts presented so far: one-stage and original two-stage.

In effect, the one-stage account sees precedent interpretation as a competition process. In this single process, the different alternatives (in this case referents) "compete" as they gain evidence from various sources of information integrated to comprehension as soon as they are available (Trueswell and Tanenhaus, 1994). As such, it is an instance of the *constraint-based* family of language comprehension models that have been successful in explaining syntactic, semantic, or referential ambiguity resolution (Elman et al., 2004; McRae and Matsuki, 2013). One key prediction from these models is that, if all sources of information point to the same alternative (in this case referents), competition should favor that alternative over all others *at every moment* during the whole process. Indeed, activation of alternatives, which in this particular case is referents, permanently increases in the same direction; if there is a systematic bias toward one alternative, and that bias is constant across the whole competition process, the less activated alternative should not be favored with respect to the other, more activated, alternative.

This pattern should be the case independently of whether or not those informational sources are available simultaneously and also regardless of the strength of association between the alternatives and the informational sources. Therefore, there should be a *monotonic increase* in looks to the target referent when the original speaker maintains a precedent until the competition process ends and one alternative "wins" (meaning that the intended referent is finally selected). The fact that the effect decreases—as can be seen in the green line in **Figure 1**—cannot be accommodated easily with this prediction.

On the other hand, the mere fact that there is an early effect of speaker-specific information in the maintained condition undermines the central claim of the original two-stage approach, as we have described so far. This approach inherits the view that contextual information, and critically speaker information (including common ground), has a role in an optional and more comprehensive process that monitors for potential misunderstandings (Keysar et al., 2000; Epley et al., 2004; Kronmüller and Barr, 2007). This account has its roots in early views on how parsing a sentence works: first, there is an interpretation based on only syntactic information and only later have a role in the semantic and contextual information (Frazier and Rayner, 1982).

As in the one-stage account, where the early integration of speaker-specific information provides evidence for it, there is also some piece of evidence that favors the original two-stage account. Indeed, even when speaker-specific information has an early influence, what has the most significant impact is whether or not there is an established precedent, independently of the speaker who established it (Kronmüller and Barr, 2015).

An Extended Two-Stage Account

There is a third possibility, which we defend here, that can reconcile the findings described above: an *extended two-stage account*. Under this idea, the information about the speaker is integrated at two different stages. In the first stage, precedent interpretation could be mediated by information about the conversational context, encoded in episodic memory, and retrieved in a *cue-driven automatic* fashion; such cues include the presence of speakers and salient perceptual features, such as their voice or gender (Horton and Slaten, 2012). Importantly, these episodic memory traces are not the same as common ground and speakers' beliefs but instead are the basis for inferring common ground (Horton and Gerrig, 2005). The second stage, we propose, might be based on *meta-representations processed by an inferential mechanism*, in line with perspective taking or "mindreading" (for a thorough discussion of mindreading, see Goldman, 2006). As with the original two-stage account, this inferential mechanism, based on information about the speaker, might be triggered as a result of the presence of ambiguity.

To the extent that our reasoning is correct, two moments for the integration of information about the speaker should be identifiable in the time course of the interpretation of precedents. These two moments are apparent in Kronmüller and Barr's meta-analysis at 900 ms, where the two lines, representing speaker-specific effects, cross each other. The present study tests this possibility, looking directly for two-stage processing of maintained precedents, both relying on speaker-specific information.

The Present Study

As presented above, the experiments on precedent interpretation so far have shown that referential precedents are interpreted in a one-stage process when they are maintained (Brown-Schmidt; Barr, 2008), but in a two-stage process when they are broken (Kronmüller and Barr, 2007). It might be the case that the second stage of integration of speaker information has not been observed on maintained precedents because the speakers' perspective is of no use when there is no (or only temporary)

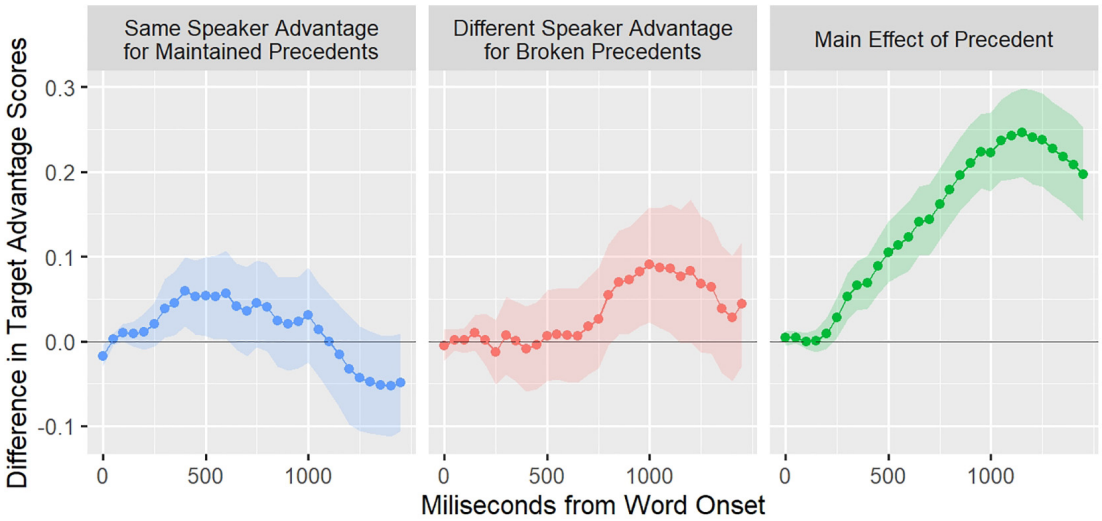


FIGURE 1 | Main results from Kronmüller and Barr's meta-analysis. Three clear effects are depicted, two of which are speaker-specific and one speaker-independent. Putting together, all of these effects have a different time profile. The blue line represents an advantage for comprehension when the same speaker maintains a precedent. The red line represents the advantage to comprehension when a different speaker uses a new expression. Finally, the green line represents the overall advantage of relying on precedents. This picture is an adaptation to Figure 3 in Kronmüller and Barr (2015).

ambiguity in the referential situation. With this in mind, we generate a situation of total referential ambiguity if only linguistic information is considered, but that could be disambiguated using information related to the speaker, information about the presence of a linguistic precedent, or both. This design allows us to test predictions from the one-stage model, the original two-stage model, and the extended two-stage processing model for maintained precedents we propose here.

Previous research may have failed to find the integration of speakers' information in two stages because this information is not necessary to resolve reference. In some cases, there was no ambiguity at all since there was always one referent that was the best candidate for the referring expression. For example, when the instruction was to select "the silver pipe," there was only one object in the referential domain that best resembled a silver pipe (Metzing and Brennan, 2003; Kronmüller and Barr, 2007). In other cases, ambiguity was momentary at either the level of the noun or the level of the adjective in the noun phrase (Brown-Schmidt, 2009; Horton and Slaten, 2012). For example, the instruction was "select the cat that's drinking milk," in a display with two images (e.g., tangrams), one mentioned before as "the cat that's drinking milk" and the other as the "cat that's sitting up." The tasks that show momentary or no ambiguity at all do not require speaker-specific information to be integrated into a second stage of processing, where common ground and speakers' perspective might be critical to disambiguate. In the present design, in contrast, due to total referential ambiguity that we created, we expect to see information related to the speaker to be used in referential resolution in a second stage as well.

In our experiment, participants saw three pictures of everyday objects on a computer screen and heard an instruction from one of two different speakers referring to one of these objects. In what we called *test trials*, ambiguity was introduced by the presence of two objects that could be named using the same word: a *target* object and a *competitor* object (see **Figure 2** for an example of a critical trial). For example, if the instruction was to "click on the bat," the possible referents were a flying mammal bat (target) and a baseball bat (competitor). Based on which referents were mentioned before the test trials, during what we called *presentation phase*, we generated four experimental conditions by combining two variables, with two levels each, in a within-subjects factorial design. The first variable was *Precedent* with two levels: *Precedent* and *No Precedent*. The second variable was *Speaker*, also with two levels: *Same* or *Different*. In what follows, we briefly describe each condition. A full example for an entire item will be presented in "Materials and Methods" section.

In the presentation phase, before each test trial, the target object (mammal bat) was either referred to or not. The competitor object (baseball bat), on the other hand, was never mentioned in this phase. Also, the target was sometimes mentioned by the same speaker as the one who gave the instructions in the test trials (female) or by a different one (male). Thus, whether or not the target object was mentioned before and by whom gives rise to our four conditions. In the *Same Speaker Precedent Condition* (*SSP*), the target object (mammal bat) was named twice before the test trial by the same speaker (female), establishing a referential precedent. In the *Different Speaker Precedent Condition* (*DSP*), by contrast, the speaker who gave the instruction on the test trials (female) was different from the one who had established the precedent in previous trials (male). In the *Same Speaker No Precedent* (*SSNP*) condition, the same speaker (male) gave the instructions in the presentation phase, but the target referent (mammal bat) was not mentioned. Instead, a different object was mentioned (the candle, the body, or the chest). Finally, in the *Different Speaker No Precedent Condition* (*DSNP*), there was a change of speaker before the test



FIGURE 2 | Test trial example. Three objects appear on the screen. Two of them are referential competitors because they can be referred to using the same word: the mammal bat and the baseball bat, which can be both refer to as the bat. The third picture is a distractor, not related with the previous two. The task is to select an object upon hearing an instruction of the type: "click on the bat," which is ambiguous in this display.

trial, and the target object (mammal bat) was not mentioned in the presentation phase. Specifically, the male speaker gave the instructions in the presentation phase, but the female speaker did it for the test trial. By changing the speaker in the presentation phase (instead of the test trial), we had the same token instruction for all four conditions in each item, allowing strong experimental control. With this design, we tested the hypotheses derived from the competing accounts.

To specify predictions, **Figure 3** shows a schematic representation of how eye-tracking data should look like for each account in all conditions. We used a measure that can summarize the competition process between the target and competitor objects. We chose a *log ratio*, which gives a positive value if there was a preference in looks to the target (mammal bat), a negative value if the preference was on the competitor (baseball bat), and zero if there was equal distribution of looks between the objects (we further explain this measure in "Materials and Methods" section). We opted for this dependent variable rather than target proportion of fixations because the latter incorporates all referents in the referential domain in the value, independently of whether or not they are of interest. Log ratios, on the other hand, only consider the referents that are of interest for testing the hypotheses. In our case, these are both referential competitors.

First, all competing accounts predict an advantage to the target object in both precedent conditions, independently of the speaker. In **Figure 3**, this effect is represented in the difference between the two green lines and the two red lines. The first two go up, whereas the second two remain at zero all the time (black thin line). Indeed, without a precedent associated with the objects, there should not be any preference for one of them, making the log ratio close to zero. Where the accounts should contrast is in the difference between the two speaker conditions when there is a precedent, namely, the SSP and DSP conditions. The one-stage account predicts an early separation of the curves since two cues are pointing to the target in the SSP condition, a linguistic cue in the form of a referential expression, and a contextual cue in the form of speaker's identity. In the DSP condition, in contrast, only the linguistic cue points to the target. Importantly, and as we argue above, both curves should increase monotonically.

The original two-stage account predicts a late separation of both precedent curves, since in the first stage only the linguistic information is processed. The contextual information is considered in the second stage through a perspective-taking process.

The pattern that would be consistent with the extended two-stage account should present two clear moments at which speaker information contributes to the preference to the target object over and beyond the existence of a precedent. The rightmost panel in **Figure 3** shows such a pattern. The early separation of SSP and DSP curves would reflect the automatic cue-driven process, whereas the late separation would be the consequence of an inferential process based on perspective taking, which would be triggered by the presence of ambiguity.

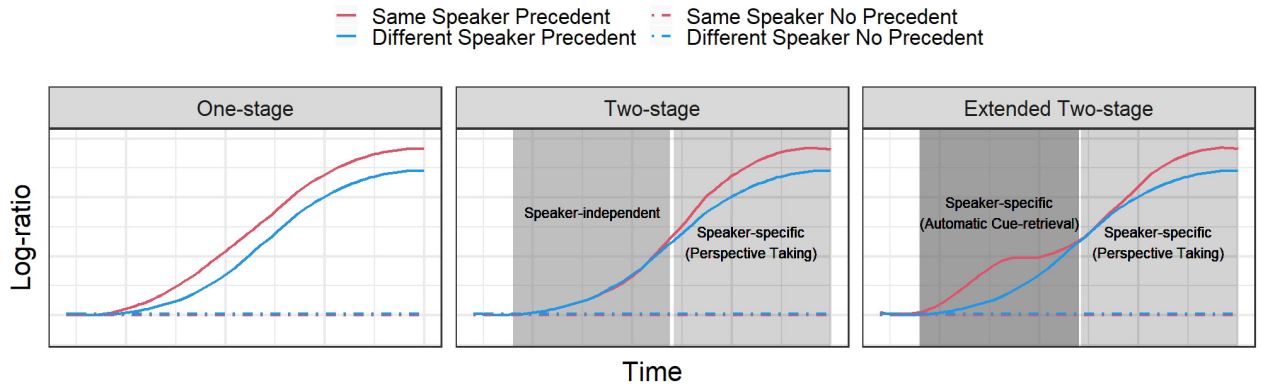


FIGURE 3 | Schematic predictions from each model. Each of the models makes different predictions with respect to the temporal pattern of the log ratio between the target and competitor objects. A value close to zero means no preference to either of the referents. A positive value reflects a preference to the target, and a negative value a preference to the competitor. The left panel shows the pattern predicted by the one-stage model, where there is an early separation between the lines representing the SSP condition and the DSP condition, favoring the former. The middle panel shows the pattern expected for the original two-stage model, where there is an early speaker-independent processing and a late perspective taking process. Finally, the panel on the right shows the predictions from the extended two-stage model, where there is an early speaker-specific processing due to automatic memory processes and a late effect related to perspective taking mechanisms. All models predict no difference in preference between referents across the whole time window when a precedent has not been previously established.

MATERIALS AND METHODS

Participants

Fifty-six undergraduate students (age range: 18–24 years) participated in this study in exchange for course credit for an introductory course in psychology at the University of California, Riverside. All of them had normal or corrected-to-normal vision and were native English speakers. Thirty-nine of the participants were female.

Design

The experiment had a 2 (Same Speaker or Different Speaker) × 2 (Precedent or No Precedent) within-subjects factorial design.

Materials

We created 24 items. **Figure 4** presents an example of an item and its corresponding elements (pictures and sound files): each item consisting of six pictures of everyday objects and 16 sound files with instructions (e.g., "Click on the bat") recorded by two different speakers. Two of the objects were referential competitors, i.e., objects that could be referred to using the same English word. For the sake of exposition, we called one these objects "target" and the other "competitor." For example, as in **Figure 2**, a mammal bat and a baseball bat can both be referred to as "the bat" (see **Figure A1** for the complete list of referential competitors). One of the remaining four pictures was used to introduce salience to the speaker manipulation in a "Speaker Change Cue" trial (see "Procedure" section and **Figure 4**). The three remaining pictures were fillers. Each item was a sequence of 10 displays, presenting three pictures at the same time. Therefore, we use different combinations of the six objects pertaining to an item to generate the 10 displays of each item. For the sake of exposition, we will refer to each display as a "trial."

On each trial, participants heard sound files, containing the instructions, depending on the experimental condition and the type of trial. Using pre-recorded speech—instead of live instructions from a speaker—allowed us to present the same instruction token in every condition. With this feature, we avoid possible confounding variables, such as paralinguistic cues (a point we come back to in the discussion). We used female and male voices for the speaker manipulation, as these have been shown to boost talker-specific effects in lexical recognition tasks (Creel et al., 2008). For half of the items, the female voice gave the instruction in the critical trial and the male voice for the other half.

Procedure

Participants were tested individually in a quiet and semi-dark room. They sat facing a projection of a computer screen on the wall. In between the projected screen and the participants, there was a table with an ISCAN ETL-400 remote eye tracker (60 Hz sampling rate). Participants were informed that they would play a communication game in which they would have to follow the directions of two different speakers. Each time, one of the speakers would instruct them to select one picture among the three images presented on the screen. They were told that each speaker recorded the instructions in a different session so that they might refer to the objects differently between them. We also told them that because speakers did not interact with one another, they were uninformed of each other's precedents. Additionally, participants were led to believe that the order of the instructions they heard was the same as the order that they were recorded.

Participants were also led to believe that when the speakers recorded the instructions, they could only see two of the three pictures on the screen. This feature was introduced for two reasons. First, we wanted to make the situation more natural and cooperative from the point of view of the listener. Since the instructions were ambiguous between two objects, a

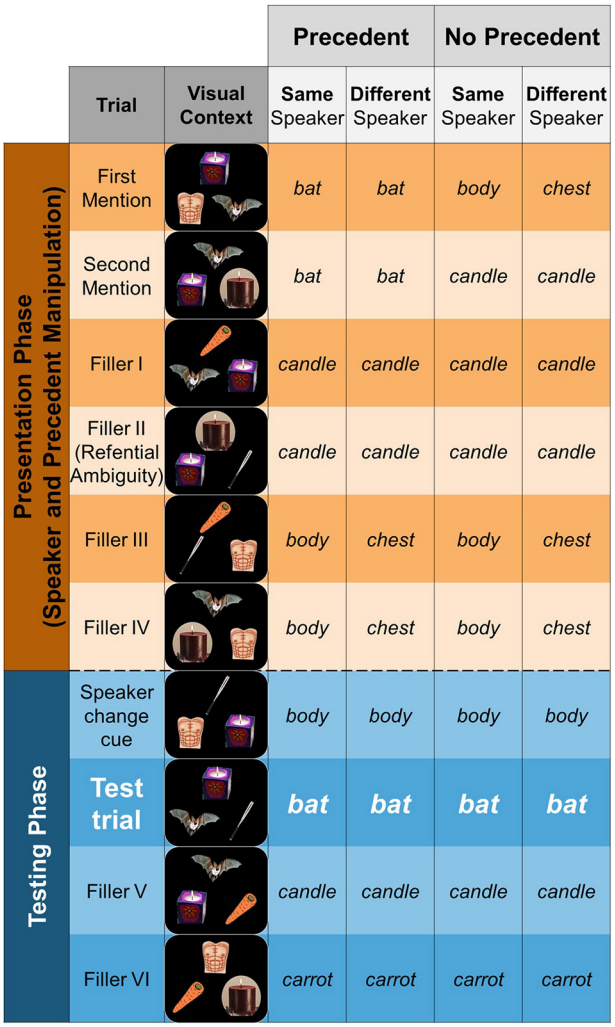


FIGURE 4 | A schematic representation of a 10-trial sequence in which a test trial is embedded (either in the eighth or ninth position). The sequence is divided into two phases: Presentation (first six trials in orange) and the Testing (subsequent four trials in blue). Precedent, or its absence, is established during Presentation, producing the Precedent vs. No Precedent conditions. During Testing, speaker is either maintained or changed (relative to Presentation), resulting in the Same vs. Different Speaker condition.

speaker normally would disambiguate their utterances by further specifying the intended referent. For example, in the presence of a mammal bat and a baseball bat, a speaker would say "the mammal bat" and not merely "the bat" if she needed to specify a referent uniquely for an addressee. The second reason concerned the implementation of a guessing game, where the participant had to guess which picture on the screen the speaker did not see. This guessing game provided data about the possible explanations that participants generated to explain why the speaker was ambiguous; specifically, if they selected one of the competitors (mammal bat or baseball bat), then it would be clear that they believed the reason was that when producing the utterance the speaker could not see the other object that was the source of the ambiguity. **Figure A2** depicts a schematic representation of the instructions given to the participants (the real instructions were given in different sequential displays on the screen).

To make the experimental situation more credible, right before the trials started, participants engaged in a role play where they gave instructions on which object to select to the experimenter. With this process, we led them to believe that they were experiencing a similar situation as the one that the speakers they were about to hear experience when recorded the instructions. The only difference was that the addressee was the experimenter instead of a real participant. After the instructions were recorded, they were played back to the participant to practice the actual task.

A Description of an Item: Phases and Type of Trials

Each test trial was embedded in a 10-trial sequence, corresponding to one item. There were three other types of trials in this sequence: the *presentation trials*, the *filler trials*, and a *speaker-change cue trial* (see **Figure 4** for a schematic representation). The sequence was divided into two phases, a *Presentation Phase*, consisting of six trials, and the *Testing Phase*, with four trials. As we described above, the precedent manipulation was implemented during the Presentation Phase via the presentation trials, where the target referent (e.g., the mammal bat) was referred to either twice (precedent conditions) or never (no precedent conditions). In both presentation trials for the precedent conditions, the speaker referred to the mammal bat as "the bat." In the presentation trials for the no precedent conditions, the speaker referred to one of the filler objects in the display (e.g., the chest and the candle in **Figure 4**). The speaker manipulation was implemented using the transition between the presentation and testing phases. In the same speaker conditions, the speaker that provided instruction in the presentation phase (e.g., female) continued giving the instructions in the testing phase. In the different speaker conditions, by contrast, the other speaker (e.g., male) provided the instructions for the presentations phase. As we said above, this feature allowed us to have the same token instructions in each item.

Speaker-change cue trials were included to add salience to the (same vs. different) speaker manipulation. On these trials, one object was referred to using a different name as in the presentation phase when there was a speaker shift. For example, if an object was called "the chest" by one speaker during the filler trials in the presentation phase, then the same object was called "the body" by the other speaker in the testing phase (see **Figure 4**). By doing this, we hoped participants would notice the shift and strengthen their belief that the speakers were unaware of one another's referential commitments. Additionally, we added a picture of the speaker that was about to give instructions. Finally, to further mask the critical trial, we added a filler trial with two referential competitors.

The test trials took place in the testing phase. As we have described, ambiguity for these trials was introduced by the presence of two objects that could be referred to using the same expression: a mammal bat (target) and a baseball bat

(competitor). Therefore, the instructions for the critical trials contained the same linguistic item that can be applied to either competitor, e.g., the word "bat" in the instruction "click on the bat." The task of the participants was to select one of the three objects using a computer gamepad. Participants' eye movements and picture selections were recorded. Given our goal in this study, the data we were interested came only from the test trials.

Finally, after the participants made their selection, the guessing game took place. In the guessing game, the three pictures appeared again on the screen, and the task of the participant was to select the one that they believed the speaker could not see when the speaker gave the instruction. In order to cover the identity of the test trials, the guessing game was also implemented after two other trials in each item during the presentation phase.

Data Analysis

Selection data (the object the listeners chose) and reaction times to that decision were analyzed using a mixed-effects regression with cross-random effects (Baayen et al., 2008), with subjects and items as random effects and Speaker and Precedent as fixed effects. We tested the effects using a model comparison approach, where a full model is compared with a restricted model without the fixed effect of interest and a chi-squared statistic is reported. The full model contained all random intercepts and slopes. Random correlations were excluded because the model would not converge (Barr et al., 2013). For the selection data, we performed a logistic regression because the total of selections of the unrelated object was 3 out of 1,344 trials, making it, in practice, a dichotomous variable. For reaction times, we log-transformed the data to ensure normality in its distribution to conduct the statistical test. We, however, report the raw data, i.e., in milliseconds.

Selection and reaction time inform us about the result of the comprehension process. However, they do not inform us about the temporal profile of the process of interpreting a referential expression. Since our hypotheses concern a specific temporal pattern determined by the underlying cognitive processes, we conducted a growth curve analysis (GCA). This analysis can capture a temporal profile by fitting a polynomial function and can provide estimates for each component of it (Mirman et al., 2008; Mirman, 2014). For this particular data, we fitted a third-degree polynomial function, capturing the linear, quadratic, and cubic components.

To compute the *p*-values for each of the terms in the polynomial function, we decided to conduct a resampling test (Carsey and Harden, 2013; see Kronmüller et al., 2017; Kronmüller and Noveck, 2019, for the use of resampling tests to conduct inferential statistics on eye-tracking data). As we mentioned above, since our main interest was the pattern of transitions between the target and competitor objects, our dependent variable was the log ratio of the looks to the competitor over the target (see Arai et al., 2007). In this measure, positive values reflect more looks to the target than to the competitor object, negative values reflect a relation in the opposite direction, and values close to zero show no preference to any of them. We computed this log ratio by fitting baseline-category multinomial logistic regression (Agresti, 2003).

The algorithm to compute the *p*-values had several steps. In the first step, we obtained the regression coefficient for each of the polynomial terms for the original dataset (Mirman, 2014). Second, we draw a Monte Carlo random sample of 9,999 permuted datasets (Carsey and Harden, 2013; Berry et al., 2016). Third, we created a null hypothesis distribution of the regression coefficients for each polynomial term, to each permuted dataset, and we stored them. Therefore, we obtained 9,999 coefficients for each polynomial term. And fourth, we compared the original regression coefficients with the null hypothesis distribution by calculating the proportion of coefficients that were larger or smaller. These proportions are the *p*-values. Since we had specific hypotheses about (a) the cubic pattern between the Same Speaker Precedent and Different Speaker Precedent conditions and (b) the linear pattern of the effect of precedent independently of the speaker, we performed one-tailed tests. We computed two *p*-values for each term, one for subjects ($p1$) and one for items ($p2$).

RESULTS

Participants followed the instructions for the experiment and were attentive to the speaker manipulation and the task. Evidence for this can be found in the relation between the object that participants selected in the main task and the guessing game. When they chose the target (e.g., mammal bat) in the main task, they selected the competitor (e.g., baseball bat) in the guessing game 90.0% of the time. And when they selected the competitor in the main task (e.g., baseball bat), they chose the target (e.g., mammal bat) in the guessing game 78.3% of the time. This pattern also shows that, even though listeners heard pre-recorded expressions, they expect the speakers to disambiguate if necessary; in other words, listeners' pragmatic expectations were not compromised during the experiment.

Target Selection Data

Participants preferred the picture named before as the target: it was selected 78.0% of the time in comparison with 54.3% of the time when there was no precedent. This main effect of Precedent is statistically significant [$\chi^2(1) = 30.341$, $p < 0.0001$]. Also, there is a reliable interaction [$\chi^2(1) = 7.342$, $p = 0.007$] driven by a higher selection of the target object when the same speaker established the precedent (82.1%) than when a different speaker established it (72.8%) [$\chi^2(1) = 6.904$, $p = 0.009$]. No statistically reliable simple effect was found between the two speaker conditions when there was no referential precedent [$\chi^2(1) = 1.737$, $p = 0.188$]. **Table 1** presents the percentage of object selection by experimental condition.

Reaction Times

Table 2 presents the mean reaction times by conditions. Participants were faster to make their selection when there was a precedent (M = 1,530 ms; Mdn = 1,212 ms) than when there was no precedent (M = 1,720 ms; Mdn = 1,375 ms). This difference is statistically reliable [$\chi^2(1) = 6.375$, $p = 0.012$]. In contrast, there was no main effect of speaker: the mean reaction

TABLE 1 | Percentage of selection of each object across conditions.

	Same speaker		Different speaker	
	Precedent	**No precedent**	**Precedent**	**No precedent**
Target	82%	52%	74%	57%
Competitor	18%	48%	26%	43%
Unrelated	0%	0.03%	0.1%	0%

TABLE 2 | Reaction times of object selection across conditions.

		Same speaker	**Different speaker**
	Main effects	1,587 ms	1,663 ms
Precedent	1,530 ms	1,536 ms	1,525 ms
No precedent	1,720 ms	1,638 ms	1,802 ms

times for the same speaker (M = 1,587 ms; Mdn = 1,267 ms) were not statistically different [($\chi^2(1) = 2.037$, $p = 0.154$] from the mean for the different speakers conditions (M = 1,663 ms; Mdn = 1,298 ms). Also, the interaction was not statistically significant [$\chi^2(1) = 0.842$, $p = 0.359$]. The mean and median for each condition were as follows: Same Speaker Precedent (M = 1,536 ms; Mdn = 1,230 ms); Different Speaker Precedent (M = 1,525 ms; Mdn = 1,209 ms); Same Speaker No Precedent (M = 1,638 ms; Mdn = 1,328 ms); and Different Speaker No Precedent (M = 1,802 ms; Mdn = 1,387 ms).

In summary, the selection data show that by the end of the interpretation process, listeners' interpretation of referential precedents has a speaker-specific component. This finding is consistent with the one-stage, original two-stage, and extended two-stage accounts. The reaction time data portray a slightly different picture. Participants' final selection took longer in the No Precedent conditions than in the Precedent conditions, independently of the speaker. Then, even when more competition would be expected in the DSP condition than in the SSP condition (since more informational cues point in the target direction in the latter than in the former), participants do not take longer to make their selection.

Notwithstanding, we can truly differentiate between the three accounts by looking at the interpretation process as it unfolds and observe the influence of the speaker information as it emerges over time. We now present the eye-tracking data.

Eye-Movement Data

Figure 5 shows the proportion of looks to each object from the onset of the critical noun in the test trial (e.g., "bat") to 1,300 ms after it. Fifty percent of trials terminated within this time window because subjects had made their selection. The graph starts at 200 ms from the onset of the critical word since it is the time that has been estimated to program an eye movement (Matin et al., 1993). Therefore, from 200 ms after the onset, eye movements can be driven by the linguistic input, in this case, the noun in the noun phrase (e.g., "bat"). The mean duration of nouns in the stimuli set was 455 ms (range 351–806 ms).

From the graphs for each condition, it can be observed that the divergence between the target and the competitor objects starts earlier in the SSP condition than in the DSP condition. It is also clear that when there is no precedent, participants cannot overcome ambiguity, looking equally to both—the target and the competitor—throughout the time window.

To test our two hypotheses, we conducted our statistical analyses on the log ratios in a specific window. **Figure 6** shows these log ratios in the time window of interest, which we define between 350 and 850 after noun onset. We determined this window following two criteria. The first one was to match the timing on which the same-speaker benefit appears in Kronmüller and Barr's (2015) meta-analysis (see **Figure 1**). The second criterion was to ensure a clean cubic pattern for the analyses so that the first time point was not be preceded by a time point with a larger log ratio and the end time point was not be followed by a lower log ratio.

One hypothesis was the specific non-monotonic pattern for the SSP condition in contrast to the DSP condition, which would reflect the influence of speaker specificity in two different stages. To test this hypothesis, we compared the SSP and DSP conditions. As can be observed, the SSP condition presents a cubic pattern, which reflects the influence of speaker information at two stages. This pattern is not observed in the DSP. The cubic component of the polynomial is observed in the SSP condition and not in the DSP condition ($p1 = 0.027$, $p2 = 0.031$). Neither the linear ($p1 = 0.703$, $p2 = 0.670$) nor the quadratic term was significant ($p1 = 0.689$, $p2 = 0.652$).

The other hypothesis was related to the effect of precedent, independently of the speaker. As can be observed in the graphs, there is an increase in log ratios (more looks to the target than to the competitor) in both precedent conditions than in the no precedent conditions. The linear component of the polynomial for the main effect of precedent is significant by subject ($p1 = 0.013$) and marginally significant by item ($p2 = 0.085$). Neither the quadratic nor the cubic was significant.

The only other effect that reached statistical significance was a main effect of speaker in the cubic term ($p1 = 0.031$; $p2 = 0.032$), favoring the Same speaker conditions, which we had not hypothesized.

DISCUSSION

Interlocutors frequently rely on referential precedents to reach mutual understanding regarding the entities being referred to across a conversation. These referential precedents carry information that goes beyond the conventional meaning of the words, as they are encoded along with information related to the history of a conversation with specific partners. In this study, we contrasted predictions from a one-stage model of referential precedents, an original two-stage model, and an extended two-stage model, and we focused on the integration of speaker-specific information. The first model, as a version of constraint-based models of language processing, predicts—in the context of our design—a monotonic increase for preferences to the target when speakers' information is added to lexical and precedent information. The original two-stage model predicts only a late integration of speaker information. The model we defend here,

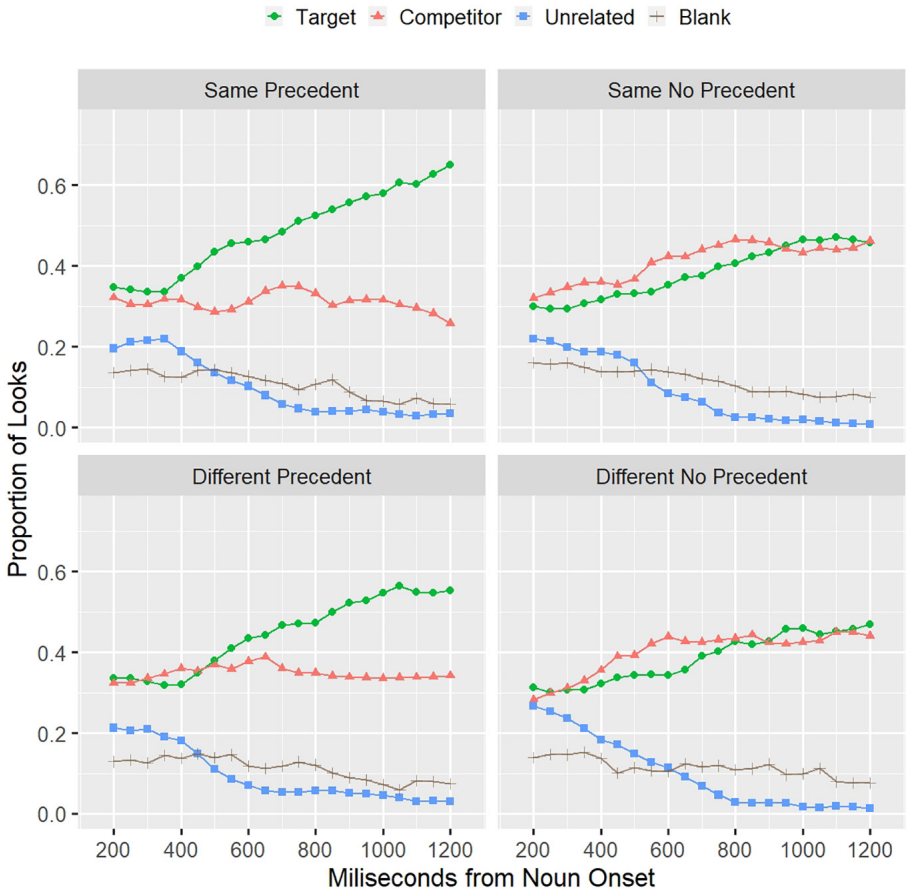


FIGURE 5 | Proportion of looks to each object and the blank part of the screen across the four conditions. Data are time-locked at the onset of the noun in the referring expression (e.g., "the bat").

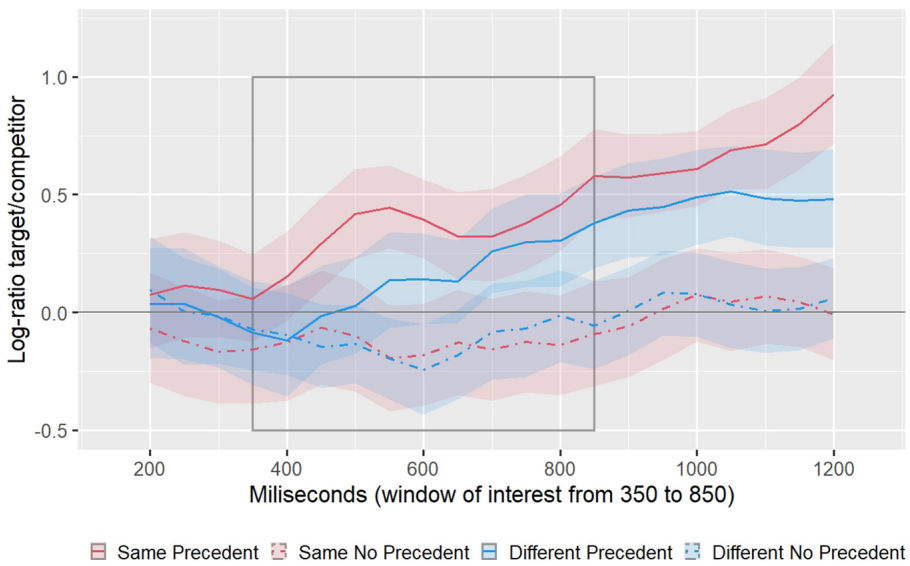


FIGURE 6 | Log ratio of fixations to the target over the competitor between 350 and 850 ms after the onset of the noun in the referring expression (e.g., "the bat"). The solid line represents the mean, and the shaded area represents 95% confidence interval, both calculated by bootstrapping subjects. The gray rectangle represents the window of interest.

the extended two-stage approach, predicts that information about the speaker has its influence at two different moments. Thus, our rationale was to distinguish these accounts based on a temporal dissociation of the effect of speaker information during the interpretation of precedents. We generate an experimental situation of referential ambiguity that could be resolved using speaker information, the presence of a precedent, or both. Consistent with the extended two-stage account, we found that speaker information has an influence at two different moments during online interpretation: descriptively speaking, we can see that at ∼350 ms after noun onset, an SSP advantage begins, only to decrease at ∼550 ms and reappear again at ∼700 ms up to the end of the time. This drifting reflects the cubic pattern predicted by the extended two-stage account for the SSP condition.

We propose that the first moment where speaker information is observed might be the result of an automatic cue-driven retrieval process of episodic memory traces that encode contextual information (among which are previous encounters with the precedent and the speaker). The second moment when the speakers' information "kicks in" can be seen as the result of mentalizing or perspective-taking mechanisms for inferring communicative intentions based on speaker's beliefs and knowledge. One result that might have been expected is a difference in reaction times between the two speaker conditions. In effect, there is evidence that speaker voice have an effect on lexical access (Goldinger, 1998). We are not clear why this effect is not observed in our data. One possible explanation is the strength of the precedent effect that might override the effect of speaker. Another possible explanation is that reaction times are rather large in our experiment compared with Goldinger's, mainly because of task differences and the existence of ambiguity in our paradigm. These large reaction times could have hidden the effect of speaker.

How Can These Results Be Interpreted in Terms of Mechanisms?

In the context of this experimental design, the non-monotonic pattern of the influence of speaker is inconsistent with a one-stage process that would reflect the functioning of a constraint-based mechanism (Brown-Schmidt, 2009; Brennan and Hanna, 2009; Brown-Schmidt and Hanna, 2011). Basically, the two sources of information that could be used to disambiguate the referential expression "the bat"—the presence of a precedent and speaker identity—pointed to the same alternative, in our example, the mammal bat. From a one-stage model, the activation reflected in eye movements toward the target object should increase systematically until the final selection, but this pattern was not observed in our data. Instead, we found an early but transient effect of speaker that disappears and appears again progressively and was maintained until final referential commitments. Similarly, a constraint-based mechanism would predict a fast resolution of the ambiguity in cases were more cues point in the direction of a referential interpretation, which was the case of the Same Speaker Precedent condition compared with the Different Speaker Precedent condition (in the former, the speaker cue was present, and in the latter, it was not present). However, there was no difference in reaction times for final selection between these two conditions, showing further inconsistent evidence for this model.

It is relevant to stress the importance of the prediction of monotonicity for the design presented here, since constraint-based models have been subject of criticisms for being unfalsifiable. This critique has been considered somehow unfair by the proponents of these models (Elman et al., 2004). The criticism points that these models could, in theory, accommodate any pattern of results, since they have many free parameters, such as the weights between constraints and the alternative interpretations, the activation of those constraints, and the timing of their availability. To address this critique in the context of sentence comprehension research, the main effort has been to explicitly set those parameters by consulting corpora and conducting behavioral experiments (McRae et al., 1998; Spivey and Tanenhaus, 1998). Unfortunately, such parameterization has not been pursued in the research on dialogue and speaker specificity in reference resolution.

It has been acknowledged that such an enterprise is hard to fulfill in this context, and thus manipulations of speaker salience—mainly through changing the experimental task for making it more interactive—have been the preferred practice to test the model (Brown-Schmidt and Hanna, 2011). We argue that our design can test direct predictions without specifying weights, activation, and availability and without relying on making speaker information more salient. We achieve this by creating a totally ambiguous situation if contextual information is not considered, and in which all constraints point in the same direction. It is important to clarify that we are not saying that these models cannot show non-monotonic patterns but that in our design they should not.

More generally, there is a conceptual issue that makes a constraint-based approach to dialogue not totally adequate and that is related with the inferential nature of communication (Grice, 1957; Sperber and Wilson, 1987; Clark, 1996). It is not clear how these models could deal with meta-representations, such as speakers' beliefs about mutual knowledge and common ground. Common ground and speakers' beliefs are not just simple representations but also representations of mental states or meta-representations, and it is not clear how they can be encoded in episodic memory (Horton and Gerrig, 2005). An interesting proposal has been put forward recently to account for the role of meta-representations on reference production and resolution, where a meta-representation, once inferred, could be stored as any other representation, and from there, it can influence language processing as any other representation (Horton and Brennan, 2016). The representational format of common ground in memory is an interesting area of research that is critical to understand perspective taking in conversation (for discussion on the role of memory in common ground during conversation, see Brown-Schmidt and Duff, 2016).

How could the present results be accommodated under an extended two-stage account? We take the view that precedent interpretation involves two different mechanisms, each one related to one of the two stages. First, interpretation of precedents can be characterized as mediated by contextual information

encoded in episodic memory. This information includes the history of the precedent (time, place, and identity of the speaker, including voice and gender) since it was first established (Metzing and Brennan, 2003). In short, it is some sort of source memory (Ryskin et al., 2015). The mechanisms for the integration of this information can be characterized as an automatic cue-driven retrieval process (Horton and Slaten, 2012; Horton and Gerrig, 2016) or episodic priming (Barr et al., 2014).

The second stage could reflect the functioning of an inferential mechanism that takes into account speakers' beliefs to explain their referential behavior. In the case of the present study, the behavior corresponds to the use of a certain linguistic form to bring into joint attention a particular entity in the context of ambiguity—where the linguistic form is insufficient to uniquely identify a referent as is required by the use of definite reference. The specific characteristics of this mechanism are still unclear. In some sense, it looks like a full mentalizing process cannot really account for the data, specifically for the big effect that the presence of a precedent has on comprehension, independently of who established it. Indeed, if participants had taken speakers' perspective in full, the Different Speaker Precedent condition should look similar as the other two No Precedent conditions; a new speaker could plausibly refer to either the mammal or the baseball bat by uttering "the bat." One possible line of explanation can be found in the philosophical discussion on the mechanisms for predicting and explaining others' behavior or *mentalizing*. In particular, *simulation theory* (Gallese and Goldman, 1998; Goldman, 2006) postulates a mechanism where people attributing mental states to others use their own mental states instead of meta-representations of the others' metal states. Translated to language interpretation, a listener could use her own mental states as proxies to the speaker's and from there could make inferences to explain the referential behavior. Using their own mental states as proxies can bias interpretation in an egocentric way (Goldman, 2006). Moreover, this simulation can be characterized as driven by heuristic reasoning (Gigerenzer et al., 1999; Todd and Brighton, 2016). This could be considered a shortcut to infer communicative intent that relies on less information but still generates *good enough* interpretations (Ferreira and Patson, 2007), which are efficient despite the time and cognitive demands of a conversation. In any case, whether full mindreading or simulation, an inferential mechanism ought to be included in an account of referential communication in dialogue. Indeed, one of the major conceptual advances in the study of communication in psychology and human sciences is to consider it an inferential process in nature (Grice, 1957; Sperber and Wilson, 1987).

Limitations and Future Research

A potential limitation of the present study is related to the "social" situation in the experiment. In order to gain experimental control, we decided to use pre-recorded expressions instead of the presence of two "live" speakers. As we mentioned before, this allowed us to measure responses to exactly the same token instruction, avoiding undesirable noise to the data. This is in line with Kuhlen and Brennan's (2013) suggestions on the adequate use of confederates in dialogue research. Concretely, they argue that "controlled" dialogue is best exploited when the confederate speaker is meant to initiate the interaction and when testing hypotheses related to an unusual situation (i.e., referential ambiguity in the present study). However, this situation is not as naturalistic as desired when studying dialogue, and this has been an important argument when testing predictions from a constraint-based mechanism. Indeed, the salience of speaker information might be weaker compared with the presence of a live speaker in the context of an interactive task (Brown-Schmidt, 2009; Brown-Schmidt, 2012; Brown-Schmidt et al., 2015). However, in the light of a recent meta-analysis, this claim should be qualified, since the argument only holds with a selective inspection of the literature, i.e., focusing on certain experiments and only some conditions (Kronmüller and Barr, 2015). Moreover, in the present experiment, we emphasized the saliency of the speaker through the inclusion of broken-precedent trials, by having speakers of different gender and by showing the speakers' pictures.

Indeed, early effects of speaker for maintained precedents have been found in interactive (Brown-Schmidt, 2009) and non-interactive settings (Kronmüller and Barr, 2007; Horton and Slaten, 2012). But they have not been found in interactive ones as well (Metzing and Brennan, 2003). The fact that we did find speaker effects is a compelling argument for the reliability of our results. Indeed, even without a real interactive situation, listeners can make speaker-specific interpretations. Moreover, higher (or lower) salience as a result of the presence of a live speaker and interactivity of the situation cannot explain the integration of speaker information in two different moments, as we show here. We do agree, however, that a more ecological situation is important to test hypotheses related to cognitive mechanisms in dialogue, but we also believe that experimental control is necessary to test hypotheses that differentiate effects at the scale of milliseconds.

The present study can motivate future research to further investigate the nature of the cognitive mechanisms involved in precedent interpretation. Given that our proposal sought to conciliate previous findings in the literature, it is necessary to generate new experimental situations to further understand the nature of the mechanisms underlying the two stages. One possibility is to generate a condition of cognitive load. Since we have characterized the integration of speaker information in the first stage as an automatic retrieval process, and in the second as an inferential one, the former should be less affected than the latter by higher memory demands. Evidence in line with this differential effect of cognitive load has been found before, but not in the context of full ambiguity (Kronmüller and Barr, 2007).

Also, there is a need to move forward and generate computational models of the mechanisms being tested. As mentioned before, modeling has been of major importance for testing predictions of models in sentence comprehension (McRae et al., 1998; Spivey and Tanenhaus, 1998). There are recent efforts to mathematically model and test predictions in the context of pragmatic inference and referential communication in dialogue (Frank and Goodman, 2012; Heller et al., 2016). However, they stress the normative character of pragmatic inference based on Bayes' rule. We believe that these efforts are important but that

a more mechanistic approximation to computational modeling is needed, for example, with ACT-R architecture (Anderson et al., 2004; Hendriks, 2016).

CONCLUSION

In conclusion, referential precedents appear to be carrying a great amount of information related to lexical, indexical, and contextual information, all of which helps in reducing ambiguity and facilitates mutual comprehension. The present research suggests that such a rich array of information distinctions calls for more than one cognitive mechanism, some of them related to memory processes, while others related to social cognition.

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 the University of California, Riverside. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EK designed the study. EK and EG analyzed the results and wrote the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Fondecyt-Chile grant (1200655) to EK. The support from ANID/PIA/Basal Funds for Centers of Excellence FB0003 to EG is also gratefully acknowledged.

REFERENCES

Agresti, A. (2003). *Categorical Data Analysis*, Vol. 482. Hoboken, NJ: John Wiley & Sons.

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439. doi: 10.1006/jmla.1997.2558

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, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. *Psychol. Rev.* 111, 1036–1060.

Arai, M., Van Gompel, R. P., and Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cogn. Psychol.* 54, 218–250. doi: 10.1016/j.cogpsych.2006.07.001

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

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., 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

Barr, D. J., and Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *J. Mem. Lang.* 46, 391–418. doi: 10.1006/jmla.2001.2815

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, 255–278. doi: 10.1016/j.jml.2012.11.001

Berry, K. J., Mielke, P. W. Jr., and Johnston, J. E. (2016). *Permutation Statistical Methods: An Integrated Approach*. Berlin: Springer.

Bögels, S., Barr, D. J., Garrod, S., and Kessler, K. (2015). Conversational interaction in the scanner: mentalizing during language processing as revealed by MEG. *Cereb. Cortex* 25, 3219–3234. doi: 10.1093/cercor/bhu116

Brennan, S. E., and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *J. Exp. Psychol.* 22, 1482–1493. doi: 10.1037/0278-7393.22.6.1482

Brennan, S. E., Galati, A., and Kuhlen, A. K. (2010). Two minds, one dialog: coordinating speaking and understanding. *Psychol. Learn. Motivat.* 53, 301–344. doi: 10.1016/s0079-7421(10)53008-1

Brennan, S. E., and Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Top. Cogn. Sci.* 1, 274–291. doi: 10.1111/j.1756-8765.2009.01019.x

Brennan, S. E., Kuhlen, A. K., and Charoy, J. (2018). Discourse and dialogue. Stevens'. *Handb. Exp. Psychol. Cogn. Neurosci.* 3, 1–57. doi: 10.1002/9781119170174.epcn305

Brown-Schmidt, S. (2009). 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. (2012). Beyond common and privileged: gradient representations of common ground in real-time language use. *Lang. Cogn. Proces.* 27, 62–89. doi: 10.1080/01690965.2010.543363

Brown-Schmidt, S., and Hanna, J. E. (2011). Talking in another person's shoes: incremental perspective-taking in language processing. *Dialogue Discourse* 2, 11–33. doi: 10.5087/dad.2011.102

Brown-Schmidt, S., Yoon, S. O., and Ryskin, R. A. (2015). People as contexts in conversation. *Psychol. Learn. Motivat.* 62, 59–99. doi: 10.1016/bs.plm.2014.09.003

Brown-Schmidt, S., and Duff, M. C. (2016). Memory and common ground processes in language use. *Top. Cogn. Sci.* 8, 722–736. doi: 10.1111/tops.12224

Carsey, T. M., and Harden, J. J. (2013). *Monte Carlo Simulation and Resampling Methods for Social Science*. Thousand Oaks, CA: Sage Publications.

Clark, H. H. (1996). *Using Language*. Cambridge, MA: Cambridge University Press.

Clark, H. H., and Carlson, T. B. (1981). "Context for comprehension," in *Attention and Performance IX*, eds J. Long and A. Baddeley (New Jersey: Erlbaum), 313–330.

Clark, H. H., and Marshall, C. R. (1981). "Definite reference and mutual knowledge," in *Elements of Discourse Understanding*, eds A. K. Joshe, B. L. Webber, and I. A. Sag (Cambridge, MA: Cambridge University Press), 10–61.

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

Creel, S. C., Aslin, R. N., and Tanenhaus, M. K. (2008). Heeding the voice of experience: the role of talker variation in lexical access. *Cognition* 106, 633–664. doi: 10.1016/j.cognition.2007.03.013

Elman, J. L., Hare, M., and McRae, K. (2004). *Cues, Constraints, and Competition in Sentence Processing. Beyond Nature-Nurture: Essays in Honor of Elizabeth Bates*. New Jersey: Lawrence Erlbaum Associates, 111–138.

Epley, N., Keysar, B., Van Boven, L., and Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *J. Personal. Soc. Psychol.* 87, 327–339. doi: 10.1037/0022-3514.87.3.327

Ferreira, F., and Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Lang. Linguist. Compass* 1, 71–83. doi: 10.1111/j.1749-818x.2007.00007.x

Frank, M. C., and Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science* 336, 998–998. doi: 10.1126/science.1218633

Frazier, L., and Rayner, K. (1982). Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cogn. Psychol.* 14, 178–210. doi: 10.1016/0010-0285(82)90008-1

Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2, 493–501. doi: 10.1016/s1364-6613(98)01262-5

Gigerenzer, G., Todd, P. M., and ABC Research Group (1999). *Simple Heuristics That Make us Smart*. Oxford: Oxford University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279. doi: 10.1037/0033-295x.105.2.251

Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Grice, H. P. (1957). Meaning. *Philos. Rev.* 66, 377–388.

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

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

Hendriks, P. (2016). Cognitive modeling of individual variation in reference production and comprehension. *Front. Psychol.* 7:506. doi: 10.3389/fpsyg.2016.00506

Horton, W. S., and Brennan, S. E. (2016). The role of metarepresentation in the production and resolution of referring expressions. *Front. Psychol.* 7:1111. doi: 10.3389/fpsyg.2016.01111

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

Horton, W. S., and Gerrig, R. J. (2016). Revisiting the memory-based processing approach to common ground. *Top. Cog. Sci.* 8, 780–795. doi: 10.1111/tops.12216

Horton, W. S., and Slaten, D. G. (2012). Anticipating who will say what: the influence of speaker-specific memory associations on reference resolution. *Mem. Cogn.* 40, 113–126. doi: 10.3758/s13421-011-0135-7

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

Keysar, B., Barr, D. J., Balin, J. A., and Paek, T. S. (1998). Definite reference and mutual knowledge: process models of common ground in comprehension. *J. Mem. Lang.* 39, 1–20. doi: 10.1006/jmla.1998.2563

Knoeferle, P., and Guerra, E. (2016). Visually situated language comprehension. *Lang. Linguist. Compass* 10, 66–82. doi: 10.1111/lnc3.12177

Kronmüller, E., and Barr, D. J. (2007). Perspective-free pragmatics: broken precedents and the recovery-from-preemption hypothesis. *J. Mem. Lang.* 56, 436–455. doi: 10.1016/j.jml.2006.05.002

Kronmüller, E., and Barr, D. J. (2015). Referential precedents in spoken language comprehension: a review and meta-analysis. *J. Mem. Lang.* 83, 1–19. doi: 10.1016/j.jml.2015.03.008

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

Kronmüller, E., and Noveck, I. A. (2019). How do addressees exploit conventionalizations? From a negative reference to an ad hoc implicature. *Front. Psychol.* 10:1461. doi: 10.3389/fpsyg.2019.01461

Kuhlen, A. K., and Brennan, S. E. (2013). Language in dialogue: when confederates might be hazardous to your data. *Psychon. Bull. Rev.* 20, 54–72. doi: 10.3758/s13423-012-0341-8

Matin, E., Shao, K. C., and Boff, K. R. (1993). Saccadic overhead: information-processing time with and with-out saccades. *Percept. Psychophys.* 53, 372–380.

McRae, K., and Matsuki, K. (2013). "Constraint-based models of sentence processing," in *Current Issues in the Psychology of Language. Sentence Processing*, ed. R. P. G. van Gompel (New York, NY: Psychology Press), 51–77.

McRae, K., Spivey-Knowlton, M. J., and Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *J. Mem. Lang.* 38, 283–312. doi: 10.1006/jmla.1997.2543

Metzing, C., and Brennan, S. E. (2003). When conceptual pacts are broken: partner-specific effects on the comprehension of referring expressions. *J. Mem. Lang.* 49, 201–213. doi: 10.1016/s0749-596x(03)00028-7

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Boca Raton, FL: CRC press.

Mirman, D., Dixon, J. A., and Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: growth curves and individual differences. *J. Mem. Lang.* 59, 475–494. doi: 10.1016/j.jml.2007.11.006

Ryskin, R. A., Yoon, S. O., and Brown-Schmidt, S. (2015). "Language, perspective, and memory," in *Emerging Trends in the Social and Behavioral Sciences*, eds R. Scott and S. Kosslyn (Hoboken, NJ: John Wiley & Sons, Inc).

Sperber, D., and Wilson, D. (1987). Précis of relevance: communication and cognition. *Behav. Brain Sci.* 10, 697–710.

Spivey, M. J., and Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: modeling the effects of referential context and lexical frequency. *J. Exp. Psychol.* 24, 1521. doi: 10.1037/0278-7393.24.6.1521

Strawson, P. F. (1950). On referring. *Mind* 59, 320–344.

Tesink, C. M., Petersson, K. M., Van Berkum, J. J., Van den Brink, D., and Buitelaar, J. K. (2008). Unification of speaker and meaning in language comprehension: an fMRI study. *J. Cogn. Neurosci.* 21, 2085–2099. doi: 10.1162/jocn.2008.21161

Todd, P. M., and Brighton, H. (2016). Building the theory of ecological rationality. *Minds Mach.* 26, 9–30. doi: 10.1007/s11023-015-9371-0

Trueswell, J. C., and Tanenhaus, M. K. (1994). "Toward a lexicalist framework of constraint-based syntactic ambiguity resolution," in *Perspectives on Sentence Processing*, eds C. Clifton Jr., L. Frazier, and K. Rayner (New Jersey: Lawrence Erlbaum Associates, Inc), 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

Willems, R. M., De Boer, M., De Ruiter, J. P., Noordzij, M. L., Hagoort, P., and Toni, I. (2010). A dissociation between linguistic and communicative abilities in the human brain. *Psychol. Sci.* 21, 8–14. doi: 10.1177/0956797609355563

Wu, S., Barr, D. J., Gann, T. M., and Keysar, B. (2013). How culture influences perspective taking: differences in correction, not integration. *Front. Hum. Neurosci.* 7:822. doi: 10.3389/fnhum.2013.00822

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.

The reviewer HK declared a past co-authorship with one of the authors EG.

Copyright © 2020 Kronmüller and Guerra. 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.

APPENDIX

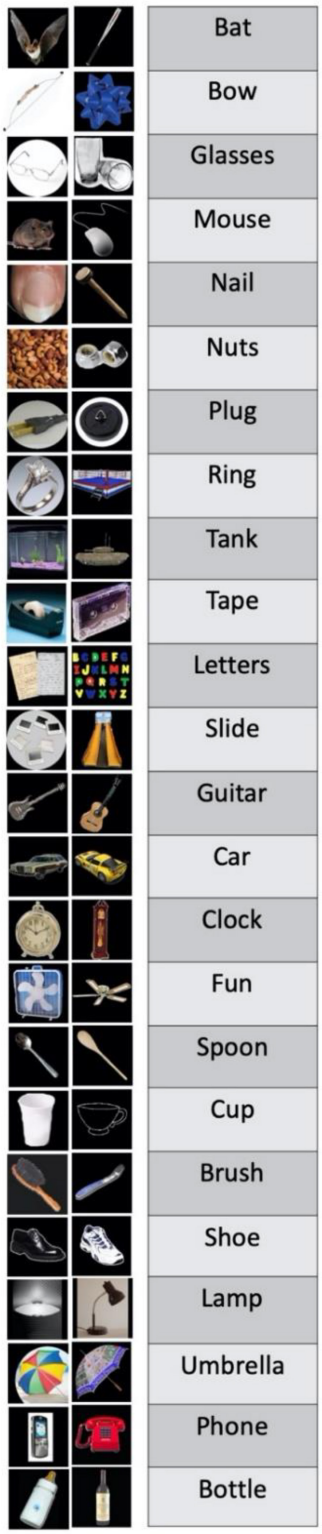


FIGURE A1 | List of referential competitors materials.

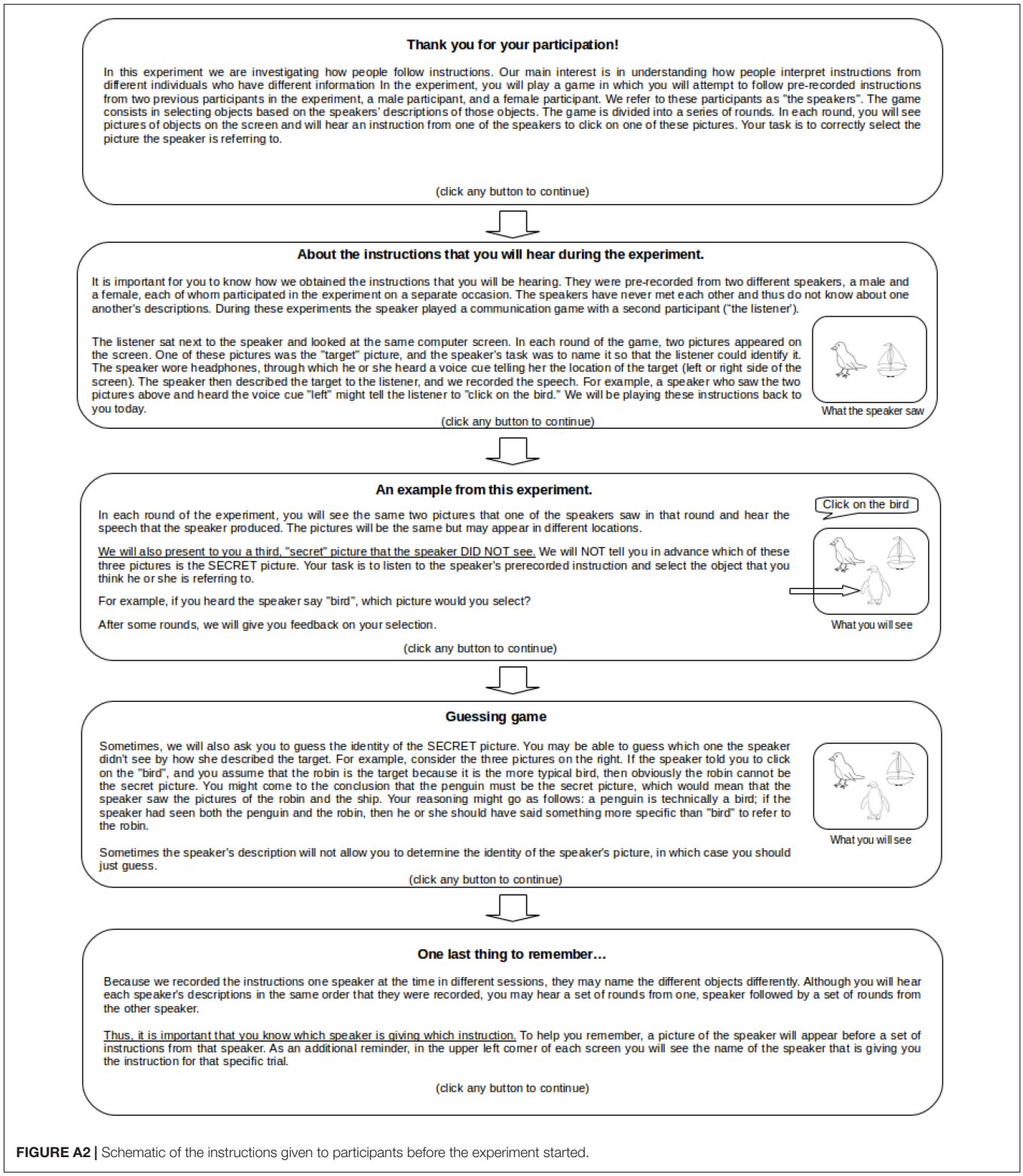


FIGURE A2 | Schematic of the instructions given to participants before the experiment started.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 January 2021
doi: 10.3389/fpsyg.2020.612601

What's New to You? Preschoolers' Partner-Specific Online Processing of Disfluency

Si On Yoon[1†], Kyong-sun Jin[2†], Sarah Brown-Schmidt[3] and Cynthia L. Fisher[4]*

[1] Department of Communication Sciences and Disorders, The University of Iowa, Iowa City, IA, United States, [2] Department of Psychology, Sungshin Women's University, Seoul, South Korea, [3] Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, United States, [4] Department of Psychology, University of Illinois, Champaign, IL, United States

Speech disfluencies (e.g., "Point to thee um turtle") can signal that a speaker is about to refer to something difficult to name. In two experiments, we found evidence that 4-year-olds, like adults, flexibly interpret a particular partner's disfluency based on their estimate of that partner's knowledge, derived from the preceding conversation. In entrainment trials, children established partner-specific shared knowledge of names for tangram pictures with one or two adult interlocutors. In each test trial, an adult named one of two visible tangrams either fluently or disfluently while children's eye-movements were monitored. We manipulated speaker knowledge in the test trials. In Experiment 1, the test-trial speaker was the same speaker from entrainment or a naïve experimenter; in Experiment 2, the test-trial speaker had been one of the child's partners in entrainment and had seen half of the tangrams (either animal or vehicle tangrams). When hearing disfluent expressions, children looked more at a tangram that was unfamiliar from the speaker's perspective; this systematic disfluency effect disappeared in Experiment 1 when the speaker was entirely naïve, and depended on each speaker's entrainment experience in Experiment 2. These findings show that 4-year-olds can keep track of two different partners' knowledge states, and use this information to determine what should be difficult for a particular partner to name, doing so efficiently enough to guide online interpretation of disfluent speech.

Keywords: disfluency, partner-specificity, common ground, pragmatic inference, eye-tracking

OPEN ACCESS

Edited by:
Carla Contemori, The University of Texas at El Paso, United States

Reviewed by:
Susan Annetta Graham, University of Calgary, Canada
Judit Bóna, Eötvös Loránd University, Hungary
Katherine White, University of Waterloo, Canada

***Correspondence:**
Kyong-sun Jin
kyongsun.jin@gmail.com

[†] These authors have contributed equally to this work

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

Received: *01 October 2020*
Accepted: *03 December 2020*
Published: *08 January 2021*

Citation:
Yoon SO, Jin K-S, Brown-Schmidt S and Fisher CL (2021) What's New to You? Preschoolers' Partner-Specific Online Processing of Disfluency. Front. Psychol. 11:612601. doi: 10.3389/fpsyg.2020.612601

INTRODUCTION

Adults understand language incrementally, integrating multiple aspects of the linguistic and non-linguistic context to assign interpretations to sentences as they unfold, and to make implicit predictions about upcoming words (Tanenhaus et al., 1995; Federmeier and Kutas, 1999; Van Berkum et al., 2005). Young children do the same, as they learn the systematic patterns that structure their native language at multiple levels of analysis (e.g., Trueswell and Gleitman, 2004). For example, both adults and toddlers are quicker to look toward a named referent if its name follows a semantically constraining verb (e.g., *eat the cake* as opposed to *move the cake*; Altmann and Kamide, 1999; Fernald et al., 2008; Mani and Huettig, 2012; Gambi et al., 2016). Similarly, both adults and toddlers use the syntactic constraints of function words to anticipate what objects are about to be named (e.g., the Spanish article *la* predicts a grammatically feminine noun, and

Where are the predicts a plural; Lew-Williams and Fernald, 2007; Lukyanenko and Fisher, 2016). Adults and children also recruit extra-linguistic information about the speaker's goals and visual perspective into online comprehension (e.g., Clark and Wilkes-Gibbs, 1986; Schober, 1993; Nadig and Sedivy, 2002; Nilsen and Graham, 2009; Yoon et al., 2012; Heller et al., 2016; Thacker et al., 2018b).

Even disfluencies in speech guide online comprehension. Disfluencies tend to occur in predictable locations in utterances, because they often reflect speaker difficulty in sentence planning or word retrieval (Ferreira, 1991; Smith and Clark, 1993; Clark and Wasow, 1998; Clark and Fox Tree, 2002; Fraundorf and Watson, 2013). As a result, listeners can use disfluencies to predict features of the upcoming speech, such as anticipating reference to something that is difficult to describe or new to the discourse (Clark and Fox Tree, 2002; Arnold et al., 2004; Arnold et al., 2007). Even 2-year-olds show this type of disfluency effect, shifting their attention to a previously unmentioned novel object when hearing *This is thee. . . uh. . .* (Kidd et al., 2011). In this paper, we explore how children use disfluency to guide online comprehension, asking whether 4-year-olds' interpretation of disfluency is influenced by their assessment of a particular speaker's knowledge state, gathered through prior conversation.

Disfluency

Disfluency is fairly common in casual speech, as speakers lengthen words, restart phrases, repeat words, or produce filled pauses (*um/uh*). Disfluency rates vary across contexts of speaking, and across speakers (e.g., Bortfeld et al., 2001; Clark and Fox Tree, 2002), and disfluencies of all kinds are less common in adult speech directed to very young children (e.g., 1 disfluency per 1000 words in speech to 2-year-olds, vs. 6 per 100 words in speech to adults; Fox Tree, 1995; Kidd et al., 2011). Disfluencies in speech to children become more frequent as children get older, presumably because adults address longer and more complex sentences to them (Kidd et al., 2011).

Two mechanisms have been proposed regarding how listeners come to interpret disfluencies; this work has focused on filled-pause disfluencies, with overt markers of disruption (*um/uh*). One proposal is that listeners interpret these disfluencies as signs of speaker difficulty, and base their expectations about what will come next on speaker- and context-specific inferences about possible causes of the difficulty [e.g., Clark and Fox Tree, 2002; Arnold et al., 2007, cf. inferences based on listeners' own difficulty (Heller et al., 2015)]. Another possibility, not contradictory to the first, is that listeners might detect the predictive value of disfluencies through distributional learning (e.g., Arnold et al., 2007; Barr and Seyfeddinipur, 2010; Kidd et al., 2011). Fillers (e.g., *um/uh.*) tend to precede reference to the discourse-new. Detecting this contingency in language experience could allow listeners to anticipate appropriate referents without the need for speaker-specific inferential processing.

Previous work yields clear evidence for inferential processing in adults' interpretation of filled-pause disfluencies (e.g., Arnold et al., 2007; Barr and Seyfeddinipur, 2010; Heller et al., 2015). For example, adult listeners expect disfluency to predict reference to a novel item that should be hard to name, but suspend this expectation if they are first informed that the speaker has a brain disorder that makes it difficult to name everyday objects (e.g., anomia; Arnold et al., 2007). Furthermore, adult listeners use their estimate of each speaker's knowledge, as established in the previous conversation, to interpret a disfluency. For example, in a task that required communication about abstract "tangram" images (Barr and Seyfeddinipur, 2010), adult listeners anticipated reference to a new rather than a previously described object when a familiar speaker became disfluent (e.g., "um. . . three blobs. . ."), but suspended this prediction when interacting with a new partner who was unfamiliar with the tangram images. Similar effects are observed in adult multiparty conversation (Yoon and Brown-Schmidt, 2014): in dialogue, listeners look toward novel objects when a familiar speaker is disfluent (e.g., "um. . . it looks like. . . the uh. . . bear"). In contrast, when a third person who is unfamiliar with the objects joins the conversation, listeners' typical expectation that disfluency predicts reference to new objects is attenuated, because the listeners attribute the speaker's disfluency to her effort to modify the referential expression to accommodate the third person's lack of shared knowledge. These findings show great flexibility in adults' interpretation of filled-pause disfluencies, suggesting a role for sophisticated inferential processing about the possible reasons for disfluency in online language interpretation.

Children's Interpretation of Disfluency

Previous evidence suggests that children as young as 2 years of age expect disfluent descriptions (e.g., "Look at thee. . . um. . .") to refer to a novel and discourse-new object as opposed to one that was familiar and discourse-old (Kidd et al., 2011; Morin-Lessard and Byers-Heinlein, 2019; Yoon and Fisher, 2020). This disfluency effect held when the two objects in view differed only in discourse status (both were familiar; Owens and Graham, 2016), but not when they differed only in novelty (neither had been previously named; Owens et al., 2018). Moreover, children's predictive use of disfluencies is not limited to a simple association between disfluency and particular types of referents (e.g., discourse-new referents). Rather, like adults, children make inferences about the possible causes of disfluencies, interpreting them flexibly based on the difficulty a particular speaker should have in naming an object (Orena and White, 2015; Thacker et al., 2018a,b). To illustrate, Orena and White (2015) introduced 3.5-year-old children to either a knowledgeable speaker who competently named everyday objects, or a forgetful speaker who often could not name ordinary objects. Children who heard the knowledgeable speaker showed the expected disfluency effect, looking preferentially toward objects that were novel and discourse-new when the speaker was disfluent (e.g., "Look! Look at thee, uhh, ..."). In contrast, children who heard the forgetful speaker did not show this pattern. This result suggests that by 3.5 years of age young children can adjust their interpretation of disfluency based on known qualities of the speaker. Because they knew this speaker had trouble naming common objects, they did not treat her disfluencies as predictors of reference to something new.

In the present work, we built on these findings to probe the flexibility of children's interpretations of disfluency. We asked for

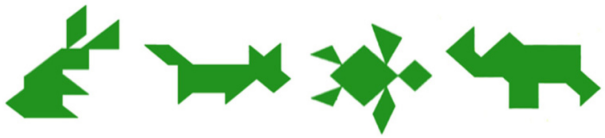


FIGURE 1 | Stimuli in Experiment 1: established labels were *bunny, dog, turtle*, and *elephant* (from left to right).

the first time whether children can track the knowledge they share with particular speakers over time, and use that assessment to interpret those speakers' disfluencies.

Establishment and Use of Common Ground

In order to make inferences about what might be difficult for a particular speaker to name, children must be able to track what the speaker does and does not know about the objects under discussion. Establishing and remembering what is in "common ground" with other people plays a fundamental part in communication, and considerable evidence suggests that young children keep track of the knowledge states of their interaction partners.

For example, in conversation, interlocutors establish common ground by developing shared labels for repeatedly mentioned entities, reflecting *conceptual pacts* for how to refer to them (Krauss and Weinheimer, 1964, 1966; Clark and Wilkes-Gibbs, 1986; Isaacs and Clark, 1987; Schober and Clark, 1989; Wilkes-Gibbs and Clark, 1992; Van der Wege, 2009; Yoon and Brown-Schmidt, 2014, 2018, 2019). To illustrate, when adult speakers describe tangram images like those in **Figure 1**, they initially produce long, elaborated descriptions (e.g., "Find the one that looks like a dog. It has a big head and you can see two legs and a tail, two ears."), but quickly shorten them through repeated use to develop a concise label for each image (e.g., "the dog"; Krauss and Weinheimer, 1964, 1966; Clark and Wilkes-Gibbs, 1986; Isaacs and Clark, 1987; Schober and Clark, 1989; Wilkes-Gibbs and Clark, 1992; Brennan and Clark, 1996; Van der Wege, 2009).

Once such a conceptual pact is established, interlocutors expect it to be honored (e.g., Metzing and Brennan, 2003). If a speaker introduces a new conceptualization of an object with an already-established description in common ground (e.g., shifting from "the shiny cylinder" to "the silver pipe" to describe the same novel object), listeners will be slower to understand this new description. Crucially, these conceptual pacts are partner-specific, reflecting their dependence on common ground between speaker and hearer. Listeners have less trouble understanding a new expression for an old object ("the silver pipe") if it is introduced by *a new speaker*, one who does not share the same conversational history (Metzing and Brennan, 2003; Brown-Schmidt, 2009; see Kronmüller and Barr, 2007, 2015 for an alternative partner-independent interpretation of the timing of these effects).

Recent reports provide evidence that children, like adults, develop conceptual pacts with their interaction partners, and show signs of expecting these pacts to be partner-specific (Matthews et al., 2010; Graham et al., 2014; Köymen et al., 2014; Branigan et al., 2016). For example, during training trials, 4-year-old children heard an experimenter use a modified noun phrase (e.g., "Look at the striped ball.") to identify a target object with two salient visual properties that could be used to identify it (e.g., a striped yellow ball; Graham et al., 2014). In later test trials, either the same experimenter or a new, naïve experimenter referred to the same object using either the original expression (e.g., "the striped ball") or a new expression ("the yellow ball"). Upon hearing the critical noun ("ball"), children were faster to identify the target object when the same speaker used the original expression compared to when the same speaker used a new expression; this advantage for the original expression disappeared if a new speaker produced the test trials. This suggests that the children expected the original speaker but not a new speaker to uphold the established conceptual pact.

Alongside these positive findings, there are suggestions in the literature that children are less reliably sensitive to partner-specific conceptual pacts than are adults. Matthews et al. (2010), for example, though reporting evidence for partner specificity in children's expectations regarding conceptual pacts, found that some children protested the use of a new term for an old object despite a partner change (e.g., "It's not a pony, it's a horse!"; see also Ostashchenko et al. (2019), for a failure to conceptually replicate Matthews et al.'s experiment). The mixed evidence on this point suggests that at least under some circumstances, children may rely on the availability of entrained labels from their own perspective, rather than successfully taking the partner's perspective in time to guide online interpretation.

A different line of evidence suggests that much younger children make speaker-specific inferences about reference, based on keeping track of common ground with their interaction partners. In one influential study (Akhtar et al., 1996), 2-year-olds played with three novel objects along with two experimenters; the objects were never named, but were referred to using neutral terms such as "this one." After the three objects had been experienced in this way, one experimenter left the room; in her absence, the other adult and the child played with a fourth object in the same manner. When the absent experimenter returned, she exclaimed excitedly "Look, I see a blicket"; all four objects were visible. Later comprehension tests suggested that children linked the new word with the fourth object, the one that was new to the speaker when she returned to the room. Tomasello and Haberl (2003) found a related result in 12- and 18-month-old infants' pronoun interpretation ("Wow, Cool! Can you give it to me?"; "it" in this context was linked with the object that was new to the speaker upon her return). The infants in these studies seem to take a speaker's excited reference to a single object as evidence that the speaker is referring to something that is new from her perspective (though all objects are old from the child's perspective; see also Saylor and Ganea, 2007). These and many other findings support the claim that the tracking of common ground with interaction partners, knowing "what's new to you," plays a key role in early language development (e.g., Tomasello, 2008; Goodman and Frank, 2016). For our purposes, these

findings also suggest that young children possess the social cognition tools needed to estimate their partners' knowledge, and perhaps in turn to use that estimate to flexibly interpret their disfluent speech.

The Present Research

In the present research, we probed the role of partner-specific inferential processing in the interpretation of disfluency. In two experiments, we manipulated speaker knowledge and tested 4-year-olds' online processing of disfluency. In Experiment 1, the speaker was either a knowledgeable partner or a naïve partner. In Experiment 2, children interacted sequentially with two different speakers; each shared particular knowledge with the child (e.g., of either animal or vehicle tangrams). In both cases, an inferential processing account of the comprehension of disfluency would predict sensitivity to each speaker's knowledge.

EXPERIMENT 1

Method

Participants

Forty-eight 4-year-olds (47.8–59.4 months; M = 51.6; 25 girls) participated in the experiment. Another ten children were excluded because of a reported language delay (1), Autism Spectrum Disorder (1), experimental errors or technical issues (6), or because they failed to complete the task (2). All were acquiring English as their native language. Children were given a book in thanks for their participation. Each child's parent gave written informed consent, and the protocol was approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign.

Apparatus

Children sat at a table in a brightly lit room, about 45 cm away from a 20-inch widescreen monitor. An experimenter stood behind the child and to the left, approximately 90-cm from the child's chair. In the task, pairs of tangram images were presented on the computer screen (**Figure 2**). The tangram images were adapted from Yoon and Brown-Schmidt (2014). Each image was approximately 10 cm tall and 10 cm wide, and separated by 22 cm of space. A digital video camera mounted above the center of the screen recorded the child's eye-movements (at a rate of 30 frames per second). Another camera behind the child and to the right recorded the experimental session, including the displayed images. Parents were instructed to wait outside the testing room.

Materials and Procedure

A child and an experimenter (E1) performed a modified version of a referential communication task (Krauss and Weinheimer, 1966). E1 first introduced the child to a secret-card game in which the child's job was to identify the matching picture on the computer monitor based on the experimenter's descriptions. For example, E1 said: "Today, we're going to play a game together. You'll have two pictures on your screen, and I have one picture on my secret card. I will tell you what I have on my card, and you find the matching picture on your screen."

Before E1 started the game, she showed the child an example of a tangram and asked the child what the tangram image looked like. Because the tangrams were abstract images, selected not to strongly suggest a particular name, children gave various answers (e.g., a dinosaur, a key, a horse). After the child answered (e.g., "a dinosaur!"), the experimenter provided another label for the tangram that differed from child's answer, and explained that tangrams did not have a specific name (e.g., "Yes, it could look like a dinosaur! But, to me it looks more like a key. See? Different people can have different names for this picture.").

The experiment consisted of two phases – Entrainment and Test (see **Figure 2**). In the Entrainment phase, the child played the secret card game with E1. On each trial, the child saw 2 tangram images on the screen, while E1 stepped back behind the child. The experimenter described the tangram picture on her secret card, and asked the child to point to the matching picture. One tangram matched the experimenter's secret card. The experimenter described her secret card following a script, but produced these scripted descriptions as naturally as possible. The script intentionally modeled the natural accumulation of common ground in conversation, adapted from the natural language production data collected in Krauss and Weinheimer (1964, 1966), Clark and Wilkes-Gibbs (1986), Isaacs and Clark (1987), Schober and Clark (1989), Wilkes-Gibbs and Clark (1992), Van der Wege (2009), Yoon and Brown-Schmidt (2014), Yoon and Brown-Schmidt (2018, 2019). Initially, the experimenter's tangram descriptions were lengthy, and they became more succinct across trials as she repeatedly described the same images. For example, one tangram was first described as "*Point to the one that looks like a bunny. It has two ears and is sitting down, stretching its arm*," but later as "*Point to the bunny*." Across trials, the child and experimenter established entrained labels for four different tangrams that were referred to three times each (**Figure 2**). Children rarely made errors but on the few trials in which they pointed to the wrong image, the experimenter gave the child another chance (e.g., "Do you think so? Let's try it again. Point to the bunny."). Descriptions in the entrainment phase included no disfluencies.

After the Entrainment phase, E1 said "Oops, I forgot to do something today. Can I go outside and check my calendar really quick?" She then left the room briefly and checked a calendar in the waiting room (leaving the door to the experiment room open while she did so). In both speaker conditions, E1 returned and told the child that she had forgotten to pick up a package. In the same-speaker condition, E1 asked another person to pick up the package for her, so that E1 could continue the game [e.g., E1: "Oh, wait, (E2's name), can you help me? Can you go upstairs and pick up a package for me? I have to finish the game with (child's name)."]. In contrast, in the different-speaker condition, E1 left the lab, after asking a new experimenter (E2) to take her place [e.g., E1: "I'll be back in 10 min, but I think (E2's name) can play the game with you. (E2's name), can you play the game with (child's name)?"]. The new experimenter (E2) agreed, but made clear that she was ignorant of the game (E2: "Sure! I don't know how to play this game but I'll do my best."). Speaker condition (Same vs. Different) was manipulated between subjects. The experimenters who played the roles of E1 and E2

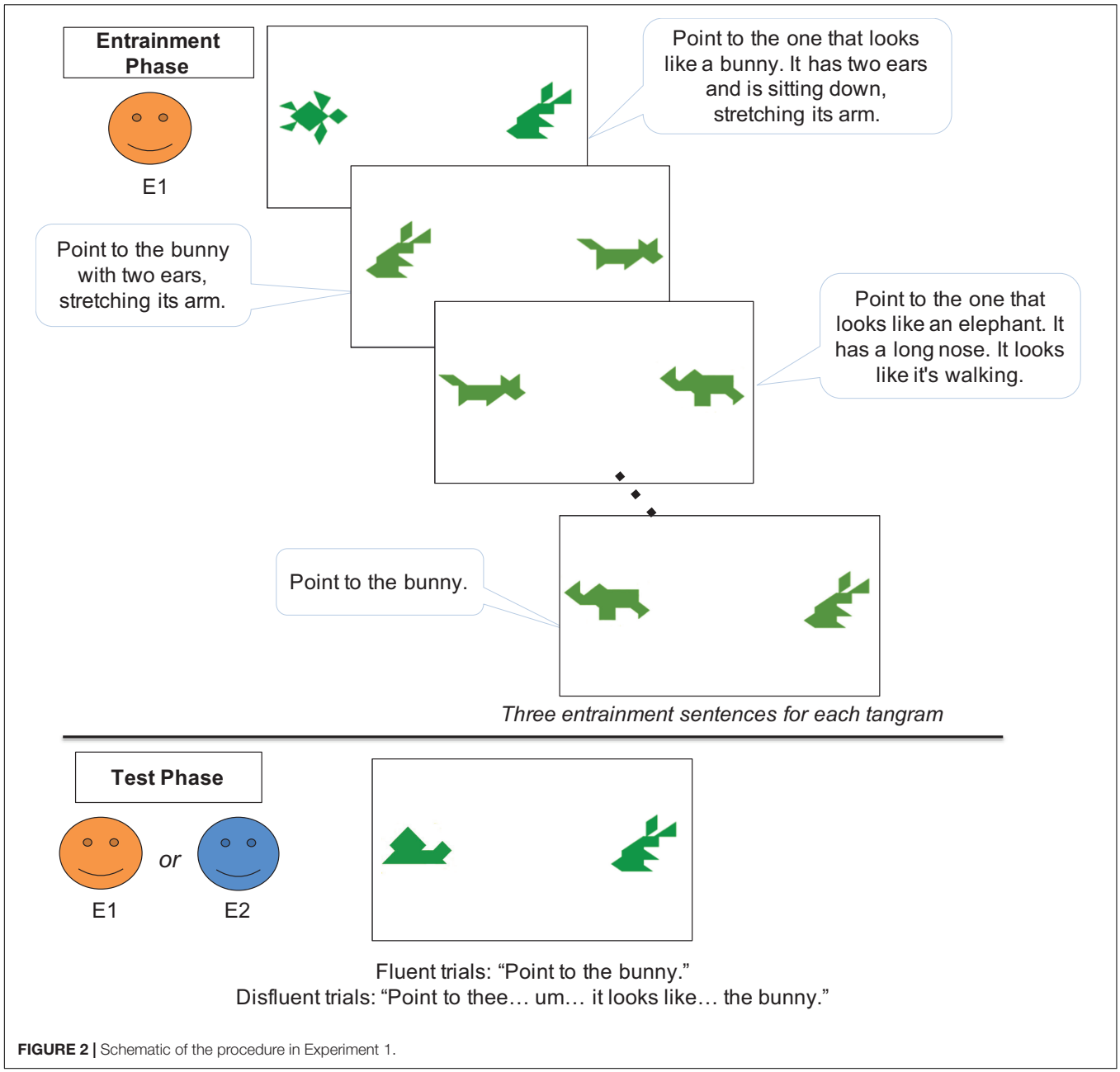


FIGURE 2 | Schematic of the procedure in Experiment 1.

were either one male and one female or two females. When both experimenters were female, they wore different vividly colored T-shirts to make them more distinctive.

In the Test phase, the task remained the same–the child pointed to the picture that matched the experimenter's description of the picture on her secret card. On each test trial, there were again 2 tangrams on the screen (**Figure 2**). One had previously been referred to during the Entrainment phase, and the other was a tangram that the child had not seen before (thus both novel and discourse-new). All 4 critical target trials referred to the old, familiar tangram. The experimenter gave fluent instructions (e.g., "Point to the bunny.") for half of the trials and disfluent instructions (e.g., "Point to thee. . . um. . .it looks like. . .the bunny.") for the other half of the trials. The disfluent instructions were modeled after naturally produced expressions by adult participants in a prior study (Yoon and Brown-Schmidt, 2014). The experimenters were trained to produce the disfluency over a period of about 3 seconds, with each part of the disfluency ("Point to thee/um/it looks like. . .") lasting approximately 1 second. The fluency of the instructions was manipulated within subjects. Regardless of fluency, the experimenter produced the same label established in the Entrainment phase (e.g., bunny). We also included 2 filler trials that described the novel and discourse-new tangram, to prevent children from ignoring novel tangrams. The experimenter gave instructions disfluently on filler trials. The location of the target

object on the screen (left vs. right) was counterbalanced. During the test trials, children were not given feedback on their accuracy, and they rarely made errors.

Predictions

Based on previous findings with 4-year-old children and young adults (Metzing and Brennan, 2003; Graham et al., 2014), we expected that children would easily interpret fluent expressions that were previously entrained, regardless of the identity of the speaker. The key question concerned how children would process a disfluency with respect to the current partner's knowledge state. If 4-year-old children are sensitive to partner-specific common ground and use this information while processing a partner's disfluency, they should interpret *the same speaker's* disfluency as indicating that the upcoming referent is novel or difficult (Arnold et al., 2004, 2007; Kidd et al., 2011). Thus, children should look at the novel/discourse-new tangram more than the familiar tangram in the same-speaker condition. The timing of this effect, specifically whether it emerges before the critical noun, would speak to the speed of this process. By contrast, in the different-speaker condition, the new experimenter is unaware of the entrained labels; thus, when the new experimenter is disfluent, the expectation that disfluency predicts a novel referent should be eliminated. If so, children's gaze should not differ between the familiar and novel tangrams following the naïve experimenter's disfluency.

Alternatively, if the child interprets the expressions from their own perspective, rather than that of the speaker, disfluent expressions should prompt them to look at the novel/discourse-new tangram more than the familiar one, regardless of the identity of the speaker.

Coding

As the critical instructions were produced live, we first marked the onset of each instruction (e.g., "Point to. . .") and of the critical noun (e.g., "bunny") to the nearest 33 ms video frame, using audio playback and a visual display of the audio waveform in Apple iMovie. The critical noun (e.g., bunny) was produced, on average, 614 ms after the onset of "Point" in fluent trials and 2,943 ms after the onset of "Point" in disfluent trials. There were no significant differences in the latency from the onset of "Point" to the onset of the critical noun across speaker conditions in either the fluent [Same Speaker: Mean = 621.5 ms (SD = 205.3), Different Speaker: Mean = 606.9 ms (SD = 118.5); $t = -0.43$, $p = 0.67$] or disfluent trials [Same Speaker: Mean = 3020.1 ms (SD = 524.3), Different Speaker: mean = 2866.7 ms (SD = 423.5; $t = -1.58$, $p = 0.12$)].

We coded children's eye-movements during Test trials from 1 second before the onset of "Point" to 3 seconds after the onset of the critical noun. Children's eye fixations (left, right, away, missing) were coded frame-by-frame (33 ms per frame) in iMovie. To ensure that coders were blind to condition, the coders viewed the video without sound, after the onset of each instruction had been marked in the video. If the child's eyes were not visible, the frame was coded as "missing." When visual fixations were coded as missing or away in more than two thirds of video frames within a critical time window defined for analysis

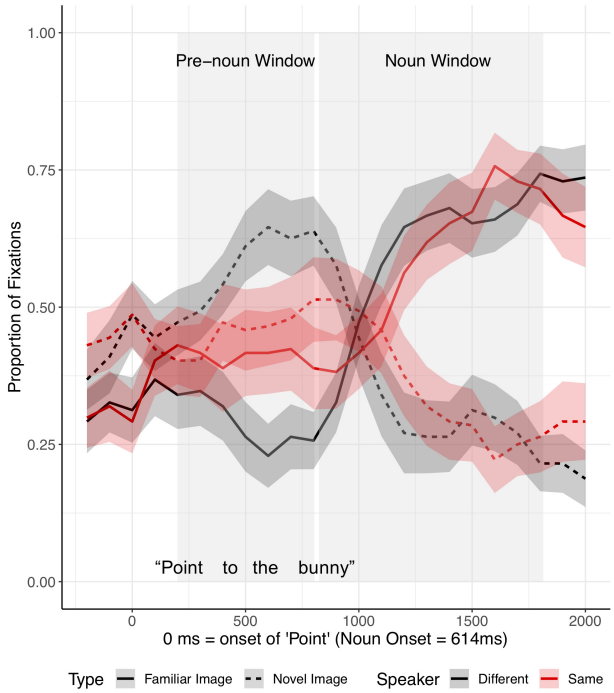


FIGURE 3 | Proportion of fixations following the onset of "Point," by Image Type (familiar vs. novel) and Speaker condition (different vs. same), for fluent trials in Experiment 1. Ribbons around the lines indicate standard errors. See **Supplementary Appendix Figures 1B**, **2B** in **Supplementary Appendix B** for figures split by speaker condition.

(see Results for more details), that time window was excluded from analysis (8.3% of time windows). Reliability was assessed by a second coder for a randomly chosen 20% of the participants. The first and second coders agreed on the children's direction of gaze for 95% of coded video frames. When the two coders did not agree, the first coder's decision was retained.

Results

Test trials were separately analyzed for fluent and disfluent descriptions, as the time before the critical noun was produced was significantly longer in the disfluent condition than the fluent condition (see Yoon and Brown-Schmidt, 2014).

Fluent Expressions

Our primary analysis focused on children's eye movements after the onset of "Point" (**Figure 3**) in two time-windows: (1) a pre-noun window extending from 200 ms after the onset of "Point" to 200 ms after critical noun onset, and (2) a noun window extending from 200 to 1200 ms after critical noun onset. The pre-noun window reflected interpretation of the description prior to the critical noun, and the noun window captured the processing of the critical noun. Both windows were offset by 200 ms, to reflect the time it takes to program and launch an eye movement (Hallet, 1986).

Target advantage scores were calculated as the empirical logit for the ratio of target fixations (fixations to familiar tangrams)

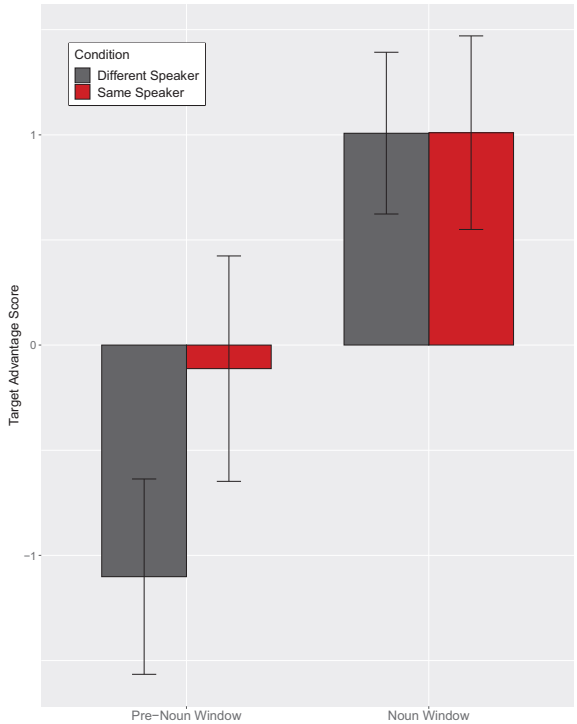


FIGURE 4 | Target advantage scores for fluent trials, calculated as the log of the ratio of target to competitor fixations, in Experiment 1. Error bars indicate standard errors.

to distractor fixations (fixations to novel tangrams) and used as the dependent measure (**Figure 4**). A target advantage score of zero indicates no preference between target and distractor, while positive values indicate a target preference and negative values indicate a distractor preference. In the pre-noun window, children looked at the distractor (novel image) more than the target image (familiar image), with this novelty preference appearing somewhat larger in the different-speaker than in the same-speaker condition. This could suggest that they expected the same speaker to continue to refer to familiar images. In the noun window, children correctly identified the target upon hearing the critical noun.

We analyzed the data in a mixed effects model with a Gaussian link function with subjects and items as random intercepts. The model included speaker identity (same vs. different speaker) and time window (pre-noun vs. noun window) as fixed effects, which were coded with mean-centered Helmert contrast codes (see **Supplementary Material** for full model details; **Supplementary Appendix Table 1**). The model revealed only a main effect of time window as target fixations increased over time within the trial ($t = -4.22$, $p = 0.004$). The identity of the speaker did not affect interpretation of fluent expressions ($t = -0.003$, $p = 1.00$).

Separate planned analyses of each window were performed, although the interaction between speaker identity and time window was not significant ($t = 1.18$, $p = 0.27$). In the pre-noun window, children looked numerically more at the novel tangram (the distractor) in the different-speaker than in the same-speaker condition; this effect was not significant ($t = -1.69$, $p = 0.10$, see **Figure 4**). Looking patterns in the noun time-window did not differ across speaker conditions ($t = 0.02$, $p = 0.98$). These results show that children interpreted fluent expressions without delay in both conditions.

Disfluent Expressions

As in the analysis of fluent expressions, we analyzed the children's eye movements from the onset of "Point." Critical noun onset occurred, on average, 2,943 ms after the onset of "Point" (**Figure 5**). Since the time between the onset of "Point" and the onset of the critical noun was longer for disfluent instructions than fluent instructions, we analyzed the eye movements in three time-windows: (1–2) two equally divided pre-noun windows extended from 200 ms after the onset of "Point" to 200 ms after noun onset (pre-noun window 1: 200–1,671 ms after the onset of "Point" vs. pre-noun window 2: 1,671–3,143 ms after the onset of "Point"), and (3) the noun window extended from 200 to 1200 ms after noun onset. The two pre-noun windows captured how children interpreted the disfluency, and the noun window reflected the processing of the critical noun phrase.

Target advantage scores were calculated as before (**Figure 6**). Before hearing the critical noun, children looked more at the (novel) distractor than at the target image in both speaker conditions. In the noun window, children in the different-speaker condition showed a stronger target advantage than did children in the same-speaker condition; thus, children identified the target image faster following a disfluency when they interacted with a different speaker as opposed to with the same speaker who participated in entrainment.

Target advantage scores were analyzed as before (**Supplementary Appendix Table 2**). Time window was coded with two Helmert contrasts. The first contrast (Window 1) tested the difference between the two pre-noun windows and the noun window, and the second contrast (Window 2) tested the difference between the first and the second pre-noun windows. The omnibus model revealed a significant main effect of Window 1, indicating increased target fixation across time within the trial ($t = -9.89$, $p < 0.0001$). Neither the main effect of speaker identity ($t = -1.35$, $p = 0.18$) nor the effect of Window 2 ($t = 0.42$, $p = 0.67$) was significant. The interaction between speaker identity and Window 1 was marginally significant ($t = 1.76$, $p = 0.08$), but the interaction between speaker identity and Window 2 was not significant ($t = -0.76$, $p = 0.45$).

Separate planned analyses of each window were performed, although the interaction between speaker identity and Window 1 was only marginally significant. During the two pre-noun windows, there was no evidence that the identity of the speaker influenced children's processing of a disfluency (pre-noun window 1: $t = -0.83$, $p = 0.41$; pre-noun window 2: $t = 0.22$, $p = 0.83$, see **Figure 6**). Children tended to look at the novel tangram more than the familiar tangram, regardless of speaker, before hearing the critical noun. A significant effect of speaker identity emerged in the noun window, such that children looked significantly less at the target (familiar) tangram in the same-speaker than in the different-speaker condition ($t = -2.23$, $p = 0.03$).

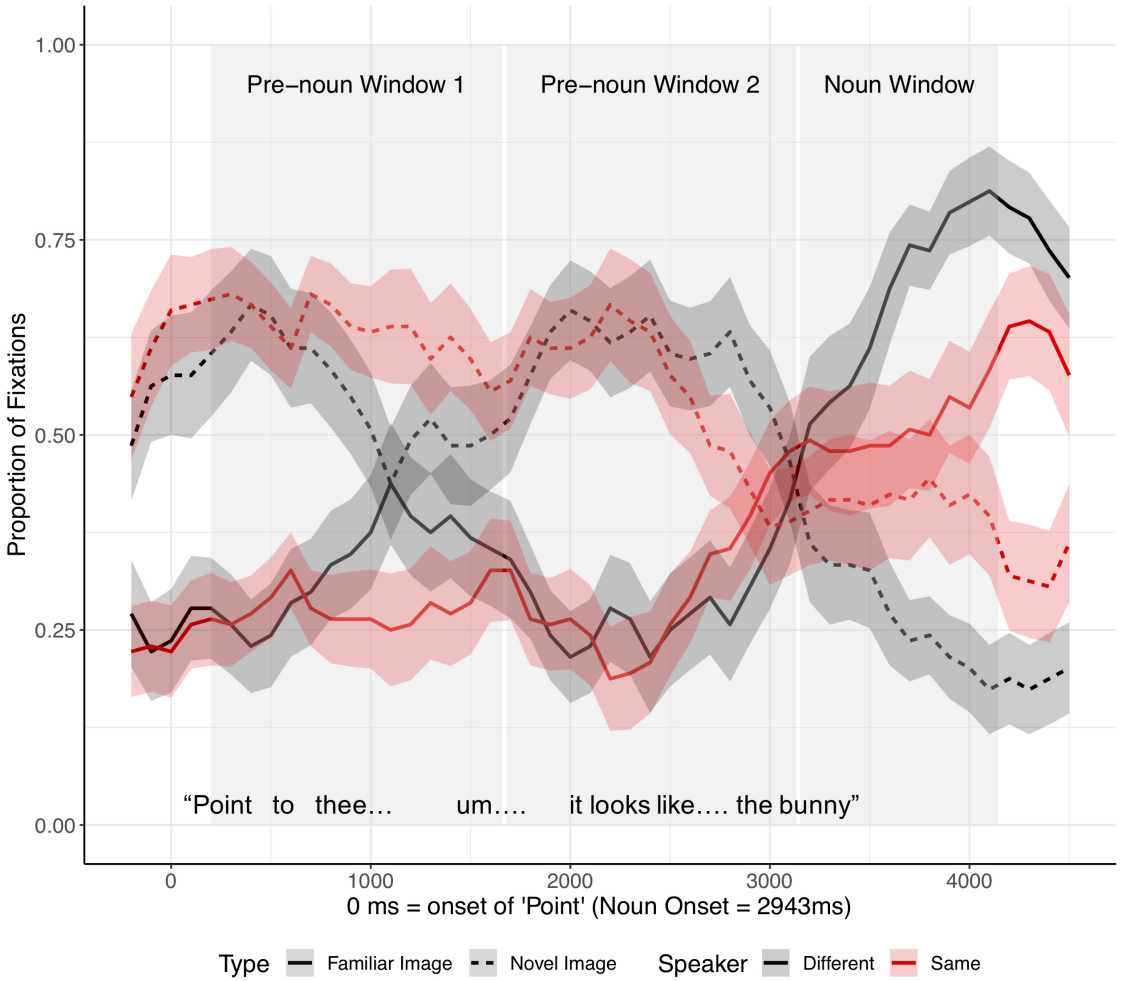


FIGURE 5 | Proportion of fixations following the onset of "Point," by Image type (familiar vs. novel image) and Speaker condition (different vs. same), for disfluent trials in Experiment 1. The target was always an old, familiar tangram that participants had seen during Entrainment, while the distractor was always a novel tangram. Ribbons around the lines indicate standard errors. See **Supplementary Appendix Figures 3B**, **4B** in **Supplementary Appendix B** for figures split by speaker condition.

This result suggests flexibility in children's online interpretation of disfluency. When interacting with their original partner, children expected disfluent descriptions to refer to new referents, and had to overcome this (violated) prediction when they heard the critical noun (e.g., following the disfluency, the noun "bunny" referred to the familiar image); they were slower or less likely to switch their gaze from the novel referent to the familiar referent. In contrast, children did not expect the new speaker's disfluencies to refer to novel referents, suggesting that they attributed the disfluency to the new speaker's lack of familiarity with both images. In conversation, children consider what is in common ground *with a particular speaker*, and use that knowledge to interpret what speakers say and how they say it.

Discussion

The results of Experiment 1 provide preliminary evidence that children are sensitive to common ground held with the current partner, and interpret disfluency with respect to the current speaker's knowledge. In the disfluent trials, children identified the target more readily when the new experimenter was disfluent compared to when the same experimenter was disfluent. This finding suggests that children's referential processing was disrupted when the familiar speaker referred disfluently to a familiar referent; in contrast, the new speaker's disfluent speech could be attributed to the speaker's lack of knowledge.

This partner effect, however, emerged in the noun window, a delay compared to findings from previous studies with young adults, for whom the comparable effect emerged prior to the noun (Yoon and Brown-Schmidt, 2014). Similarly, prior work with 2- to 4-year old children also revealed sensitivity to disfluency prior to the noun (Kidd et al., 2011; Orena and White, 2015; Yoon and Fisher, 2020). One possible explanation for the delayed effects of partner on disfluent trials is that children are less efficient at taking another's perspective in online language processing (see Epley et al., 2004). Alternatively, it may be that the children are about as efficient as adults, but

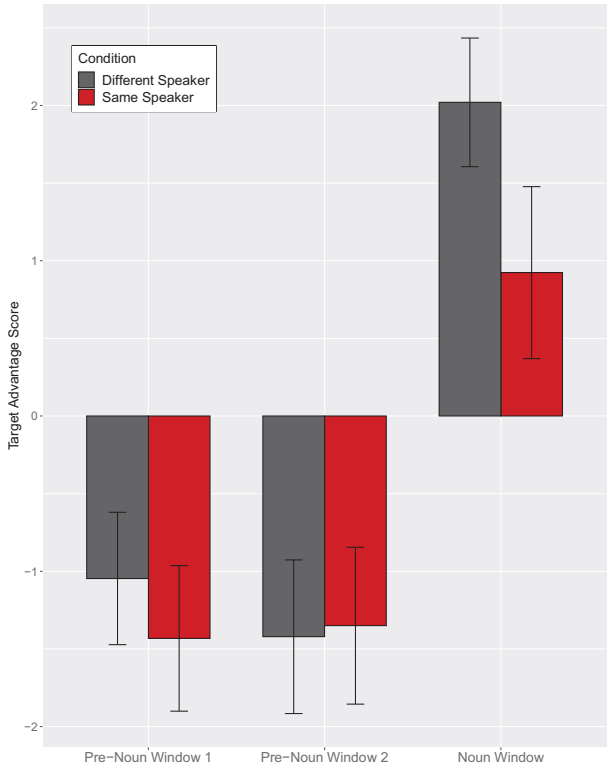


FIGURE 6 | Target advantage scores for disfluent trials, calculated as the log of the ratio of target to competitor fixations, in Experiment 1. Error bars indicate standard errors.

the presence of the novel tangram in the test trials captured their attention, delaying the emergence of the partner effect in our data. However, another alternative explanation of the children's apparent partner-sensitivity in Experiment 1 is that the "disfluency = new" association was attenuated in the different-speaker condition because the introduction of the new partner signaled an abrupt context change, potentially slowing memorial access to which items were old or new from the child's own perspective (see Smith and Vela, 2001; Brown-Schmidt et al., 2015). One way to address these alternative interpretations is to create experimental situations in which we can test for the partner specificity of interpretation while holding constant experimenter and tangram familiarity. This is the aim of Experiment 2.

EXPERIMENT 2

In Experiment 1, children successfully learned image labels with a partner in conversation and used this partner-specific information when they processed disfluent speech, although the key interaction of speaker and time-window was marginal. In Experiment 2, we provided a stronger test of partner-specificity by asking if children could develop distinct representations of common ground held with different partners and use these representations to guide online language processing in a partner-specific manner. In Experiment 2, in the test trials, both of the tangrams were familiar to the child, but only one of them was familiar to the speaker. We accomplished this by establishing names in entrainment trials for two sets of tangrams: animal tangrams were entrained with one experimenter, and vehicle tangrams were entrained with the other experimenter. At Test, the children viewed one animal and one vehicle tangram on each trial. Critically, while both tangrams were familiar from the child's perspective, only one of the tangrams was familiar to the experimenter. We tested the children's on-line processing of disfluency to examine if they interpreted the speaker's disfluency based on the current speaker's knowledge state, rather than on their own knowledge.

Method

Participants

Forty-eight 4-year-olds (48.0–59.2 months; M = 52.0; 25 girls) participated in the experiment. Another eleven children were excluded because of a reported autism spectrum diagnosis (1), cerebral palsy (1), experimental errors or technical issues (3), insufficient eye movement data[1] (3), or because they failed to complete the task (3). All were acquiring English as their native language. None of the children in Experiment 2 had participated in Experiment 1.

Materials and Procedure

The materials and procedure of Experiment 2 were similar to those of Experiment 1 except for the following changes: Experiment 2 consisted of three phases rather than two – two separate Entrainment phases followed by a Test phase (see **Figure 7**). In each of the two entrainment phases, the child interacted with one of two experimenters (first E1 then E2). Each experimenter introduced a set of novel tangrams that belonged to either an animal category or a vehicle (things-that-go) category. Thus across the two entrainment phases, one experimenter was associated with animal tangrams and the other was associated with vehicle tangrams. The tangram images from the two categories also differed in color (animal tangrams were orange and vehicles were blue), in order to better distinguish them (see Yoon et al., 2019 for a similar technique). At Test, one of the experimenters (either E1 or E2) continued to play the card game with the child, in which the child was shown an animal tangram and a vehicle tangram that they had previously seen during entrainment. Finally, we increased the number of entrainment trials within each phase (from 12 to 16) to help children establish partner-specific common ground with the two individuals. We also increased the number of test trials (from 4 to 8) to increase statistical power in Experiment 2. As in Experiment 1, the Experimenters were either one male and one female, or two females. When the Experimenters were two females, in order to better distinguish them, they wore two different vividly colored T-shirts.

In the first entrainment phase, E1 described the picture on her secret card and the child pointed to the matching

[1]If more than 25% of the time windows for a particular child were excluded, the participant was dropped from the analysis.

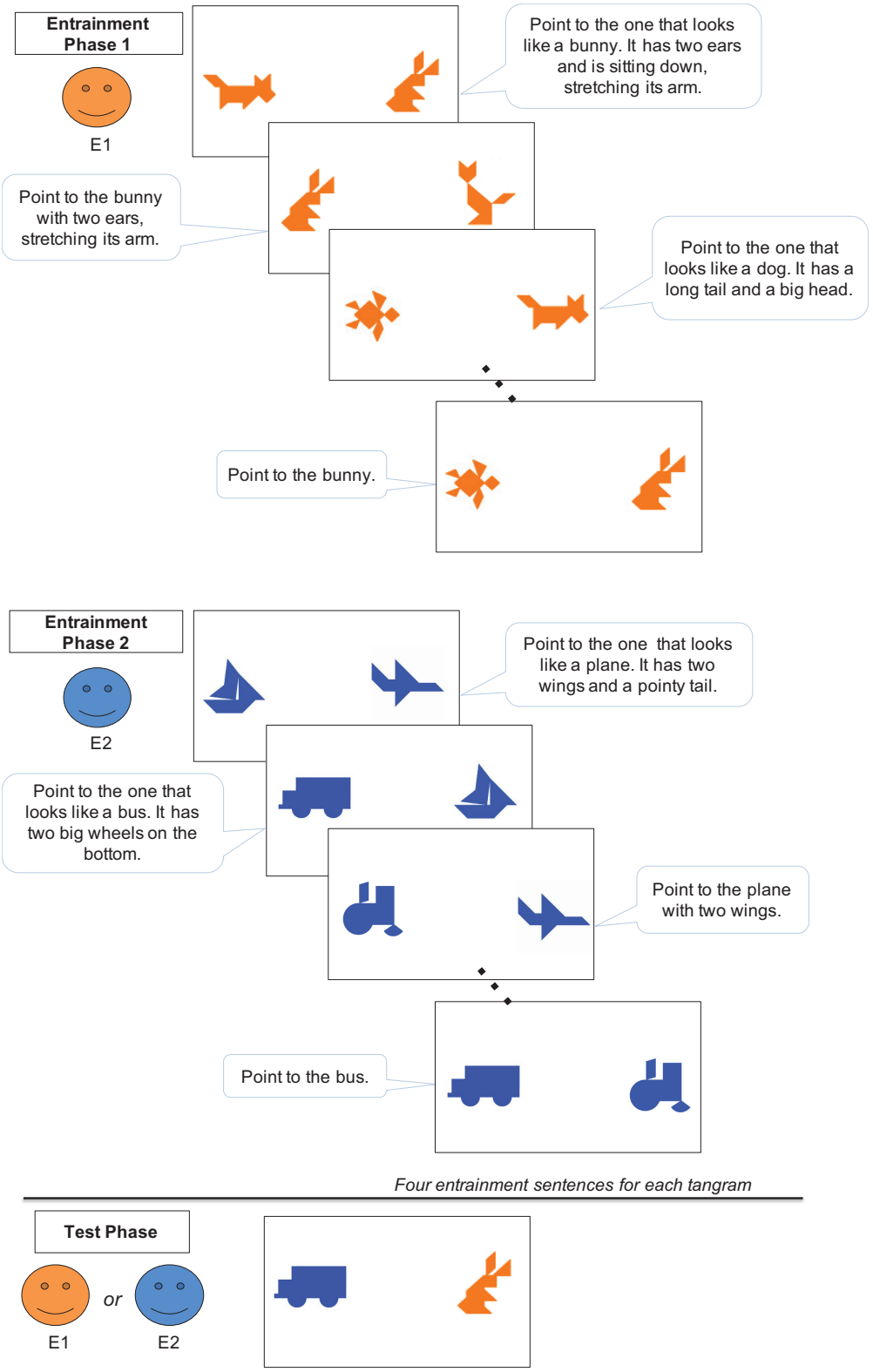


FIGURE 7 | Schematic of the procedure in Experiment 2.

picture on the screen. The child was shown tangrams of one category, either animals or vehicles (which category came first was counterbalanced across children). As in Experiment 1, the experimenter's descriptions shortened across trials, establishing a single-word label for each tangram by the last trial (e.g., "bunny"). There were four tangrams in each category that were repeated four times each; thus each child completed a total of 16 trials in each Entrainment phase. After the first Entrainment phase, E1 left the room [e.g., E1: "That's it for me. It was really cool! I'll switch with (E2's name)."] and E2 entered the room to continue the game. The entrainment trials with E2 were identical, except that E2 now described four tangrams from the other category.

Following the two Entrainment phases, either E1 or E2 performed the Test phase (saying "It's my turn again! Let's finish the game."). Before the test started, the experimenter reminded the child that she knew the tangrams from one category but not the tangrams from the other category [e.g., "Now, I have both orange animals and blue things-that-go. I had played the orange animals (or blue things-that-go) with you, so I know them very well. However, I don't know what the blue things-that-go (or orange animals) are. I'll try my best."]. During the Test, the child viewed one animal and one vehicle tangram on the screen. The critical within-subjects manipulations at test were the experimenter's fluency (fluent vs. disfluent) and the experimenter's familiarity with the target (familiar vs. unfamiliar to the speaker). Note that from the child's perspective, both of the tangrams were familiar, whereas from the experimenter's perspective, only one of the tangrams was familiar. There were a total of 8 Test trials; half of the trials were fluent and half were disfluent. The order of the two Entrainment phases (animal vs. vehicle tangrams), the identity of the speaker during Test (E1 vs. E2 from Entrainment), and the location of the target object on the screen (left vs. right) were all counterbalanced across participants.

Following the test phase, the child performed a memory test. The experimenter who did not participate in the Test showed the child all 8 tangram cards at once – 4 animal and 4 vehicle cards. The child was asked to pick out the cards that the current experimenter knew.

Predictions

Consistent with the results of Experiment 1, we expected children to readily identify the target object when the description was fluent, regardless of target type or speaker's knowledge state. In Experiment 2, we held constant the familiarity of the tangrams and asked if children could take into consideration the current speaker's perspective when interpreting disfluent expressions. If they do take the speaker's knowledge into consideration in processing her disfluency, then children should look more at the object that is unfamiliar to the current speaker when they hear a disfluent description. If so, we would expect positive target advantage scores when the target is unfamiliar to the speaker and negative target advantage scores when the target is familiar to the speaker. The timing of this effect will speak to the question of whether children's use of perspective is delayed (Epley et al., 2004).

Alternatively, if children do not consider the speaker's knowledge but rely on their egocentric knowledge when processing disfluency, we would expect an equivalent pattern of fixations, regardless of target familiarity.

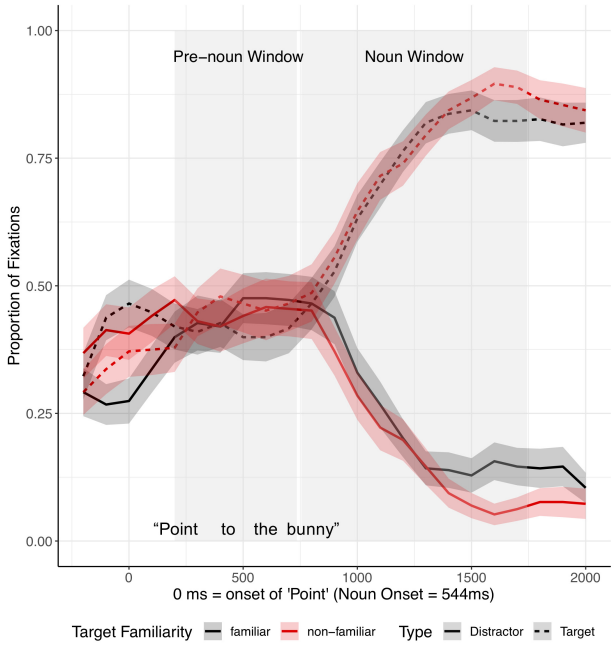


FIGURE 8 | Proportion of fixations after the onset of "Point," by Image type (distractor vs. target) and Target Familiarity (familiar vs. unfamiliar from the speaker's perspective), for fluent trials in Experiment 2. Ribbons around the lines indicate standard errors. See **Supplementary Appendix Figures 5B**, **6B** in **Supplementary Appendix B** for figures split by target familiarity.

Coding

As in Experiment 1, we first marked the onset of "Point" and of the critical noun (e.g., bunny) in each trial, as the instructions were produced live by the experimenter. The latency of the critical noun after the onset of "Point" was on average 543.58 ms in the fluent trials and 3189.24 ms in the disfluent trials (see **Figures 8**, **9**). There was no significant difference in the latency of the critical noun across conditions familiar (F) vs. unfamiliar (UF) object from the experimenter's perspective] for either fluent [F: Mean = 539. 2 ms (SD = 119.4); UF: Mean = 547.9 ms (SD = 115.8); $t = -0.51$, $p = 0.61$] or disfluent expressions [F: Mean = 3170.9 ms (SD = 690.8); UF: Mean = 3241.3 ms (SD = 633.3); $t = -0.78$, $p = 0.43$].

Children's eye-movements during the Test phase were coded manually using the same procedure as in Experiment 1. 8.0% of the time windows were excluded from analysis because more than two thirds of video frames were coded as missing or away. Reliability between the coders was high (96%, assessed for 20% of the participants). Performance on the memory test was coded as a binary measure – whether each tangram was correct or not. If the child chose four tangrams from the same category that the experimenter had previously described and none from the other category, all 8 tangrams were coded as correct and accuracy was 8/8. If the child missed one tangram from the

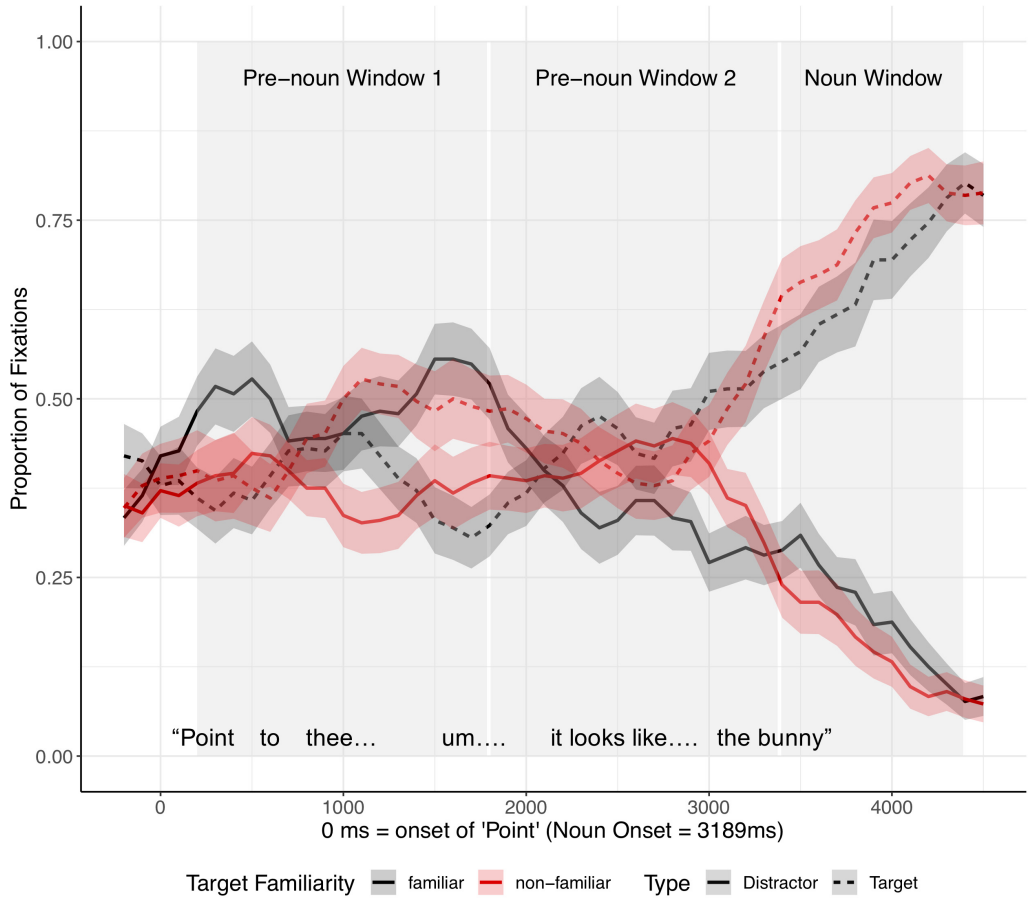


FIGURE 9 | Proportion of fixations after the onset of "Point," by Image type (distractor vs. target) and Target Familiarity (familiar vs. unfamiliar from the speaker's perspective), for disfluent trials in Experiment 2. Ribbons around the lines indicate standard errors. See **Supplementary Appendix Figures 7B**, **8B** in **Supplementary Appendix B** for figures split by target familiarity.

correct category and replaced it with another tangram from the other, incorrect category, then we coded those 2 cards as incorrect, and accuracy was 6/8.

Results

As before, we analyzed fluent and disfluent trials separately, because of the large differences in latency between disfluent and fluent trials.

Fluent Expressions

Eye movements in response to fluent expressions were analyzed in two time-windows: (1) a pre-noun window, from 200 ms after onset of "Point" to 200 ms after critical noun onset, and (2) a noun window, from 200 to 1200 ms after noun onset. As in Experiment 1, the dependent measure, target advantage score, was calculated as the empirical logit of the ratio of target fixations to competitor fixations (**Figure 10**). As **Figure 10** shows, looks to the distractor and target were approximately equal in both conditions during the pre-noun window. Upon hearing the critical noun, children looked at the target image more than the distractor image.

A mixed-effects model with a Gaussian link function included target familiarity (familiar vs. unfamiliar object from the experimenter's perspective) and time window as fixed effects (**Supplementary Appendix Table 3**). The model revealed a significant main effect of time window (t = −7.26, p < 0.0001), but the main effect of target familiarity (t = 1.36, p = 0.18) and the interaction between time window and target familiarity (t = −0.95, p = 0.35) were not significant. Planned comparisons for each time window did not show significant effects of target familiarity on target advantage scores (pre-noun window: t = −0.02, p = 0.99; noun window: t = 1.41, p = 0.17).

In sum, for fluent expressions, children showed no clear preference for either object prior to the noun. Then, upon hearing the critical noun, children quickly identified the target. This finding is consistent with the results of Experiment 1.

Disfluent Expressions

As in Experiment 1, we analyzed the eye gaze data in the disfluent trials in three time-windows (**Figure 11**): (1–2) two equally divided pre-noun windows from 200 ms after onset of "Point"

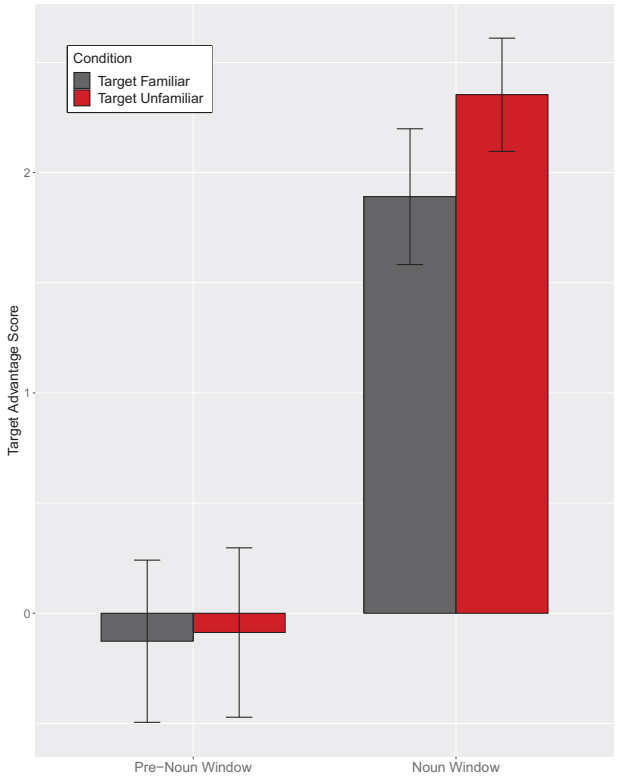


FIGURE 10 | Target advantage scores for fluent trials, calculated as the log of the ratio of target to competitor fixations, in Experiment 2. Test trials were separated based on target familiarity (whether the target was familiar vs. unfamiliar from the speaker's perspective). Error bars indicate standard errors.

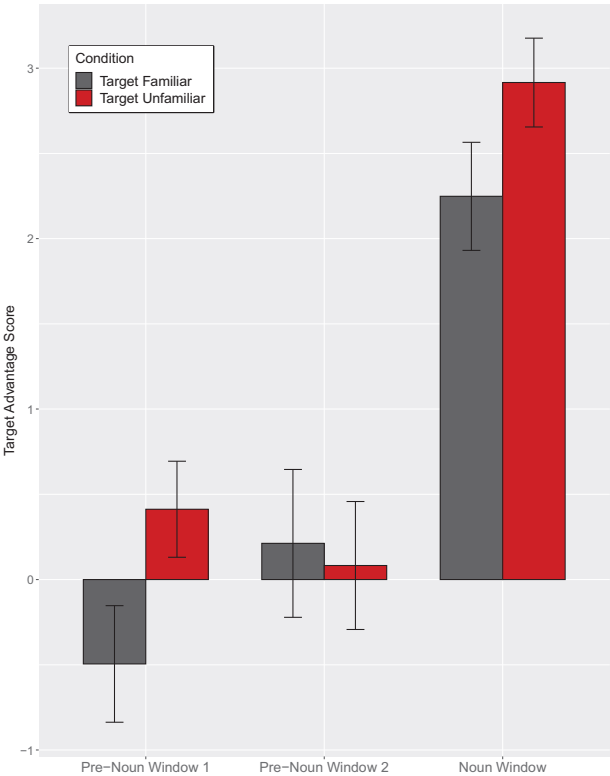


FIGURE 11 | Target advantage scores for disfluent trials, calculated as the log of the ratio of target to competitor fixations, in Experiment 2. Test trials are separated based on target familiarity (familiar vs. unfamiliar object from the speaker's perspective). Error bars indicate standard errors.

to 200 ms after noun onset, and (3) a noun window from 200 to 1200 ms after noun onset.

As **Figure 11** shows, in the first pre-noun window, target advantage scores differed across the two target familiarity conditions, indicating that in both target conditions children looked more at the image that was unfamiliar from the speaker's perspective. That is, when the target was familiar from the speaker's perspective, the target advantage score was negative, showing anticipatory looks to the distractor (which was unfamiliar to the speaker). When the target was unfamiliar to the speaker, the target advantage score was positive, showing anticipatory looks to the target. Note that children did not yet know which image was the target in the pre-noun windows. In the second pre-noun window, children looked about equally at the distractor and the target. Lastly, in the noun window, children were faster to identify the target when the target was unfamiliar as opposed to familiar from the speaker's perspective.

A mixed-effect model included target familiarity (familiar vs. unfamiliar target from the speaker's perspective) and time window as fixed effects. Time window was coded with Helmert coding. The first contrast (Window 1) tested the difference between the two pre-noun windows and the noun window and the second contrast (Window 2) tested the difference between the first and the second pre-noun window. This model revealed a significant main effect of target familiarity ($t = 2.24$, $p = 0.03$) and Window 1 ($t = -12.23$, $p < 0.001$), and a significant interaction between target familiarity and Window 2 ($t = 2.10$, $p = 0.04$) (**Supplementary Appendix Table 4**). The main effect of Window 2 ($z = -0.79$, $p = 0.43$) and the interaction between target familiarity and Window 1 ($z = -0.63$, $p = 0.53$) were not significant. Planned comparisons revealed a significant effect of target familiarity in the first pre-noun window ($t = 2.70$, $p = 0.01$), showing that children looked more at the object that was unfamiliar from the speaker's perspective from early stages of disfluency processing. This effect was not significant in the second pre-noun window ($t = -0.39$, $p = 0.70$), but it re-emerged during the noun window ($t = 2.21$, $p = 0.03$). This late emerging effect upon hearing the critical noun was consistent with the results in Experiment 1.

These results suggest that children successfully attributed the speaker's disfluency at test to that speaker's attempt to refer to an unfamiliar tangram. This effect emerged in the first half of the disfluency, disappeared during the second half of the disfluency, and then re-emerged during the noun window. The disfluencies in this study, approximately 3200 ms, were longer than disfluencies used in previous studies (e.g., ∼2 s in Kidd et al., 2011), which may have contributed to the disappearance of the effect in the second time-window.

Memory Test

Accuracy on the memory test was good overall, 6.27/8 on average (range: 0/8–8/8; Median = 8; SD = 2.63), indicating that the children successfully recalled which partner was associated with which set of tangrams.

Discussion

We found that 4-year-old children were able to learn two distinct representations of common ground with different partners, and use these representations when later encountering a partner's disfluent speech. Children attributed the disfluency to the current speaker's lack of knowledge, and therefore directed anticipatory looks toward the tangram that was unfamiliar to the speaker during the disfluency period. Because all tangrams were familiar from the children's own perspective, this result is strong evidence that children successfully took their partner's knowledge state into consideration, rather their own knowledge, in online interpretation of a disfluency. This partner-specific interpretation of disfluency as signaling reference to new information appeared early in the processing of the speaker's expressions. The fact that Experiment 2 employed a design in which both tangram categories and both Experimenters were familiar to the children may have supported their early use of this information. Children's memory performance was generally good, showing that children typically remembered what information had been shared with whom in our context.

GENERAL DISCUSSION

The results of two experiments show that 4-year-old children can successfully represent two different partners' knowledge states and use this information appropriately when interpreting disfluency in live conversation. This finding demonstrates that 4-year-old children process disfluency based on their estimate of their partner's knowledge, rather than on their egocentric knowledge.

In Experiment 1, when interpreting a disfluent expression, children tended to look at the novel and discourse-new object rather than the familiar object when they interacted with the knowledgeable speaker, but not when they interacted with the naïve speaker. This pattern must be interpreted with caution because the key interaction was marginally significant, but these signs that children's interpretation was tailored to the knowledge of the speaker show sensitivity to speaker perspective. This finding points to a role for inferential processing in the interpretation of disfluency, as opposed to simple associations between disfluency and novel/discourse-new referents. In Experiment 2, children viewed two familiar objects, only one of which was unfamiliar and discourse-new to the speaker. When interpreting disfluent expressions, children showed an early preference to gaze at the image that was unfamiliar to the speaker, much as adults do (Arnold et al., 2007; Yoon and Brown-Schmidt, 2014). Taken together, the current findings provide further evidence that children's interpretation of disfluent speech is speaker-specific, suggesting that children make rapid and sophisticated inferences about a speaker's use of disfluency based on the current speaker's knowledge state in live conversation.

The results of Experiment 2 are particularly striking. The current findings show that children can develop two distinct representations of common ground while communicating, and flexibly retrieve them to guide online processing of disfluent speech. This is consistent with recent findings demonstrating 4-year-old children's sensitivity to two speakers' differing visual perspectives during online reference resolution (Khu et al., 2020). Just as 4-year-olds consider what objects each speaker can or cannot see when interpreting her words, our findings show they also consult their memory to determine what objects are known or unknown to each speaker from prior conversation.

A possible alternative interpretation of the current findings is that the phrase "... looks like..." in disfluent test trials, independent of other features of the extended disfluency, might have reminded children of the first round of entrainment, suggesting that the speaker was naming a new referent. The extended disfluencies in our experimental scripts were intentionally modeled on naturally produced instructions in an earlier study (Yoon and Brown-Schmidt, 2014). These utterances therefore have high ecological validity, but the fact that they included multiple types of disfluency (e.g., elongated "the," filled pauses, and "looks like") makes it difficult for us to tell which type of disfluency drove the effects in our current studies. However, we can offer two considerations that help rule out this alternative interpretation. First, although the disfluent test instructions and the first instruction in the entrainment phase both included "looks like," they differed in other respects. In particular, the entrainment trials included no filled pauses or elongated definite articles (e.g., thee ... um). These markers of disfluency were first introduced during the test trials. Second, as shown in **Figures 10**, **11**, in Experiment 2 children showed their disfluency effect (looking at the image that was unfamiliar to the current speaker) in pre-noun window 1 of the extended disfluency, before they heard the phrase "looks like." This suggests that other markers of disfluency ("thee ... um ...") drove our key effects.

Several factors may have contributed to children's success in tracking and using partner-specific common ground online in our task. First of all, we used abstract tangram images that did not invite obvious names, rather than easily namable everyday objects. Tangrams are often used in studies of common ground in conversation for exactly this reason, because they lack conventional labels that any competent speaker would be expected to know (Krauss and Weinheimer, 1964, 1966; Clark and Wilkes-Gibbs, 1986; Wilkes-Gibbs and Clark, 1992; Yoon and Brown-Schmidt, 2014). As noted earlier, in a prior study of children's understanding of conceptual pacts that used everyday objects (e.g., horse/pony) as stimuli, some children complained when even a new speaker used a new label (Matthews et al., 2010). Children may have interpreted the experimenter's initial labeling of each object as information about its conventional category label. In contrast, children in our studies collaborated with the experimenters to label less codable tangram images across a series of entrainment trials. This process, and the relative difficulty of describing these abstract images, may have alerted children to the need to consult their partners' knowledge states.

Another factor that might have supported children's success in our tasks is that we provided additional cues in both experiments that might have alerted children to the need to consult partner-specific common ground. These included the new experimenter's verbal statement before the test trials in Experiment 1 that she was not familiar with the tangram images, and the category and color differences between the tangrams shared with each interlocutor in Experiment 2 (e.g., orange animals vs. blue vehicles). Access to the appropriate memory representations is required to use jointly shared knowledge in conversation; failures of memory may cause poor perspective-taking (Horton and Gerrig, 2005). In a study of audience design in language production, Horton and Gerrig (2005) showed that speakers more reliably tailored their referential expressions to the knowledge states of particular listeners when the experiences shared with the each listener were organized by category and thus easier to tell apart (e.g., one partner shared images of frogs, and another shared images of fish), as opposed to when the experiences shared with each listener were of the same kinds (both shared images of frogs and fish). This finding prompted us to provide hints such as the category and color cues in Experiment 2, to support children's memory for what knowledge should be attributed to each interaction partner.

An open question is whether the interactive paradigm that allowed the children to freely interact with the speakers played a crucial role in our current findings. Adults sometimes show stronger sensitivity to contextual cues during live interactions than in non-interactive settings (Horton and Spieler, 2007; Brown-Schmidt, 2009; Yoon and Stine-Morrow, 2019). Children may also be more susceptible to cues made available through live interaction and if so, those cues could have been particularly advantageous in Experiment 2 when they needed to establish multiple representations of common ground. Interestingly, the studies that yield evidence for infants' tracking of partner-specific common ground in offline tasks also rely on live interaction to create a memorable interaction history (e.g., Akhtar et al., 1996; Tomasello and Haberl, 2003).

Another open question is how much children benefited from the various hints we provided during the experimental interactions, and whether such supporting hints are necessary for young children to maintain and use partner-specific common ground in online comprehension. Young adults can establish and maintain multiple distinct representations even without such explicit hints to do so (see Yoon and Brown-Schmidt, 2018, 2019; Yoon and Stine-Morrow, 2019). When and how children develop the ability to routinely retain and use multiple representations of common ground remains an important open question regarding the development of partner-specific language processing. Interestingly, it may be that disfluency itself serves as a cue to the need to consult partner-specific common ground. Given the flexibility with which adult listeners interpret filled pauses, Barr and Seyfeddinipur (2010) speculated that *um*'s and *uh*'s might draw the listener's attention to the knowledge and goals of the speaker. When a speaker is disfluent, the listener must infer the cause of the disfluency in order to interpret it appropriately as a cue for how the rest of the sentence might unfold. Disfluency could reflect the abilities of the speaker (as in the anomic or forgetful speakers described in the Introduction; Arnold et al., 2007; Orena and White, 2015), distraction by a concurrent task (Yoon and Fisher, 2020), or the conversational history shared with individual interaction partners, as in the present studies.

CONCLUSION

Across two experiments, we investigated how children interpret disfluency when they interact with multiple partners who share different knowledge sets with them. We found that 4-year-old children were able to establish multiple representations of shared knowledge through live interaction, and then use this partner-specific information when processing a speaker's disfluent referring expressions. Children treated disfluency as a predictor of reference to images that were novel and discourse-new from the perspective of the speaker rather than of the child.

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 the University of Illinois at Urbana-Champaign. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors contributed to planning and carrying out the experiments, writing the manuscript, and approved the final version of the manuscript for submission.

FUNDING

This work was supported by the NIH grant R01HD054448 to CF and an Old Gold Fellowship at the University of Iowa to SY.

ACKNOWLEDGMENTS

We would like to thank the parents and children who participated in the research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.612601/full#supplementary-material

REFERENCES

Akhtar, N., Carpenter, M., and Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Dev.* 67, 635–645. doi: 10.2307/1131837

Altmann, G. T. M., 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

Arnold, J. E., Hudson Kam, C. L., and Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 914–930. doi: 10.1037/0278-7393.33.5.914

Arnold, J. E., Tanenhaus, M. K., Altmann, R., and Fagnano, M. (2004). The old and thee, uh, new. *Psychol. Sci.* 15, 578–582. doi: 10.1111/j.0956-7976.2004.00723.x

Barr, D., and Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Lang. Cogn. Process.* 25, 441–455. doi: 10.1080/01690960903047122

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., and Brennan, S. E. (2001). Disfluency rates in conversation: effects of age, relationship, topic, role, and gender. *Lang. Speech* 44, 123–147. doi: 10.1177/00238309010440020101

Branigan, H. P., Bell, J., and McLean, J. F. (2016). Do you know what I know? The impact of participant role in children's referential communication. *Front. Psychol.* 7:213. doi: 10.3389/fpsyg.2016.00213

Brennan, S. E., and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 482–493. doi: 10.1037/0278-7393.22.6.1482

Brown-Schmidt, S. (2009). 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., Yoon, S., and Ryskin, R. A. (2015). People as contexts in conversation. *Psychol. Learn. Motiv.* 62, 59–99. doi: 10.1016/bs.plm.2014.09.003

Clark, H. H., and Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition* 84, 73–111. doi: 10.1016/S0010-0277(02)00017-3

Clark, H. H., and Wasow, T. (1998). Repeating words in spontaneous speech. *Cogn. Psychol.* 37, 201–242. doi: 10.1006/cogp.1998.0693

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

Epley, N., Morewedge, C. K., and Keysar, B. (2004). 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

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

Fernald, A., Zangl, R., Portillo, A. L., and Marchman, V. (2008). "Looking while listening: using eye movements to monitor spoken language comprehension by infants and young children," in *Language Processing in Children*, eds I. Skerina, E. Fenandez, and H. Clahsen (Amsterdam: Benjamins). doi: 10.1075/lald.44.06fer

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *J. Mem. Lang.* 30, 210–233. doi: 10.1016/0749-596X(91)90004-4

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *J. Mem. Lang.* 34, 709–738. doi: 10.1006/jmla.1995.1032

Fraundorf, S. H., and Watson, D. G. (2013). Alice's adventures in um-derland: psycholinguistic sources of variation in disfluency production. *Lang. Cogn. Process.* 29, 1083–1096. doi: 10.1080/01690965.2013.832785

Gambi, C., Pickering, M. J., and Rabagliati, H. (2016). Beyond associations: sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition* 157, 340–351. doi: 10.1016/j.cognition.2016.10.003

Goodman, N. D., and Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends Cogn. Sci.* 20, 818–829. doi: 10.1016/j.tics.2016.08.005

Graham, S. A., Sedivy, J., and Khu, M. (2014). That's not what you said earlier: preschoolers expect partners to be referentially consistent. *J. Child Lang.* 41, 34–50. doi: 10.1017/S0305000912000530

Hallet, P. E. (1986). "Eye movements," in *Handbook of Perception and Human Performance*, Vol. 1, eds K. R. Boff, L. Kaufman, and J. P. Thomas (New York, NY: Wiley), 10.1–10.112.

Heller, D., Arnold, J. E., Klein, N., and Tanenhaus, M. K. (2015). Inferring difficulty: flexibility in the real-time processing of disfluency. *Lang. Speech* 58, 190–203. doi: 10.1177/0023830914528107

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

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

Horton, W. S., and Spieler, D. H. (2007). Age-related differences in communication and audience design. *Psychol. Aging* 22, 281–290. doi: 10.1037/0882-7974.22.2.281

Isaacs, E., 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

Khu, M., Chambers, C. G., and Graham, S. A. (2020). Preschoolers flexibly shift between speakers' perspective during real-time language comprehension. *Child Dev.* 91, e619–e634. doi: 10.1111/cdev.13270

Kidd, C., Katherine, S. W., and Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Dev. Sci.* 14, 925–934. doi: 10.1111/j.1467-7687.2011.01049.x

Köymen, B., Schmerse, D., Lieven, E., and Tomasello, M. (2014). Young children create partner-specific referential pacts with peers. *Dev. Psychol.* 50, 2334–2342. doi: 10.1037/a0037837

Krauss, R. M., and Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction. A preliminary study. *Psychon. Sci.* 1, 113–114. doi: 10.3758/BF03342817

Krauss, R. M., and Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *J. Pers. Soc. Psychol.* 4, 343–346. doi: 10.1037/h0023705

Kronmüller, E., and Barr, D. J. (2007). Perspective-free pragmatics: broken precedents and the recovery-from-preemption hypothesis. *J. Mem. Lang.* 56, 436–455. doi: 10.1016/j.jml.2006.05.002

Kronmüller, E., and Barr, D. J. (2015). Referential precedents in spoken language comprehension: a review and meta-analysis. *J. Mem. Lang.* 83, 1–19. doi: 10.1016/j.jml.2015.03.008

Lew-Williams, C., and Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychol. Sci.* 18, 193–198. doi: 10.1111/j.1467-9280.2007.01871.x

Lukyanenko, C., and Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition* 146, 349–370. doi: 10.1016/j.cognition.2015.10.012

Mani, N., and Huettig, F. (2012). Prediction during language processing is a piece of cake – But only for skilled producers. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 843–847. doi: 10.1037/a0029284

Matthews, D., Lieven, E., and Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Dev. Psychol.* 46, 749–760. doi: 10.1037/a0019657

Metzing, C., and Brennan, S. (2003). When conceptual pacts are broken: partner-specific effects on the comprehension of referring expressions. *J. Mem. Lang.* 49, 201–213. doi: 10.1016/S0749-596X(03)00028-7

Morin-Lessard, E., and Byers-Heinlein, K. (2019). Uh and euh signal novelty for monolinguals and bilinguals: evidence from children and adults. *J. Child Lang.* 46, 522–545. doi: 10.1017/S0305000918000612

Nadig, A. S., and Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's online reference resolution. *Psychol. Sci.* 13, 329–336. doi: 10.1111/j.0956-7976.2002.00460.x

Nilsen, E., and Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cogn. Psychol.* 58, 220–249. doi: 10.1016/j.cogpsych.2008.07.002

Orena, A. J., and White, K. S. (2015). I forget what that's called! Children's online processing of disfluencies depends on speaker knowledge. *Child Dev.* 86, 1701–1709. doi: 10.1111/cdev.12421

Ostashchenko, E., Deliens, G., Geelhand, P., Bertels, J., and Kissine, M. (2019). Referential processing in 3- and 5-year-old children is egocentrically anchored.

J. Exp. Psychol. Learn. Mem. Cogn. 45, 1387–1397. doi: 10.1037/xlm0000659

Owens, S. J., and Graham, S. A. (2016). Thee, uhh disfluency effect in preschoolers: a cue to discourse status. *Br. J. Dev. Psychol.* 34, 388–401. doi: 10.1111/bjdp.12137

Owens, S. J., Thacker, J. M., and Graham, S. A. (2018). Disfluencies signal reference to novel objects for adults but not children. *J. Child Lang.* 45, 581–609. doi: 10.1017/S0305000917000368

Saylor, M. M., and Ganea, P. (2007). Infants interpret ambiguous requests for absent objects. *Dev. Psychol.* 43:696. doi: 10.1037/0012-1649.43.3.696

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition* 47, 1–24. doi: 10.1016/0010-0277(93)90060-9

Schober, M. F., and Clark, H. H. (1989). Understanding by addressees and overhearers. *Cogn. Psychol.* 21, 211–232. doi: 10.1016/0010-0285(89)90008-X

Smith, S. M., and Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychon. Bull. Rev.* 8, 203–220. doi: 10.3758/BF03196157

Smith, V. L., and Clark, H. H. (1993). On the course of answering questions. *J. Mem. Lang.* 32, 25–38. doi: 10.1006/jmla.1993.1002

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* 16, 1632–1634. doi: 10.1126/science.7777863

Thacker, M. T., Chambers, C. G., and Graham, S. A. (2018a). Five-year-olds' and adults' use of paralinguistic cues to overcome referential uncertainty. *Front. Psychol.* 9:143. doi: 10.3389/fpsyg.2018.00143

Thacker, M. T., Chambers, C. G., and Graham, S. A. (2018b). When it is apt to adapt: flexible reasoning guides children's use of talker identity and disfluency cues. *J. Exp. Child Psychol.* 167, 314–327. doi: 10.1016/j.jecp.2017.11.008

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/7551.001.0001

Tomasello, M., and Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Dev. Psychol.* 39, 906–912. doi: 10.1037/0012-1649.39.5.906

Trueswell, J., and Gleitman, L. (2004). "Children's eye movements during listening: deveopmental evidence for a constraint-based theory of sentence processing," in *The Interface of Language, Vision, and Action: Eye Movements and the Visual Word*, eds J. Henderson and F. Ferreira (New York, NY: Psychology Press), 319–346.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 443–467. doi: 10.1037/0278-7393.31.3.443

Van der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. *J. Mem. Lang.* 60, 448–463. doi: 10.1016/j.jml.2008.12.003

Wilkes-Gibbs, D., and Clark, H. H. (1992). Coordinating beliefs in conversation. *J. Mem. Lang.* 31, 183–194. doi: 10.1016/0749-596X(92)90010-U

Yoon, S., and Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 919–937. doi: 10.1037/a0036161

Yoon, S., and Brown-Schmidt, S. (2018). Aim low: mechanisms of audience design in multiparty conversation. *Discourse Process.* 55, 566–592. doi: 10.1080/0163853X.2017.1286225

Yoon, S., and Brown-Schmidt, S. (2019). Audience design in multiparty conversation. *Cogn. Sci.* 43:e12774. doi: 10.1111/cogs.12774

Yoon, S., and Fisher, C. (2020). "Children's attribution of disfluency to different sources," in *Proceedings of the 42nd Annual Virtual Meeting of the Cognitive Science Society*, Toronto, ON.

Yoon, S., Koh, S., and Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychon. Bull. Rev.* 19, 699–707. doi: 10.3758/s13423-012-0262-6

Yoon, S., and Stine-Morrow, E. A. L. (2019). Evidence of preserved audience design with aging in interactive conversation. *Psychol. Aging* 34, 613–623. doi: 10.1037/pag0000341

Yoon, S. O., Brown-Schmidt, S., and Duff, M. C. (2019). "When is declarative memory necessary for audience design? Evidence from amnesia," in *Proceddings of the Spoken Presentation at the 32nd Annual CUNY Conference on Human Sentence Processing*, Boulder, CO.

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

frontiers in Communication

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fcomm.2020.606616

Free Indirect Discourse in Non-fiction

Andreas Stokke [1,2]*

[1] Department of Philosophy, Uppsala University, Uppsala, Sweden, [2] Swedish Collegium for Advanced Study, Uppsala, Sweden

This paper considers some uses of Free Indirect Discourse within non-fictional discourse. It is shown that these differ from ordinary uses in that they do not attribute actual thoughts or utterances. I argue that the explanation for this is that these uses of Free Indirect Discourse are not assertoric. Instead, it is argued here that they are fictional uses, that is, they are used with fictional force like utterances used to tell a fictional story. Rather than making assertions about the actual world, these occurrences of Free Indirect Discourse introduce localized fictional scenarios from which audiences are meant to learn factual information. As such, they exhibit some of the ways in which the involvement of perspective in historical fiction has been shown to facilitate learning and retention of information.

Keywords: free indirect discourse (FID), perspective, non-fiction, narrative, direct discourse, fiction

1. INTRODUCTION

Ways of reporting speech and thought have long been studied in both philosophy and linguistics. Familiar ways of reporting are Direct Discourse, Indirect Discourse, and Free Indirect Discourse, as in (1)–(3).

(1) **Direct Discourse** (DD)
Ellen made a decision. "Yes! I will tell him later today," she said/thought.

(2) **Indirect Discourse** (ID)
Ellen made a decision. She said/thought that she would tell him later that day.

(3) **Free Indirect Discourse** (FID)
Ellen made a decision. Yes! She would tell him later today(, she said/thought).

Whereas it is clear that DD and ID occur in both fictional and non-fictional discourse, FID has traditionally been associated with fictional discourse, or at least seen as a "literary style[1]."

This paper examines some uses of FID in non-fictional discourse. Three examples are given in (4)–(6).

(4) [Discussing the ancient Roman historian Livy's treatment of the Romulus and Remus myth:] Could it be that a local whore rather than a local wild beast had found and tended the twins? Whatever the identity of the *lupa*, a kindly herdsman or shepherd soon found the boys and took them in. Was his wife the prostitute? Livy wondered (Beard, 2015, p. 59).

(5) [Discussing Julius Caesar's motivations for his invasion of Britain:] Sitting in sunlit Rome at the height of his powers, a little giddy with invincibility, Caesar must have imagined a nice little sideshow, a triumph on the cheap. Faced with the glittering armour of the legions and the eagle standards, the barbarians would simply line up to surrender. They would understand that history always fought on the side of Rome (Schama, 2000, p. 29).

[1] See e.g., (Banfield, 1982; Doron, 1991; Schlenker, 2004; Currie, 2010; Eckardt, 2015; Maier, 2015).

OPEN ACCESS

Edited by:
Sonja Zeman,
Ludwig Maximilians Universität München, Germany

Reviewed by:
Anne Colette Reboul,
Université Claude Bernard Lyon 1, France
Stefan Hinterwimmer,
University of Wuppertal, Germany

***Correspondence:**
Andreas Stokke
andreas.stokke@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *15 September 2020*
Accepted: *08 December 2020*
Published: *12 January 2021*

Citation:
Stokke A (2021) Free Indirect Discourse in Non-fiction. Front. Commun. 5:606616. doi: 10.3389/fcomm.2020.606616

(6) [Describing Darwin's reaction to Wallace's discovery of the principle of evolution:] He racked his brain to recall whether or not he had written something in a letter that tipped Wallace off. But he couldn't recall a thing.
Oh, Lyell had warned him …Lyell had warned him…and now all my work, all my dreams—*all my dreams*—
Then he caught hold of himself. He mustn't give in to this horrible feeling overwhelming his solar plexus. There wassomething more important than priority and glory and applause and universal admiration and an awesome place in history …namely, his honor as a Gentleman and a scholar (Wolfe, 2016, p. 30–31).

The chief aim of this paper is to account for two observations about occurrences of FID in non-fiction like those above. First, they do not convey attributions or reports of actual thoughts or utterances. For instance, (4) does not convey that Livy actually said or thought, "Can it be that a local whore rather than a local wild beast found and tended the twins?"[2]. Similarly, it would be wrong to read (5) as claiming that Caesar actually thought, "The barbarians will simply line up to surrender." Nor does (6) communicate, for instance, that Darwin actually thought, "Oh, Lyell warned me," and so on (6) also includes DD. I return to this in 2.2 below).

Second, these occurrences of FID nevertheless do impart information about actual people or events. In the case of (4) we are supposed to understand, roughly, that Livy wondered how to understand the figure of the "lupa" in the myth. (5) is arguably intended to convey something like that Caesar thought the Britons were uncivilized and would be easily conquered. Correspondingly, audiences are meant to learn from (6), roughly, that Darwin was vexed by Wallace's discovery.

I will argue that both observations can be explained by seeing such instances of FID as *fictional uses*. This means that they are not used with assertoric force, that is, they are not assertions about the actual world. Rather, these instances of FID are used in the same way as the utterances used to tell a fictional story, as in the text of a novel.

Consider, for example, the first sentences of A.S. Byatt's *The Children's Book*:

(7) Two boys stood in the Prince Consort Gallery, and looked down on a third. It was June 19th, 1895. (Byatt, 2009, p. 5)

In writing (7), Byatt was not asserting that on 19 June 1895 two boys were standing in the Prince Consort Gallery looking down on a third boy. (7) is not put forward as a claim about what was actually the case. Rather, (7) is used non-assertorically. While there are many ways in which sentences can be used non-assertorically, in this case, following the standard approach, (7) is used with *fictional force*, that is as part of telling a fictional story. Roughly, that is, (7) does not make an assertion about the actual world, but makes its content true in the fiction *The Children's Book*.

Similarly, on the view I will develop here, (4) introduces a fictional scenario in which Livy thinks, "Can it be that a local whore rather than a local wild beast found and tended the twins?" Correspondingly, (6) introduces a fictional scenario in which Darwin thinks, "Oh, Lyell warned me," and so on. In each case, the fiction is about an actual person, Livy or Darwin, just as, for example, parts of *War and Peace* are about Napoleon[3]. So, while these uses of FID occur in non-fictional discourse, they attribute fictional thoughts or utterances to actual people. This explains the first observation by seeing these uses of FID as non-assertoric, and hence, as not attributing actual thoughts.

Given this, the second observation will be explained as an instance of the way in which audiences routinely learn things about the actual world from fictions. Fictional uses of FID in non-fiction exemplify the way in which "manipulation of point of view" (Friend, 2007b, p. 41) can be used to facilitate audiences' ability to comprehend and retain what is being communicated. Such instances of FID allow audiences to grasp information about actual people and events in a way that would not necessarily be achieved by flatly asserting things about the relevant people or events.

Section 2 reviews some important features of FID, specifically with respect to the examples under examination. I follow the standard view in arguing that both FID and DD, when reporting thoughts, attribute *inner speech*. Further, I suggest that FID and DD are truth-conditionally equivalent. I then flesh out the two main observations concerning the uses of FID in non-fiction that I am interested in.

Section 3 proposes a view of these occurrences of FID on which they are used with fictional force. I suggest that fictional discourse interacts with a species of common ground information that I call a *fictional record*, and I show how this view captures the non-assertoric nature of the relevant uses of FID.

Section 4 reviews some findings highlighted by Friend (2007b) concerning cognitive benefits of learning from historical fiction. I then propose that the occurrences of FID in non-fictional contexts under discussion can be seen as small-scale instances of historical fiction, and I comment on the consequences of such a view concerning learning factual information from these ways of reporting the thoughts and attitudes of historical figures.

2. FREE INDIRECT DISCOURSE, DIRECT DISCOURSE, AND INNER SPEECH

2.1. Free Indirect Discourse

The hallmark of FID is that it blends DD and ID in a relatively well-understood way. As is routinely observed, in FID tenses and person-features of pronouns are as in ID, while everything else—including indexicals, exclamations, and speaker-oriented expressions—is as in DD[4]. Take our examples in (1)–(3).

[2]It is well-known that DD allows reporting in a different language from the one used by the subject of the report, see e.g., Cappelen and Lepore (2007, p. 43–44). E.g., a DD report in English may report an utterance that was originally in German. I assume throughout this paper that it is clear that the same applies to FID, and hence that such a discrepancy is irrelevant to the topics discussed here.

[3]Cf. Currie (1990), Friend (2007a), Kripke (2011), and Stokke (2020).

[4]See, e.g., Banfield (1982, ch. 2), Sharvit (2008, 354), Schlenker (2004, 283–284), Eckardt (2015, 3–4), Maier (2015, 347–348).

(1) Ellen made a decision. Yes! I will tell him later today, she said/thought.
(2) Ellen made a decision. She said/thought that she would tell him later that day.
(3) Ellen made a decision. Yes! She would tell him later today(, she said/thought).

The FID report in (3) represents Ellen's utterance or thought by referring to Ellen with the 3rd person, and by using the past tense, as in the ID report in (2). At the same time, the FID report preserves the exclamation *Yes!* as it occurs in DD, as in (1). Similarly, (3) uses *today* to refer to the day of Ellen's decision, as in DD, rather than to the day the report is made, as would be the case in ID.

These features of FID are displayed by our examples. For instance, (6) refers to Darwin with the 3rd person while using the past tense as in ID. But, moreover, (6) preserves the exclamation *Oh* as it might appear in DD, as well as marking disfluency with "...".[5]

Further, as originally observed by Banfield (1982, ch. 2), FID behaves like DD in allowing direct questions, as seen from (8)[6].

(8) a. Where was he today? Ellen wondered.
b. *Ellen wondered where was he today?
(cf. Schlenker, 2004, p. 283–284)

Correspondingly, consider the fragment of (4) in (9).

(9) Was his wife the prostitute? Livy wondered.

(9) is parallel to (8a). (9) presents an FID report about what Livy wondered—roughly, whether the wife of the herdsman or shepherd was the prostitute.

Now consider the fragment of (5) in (10).

(10) Faced with the glittering armour of the legions and the eagle standards, the barbarians would simply line up to surrender. They would understand that history always fought on the side of Rome.

It might be thought that (10) can be understood as ID, given that there are no indexicals, exclamations, or other elements the behavior of which we could cite as evidence for FID. In particular, one might suggest that (10) should be understood as an instance of the phenomenon known as *Unembedded Indirect Discourse* (UID), described by, among others, Bary and Maier (2014). Briefly, instances of UID are ID but without the occurrence of a matrix like "*x* thought/said that..." Similarly, one might think that (10) is just ID but where such a matrix is left out.

However, there is evidence for FID, as opposed to UID, in (10). Consider the definite description *the barbarians*. As has been noted by Banfield (1982), Schlenker (2004), Bary and Maier (2014), and others, such definite descriptions behave differently in ID and FID, respectively. As Bary and Maier (2014) say,

> Definite descriptions in FID are protagonist-oriented (i.e., interpreted from the perspective of the reported speaker), whereas in indirect discourse they can be both protagonist-oriented and narrator-oriented (interpreted from the perspective of the actual speaker) (Bary and Maier, 2014, p. 82).

This means that if (10) is ID, and in particular UID, *the barbarians* should permit two readings, one on which it is attributed to Caesar, and one on which it is attributed to the speaker, Schama, but not to Caesar. That is, there should be a reading of (5) on which Schama thinks of the Britons as "barbarians," but Caesar does not. This reading is strictly speaking possible, but it is clearly neither the intended nor preferred one. Rather, the default reading of (5) is one on which Caesar, but not Schama, thinks of the Britons as "barbarians." In turn, this suggests that the default reading of (10) is the FID reading.

For these reasons, I take it that (4)–(6) do include occurrences of FID. The goal for what follows will be to sketch an account of these examples on the assumption that the relevant passages are interpreted as FID. Specifically, the aim will be to account for two main observations. First, in contrast to ordinary uses of FID in non-fiction, they do not attribute actual thoughts. Second, at the same time, these uses of FID are clearly intended to convey something about the thoughts or attitudes of the relevant people—Livy, Caesar, Darwin. In the rest of this section, I elaborate on each of these points in turn.

2.2. Fictional Uses of Direct Discourse in Non-Fiction

Before focusing exclusively on cases of FID like those in (4)–(6), I want to comment briefly on a parallel phenomenon involving DD. There are instances of DD in non-fiction that exhibit similar behavior to the occurrences of FID noted above. Here are two examples:

(11) Maelzel proposed that Beethoven compose a piece of music celebrating Wellington's victory at Vitoria [...]. Beethoven who was depressed and had nothing else to do said, "Yeah, sure, whatever." ("Beethoven: *Wellington's Victory* (1813)." *Music as a Mirror of History*. 2016. The Teaching Company. Audio).
(12) He goes to Ireland because it's his family basically who are conquering it, and they thought, "Yippee! This is a chance to build up our own territory outside the orbit of the sort of Anglo-Welsh problems. We can now go to Ireland and become terribly powerful there." And of course then Henry II thinks, "Oh my goodness, I can't miss out on this either." ("Gerald of Wales." *In Our Time*. 4 October 2012. BBC. Radio).

The observations I make in this paper about the cases involving FID apply, *mutatis mutandis*, to these instances of DD. The latter are also plausibly regarded as fictional uses. The speaker of (11) is not asserting that Beethoven actually said, "Yeah, sure, whatever," but is introducing a fictional scenario in which Beethoven makes that utterance. In turn, listeners are expected to learn, roughly,

[5] I use [...] to indicate omissions from quoted texts. The occurrences of "..." in (6) are from the original.
[6] (8) is grammatical and intelligible if read as FID—as in "Ellen wondered. Where was he today?"—yet this is irrelevant for the point at hand. For an overview of other relevant properties of FID, see Schlenker (2004, p. 283–284)

that Beethoven accepted but was not enthusiastic. And similarly for (12). I cannot undertake analyses of these cases of DD in this paper. Yet it should be flagged here that I take my account to apply to these, too.

To be sure, both (11) and (12) are spoken examples. Yet there are cases like these even with written DD. For instance, such occurrences of DD are found in (6), here underlined:

(6) [Describing Darwin's reaction to Wallace's discovery of the principle of evolution:] He racked his brain to recall whether or not he had written something in a letter that tipped Wallace off. But he couldn't recall a thing.

Oh, Lyell had warned him …Lyell had warned him …and now all my work, all my dreams—*all my dreams*—

Then he caught hold of himself. He mustn't give in to this horrible feeling overwhelming his solar plexus. There was something more important than priority and glory and applause and universal admiration and an awesome place in history …namely, his honor as a Gentleman and a scholar (Wolfe, 2016, p. 30–31).

At the same time, it is not implausible to think that uses of DD of this kind in written texts may be more rare than the corresponding uses of FID, in particular, if embedded under a matrix like "*x* said/thought...," which is absent from (6). I have no explanation for this contrast, if it exists, in this paper. One possible hypothesis might be that, since FID is conventionally associated with written, fictional discourse, it is more likely to be interpreted as fictional, even when appearing in non-fiction. Another suggestion is that DD presents a more verbatim or iconic representation of speech or thought and hence is harder to interpret as non-assertoric. However, these issues must be left to future work.

2.3. Free Indirect Discourse and Inner Speech

We have seen that our cases of FID do not attribute actual speech or thought. To spell this out further, note that standard uses of FID attribute what is sometimes called "occurrent," or "conscious," thoughts. As I will say, FID attributes *inner speech*[7]. In this respect, FID patterns with DD and contrasts with ID.

It is uncontroversial that ID *think* reports can report non-occurrent thoughts. That is, such reports can be true even if the subject has not consciously had the particular thought picked out by the complement clause. For instance, (13) can be true even if John has never thought to himself, "There's life in other galaxies."

(13) John thinks that there's life in other galaxies.

By contrast, the corresponding DD report in (14) clearly requires such an occurrent thought.

(14) "There's life in other galaxies," John thought.

While there may be different ways of cashing this out, I will say that (14) entails that "There's life in other galaxies" occurred in John's inner speech. That is, (14) is false if John did not think to himself, "There's life in other galaxies." For instance, suppose that it can be inferred from John's behavior and other beliefs that he thinks there's life in other galaxies, although John does not know the word *galaxy*. In that case (14) is clearly false, while (13) is true.

As has been realized since Banfield (1982), FID is like DD in this respect[8]. Here I follow Abrusn's 2020, 10–11 summary of the evidence for this conclusion. First, in FID "It is possible to add x thought/said, as an afterthought or interjection:"

(15) Tomorrow was her sixth year anniversary with Spencer, she thought, and it had been the best six years of her life (Maier, 2015).

Second, "Exclamatives and interrogatives are allowed:"

(16) She stood up in a sudden impulse of terror. Escape! She must escape! (Joyce, *Eveline*, cited in Abrsan, 2020.)

Third, "Hesitation, disfluency is allowed:"

(17) She wondered if he was still asleep, how did she even fall asleep and on top of him?!... Was he... shirtless? Oh,... he was... (Maier, 2015)

Fourth, "The protagonist's nonstandard dialect can be retained:"

(18) He [Big Boy] remembered the day when Buck, jealous of his winning, had tried to smash his kiln. Yeah, that ol sonofabitch! [...] Yeah, po ol Buck wuz dead now (Maier, 2015).

These observations suggest that FID reports attribute inner speech, as do DD *think* reports[9].

A consequence of this is the following generalization:

FID-DD Equivalence
FID and DD are truth-conditionally equivalent.

In other words, an FID report is true if and only if the corresponding DD report is. For example, (19a) is true if and only if (19b) is.

(19) a. Yes! She would tell him later today, Ellen said/thought.
b. "Yes! I will tell him later today," Ellen said/thought.

With respect to the versions with *thought*, both report an occurrence of "Yes! I will tell him later today" in Ellen's inner speech. Hence, if Ellen did not think to herself, "Yes! I will tell him later today," both are false. Similarly, if used with *said*, both are true if and only if Ellen uttered, "Yes! I will tell him later today."

We are not claiming that there are no differences between FID and DD reports. Indeed, we are not claiming that there may not be differences in what such reports communicate. What we are claiming is just that there is no truth-conditional difference. This claim is analogous to the uncontroversial claim that (20a–b) are truth-conditionally equivalent.

[7]There is a vast literature confirming the phenomenon of inner speech and its relation to thought, dating from at least the last four decades. For a useful overview of some recent work, see Langland-Hassan and Vicente (2018).

[8]Cf. Fludernik (1993), Maier (2015), and many others.

[9]See Huddleston and Pullum (2002, 1030) for a similar point.

(20) a. *Uttered in London*: It's foggy here.
b. It's foggy in London.

No one will deny that there are significant differences between the two, but everyone will agree that there is no truth-conditional difference in the sense that (20a) is true if and only if (20b) is.

FID-DD Equivalence has consequences for our examples of FID in non-fiction. For instance, it implies that (9) is true if and only if (21) is.

(9) Was his wife the prostitute? Livy wondered.

(21) "Is his wife the prostitute?" Livy wondered (to himself).

I take this to be correct. If Livy did indeed think to himself, "Is his wife the prostitute," then surely (9) is also true, and *vice versa*. But if Livy did not think to himself, "Is his wife the prostitute?" then clearly both are false.

This suggests that the FID reports in the examples of non-fictional discourse we have looked at are not assertoric. Given FID-DD Equivalence, if (9) is an assertion about what actually happened, then it is claiming that (9), and hence (21), are actually true. Yet this is not how (9) is interpreted, as it occurs in (4). Readers do not think that Beard is claiming that Livy actually wondered to himself, "Is his wife the prostitute?" in the same way that she might claim, for instance, that Livy wrote *Ab Urbe Condita*, that he was born in modern-day Padua, or that he was a friend of Augustus.

Similarly, given FID-DD Equivalence, (10) entails that Caesar thought to himself, "The barbarians will simply line up to surrender" and "They will understand that history always fights on the side of Rome." Yet audiences do not interpret these occurrences of FID as asserting that such inner speech events actually took place. You cannot object to this use of FID that Schama has no evidence that Caesar actually thought to himself, "The barbarians will simply line up to surrender," and so on. Correspondingly, (6) does not assert that Darwin actually thought to himself, "Oh, Lyell warned me," and so on.

We should conclude that, in all of our examples, FID is being used non-assertorically in that they do not attribute actual occurrences of inner speech. To be sure, it might be suggested instead that our examples show that FID-DD Equivalence is false, and that assertoric uses of FID sometimes do not attribute inner speech. However, the evidence summarized in (15)–(18), and earlier, disfavors this reaction. Assertoric uses of FID attribute inner speech. Yet our cases of FID do not. They are not assertoric uses.

2.4. Attributing Attitudes

The second observation we noted was that, even though they differ from standard, non-fictional uses of FID in not attributing inner speech, these occurrences of FID do convey information about the actual historical figures. At a certain level of abstraction, it is natural to say that (4) conveys that Livy wondered about the role of the she-wolf (the "lupa") in the Romulus and Remus myth, or that Caesar anticipated an easy victory over the Britons. Yet, although true, this leaves out some important aspects of the phenomenon.

Consider again (10).

(10) Faced with the glittering armour of the legions and the eagle standards, the barbarians would simply line up to surrender. They would understand that history always fought on the side of Rome.

While (10) does not assert that, for instance, Caesar thought to himself, "The barbarians will simply line up to surrender," clearly it does convey more than just that Caesar imagined an easy victory. For example, (10) conveys that Caesar thought of the Britons as "barbarians," that he thought the armor and the eagle standards of the Roman army would intimidate them, that Rome had a special place in history, and so on.

To illustrate further, it is useful to compare our examples with other instances of FID in non-fiction. Consider the example in (22), discussed by Fludernik (1993). (We will consider other cases in 3.2 below.)

(22) Reform of the Lords was a long-pursued mirage, Mr Powell told a House which started to fill up as news that he was on his feet spread. If it was found that the Lords really curbed the Commons, MPs would not stand for it for long. (Transcription of British parliamentary debate from *Survey of English Usage*; cited in Fludernik, 1993, 88).

As Fludernik notes, (22) clearly includes FID. But moreover, the FID in (22) is used non-fictionally. That is, the FID report in (22) makes an assertion about what Powell actually said. It asserts that Powell actually said, "Reform of the Lords is a long-pursued mirage," and so on.

This is the contrast with our examples. As we have emphasized, in our cases, FID is not used to make assertions about what the relevant individuals actually said or thought. At the same time, as we said above, these occurrences of FID do convey something about the actual individuals' attitudes. (10) conveys things about Caesar's attitudes beyond just suggesting that he thought the Britons would be easily conquered. For instance, (10) conveys that Caesar thought of the Britons as "barbarians." Similarly, (22) conveys that Powell called the reform a "mirage." Both are consequences of FID-DD Equivalence.

What we want to account for, then, is not just that (10) lacks assertoric force in that it does not attribute an actual event of inner speech to Caesar. We want to explain that, like ordinary, assertoric uses of FID such as (22), (10) conveys something about Caesar's attitudes. And similarly for the our other examples, (4) and (6).

In the next two sections I spell out the account I favor. I start by explaining what is meant by fictional force, as opposed to assertoric force.

3. FICTIONAL FORCE AND FICTIONAL RECORDS

3.1. Fictional Force

It is common to distinguish between different ways of using sentences pertaining to fictions. Take the example of (23) concerning the movie and play *Amadeus*.

(23) Salieri commissioned the *Requiem*. (Predelli, 2005, ch. 2)

We can distinguish between three ways of using (23). A *non-fictional use* of (23) is an assertion about the actual world. If used in this way, (23) is false, since Salieri did not commission Mozart's *Requiem* in the actual world. On a *fictional use* (23) is part of telling a fictional story. Imagine, for example, that (23) occurs as part of a novelization of the movie, or the like. Finally, one can use (23) *metafictionally*, as making a claim about a particular fiction. For instance, one can use (23) to make the claim that in *Amadeus* Salieri commissioned the *Requiem*.

For our purposes, the important difference is between fictional and non-fictional uses. I adopt the standard view on this distinction, endorsed by Searle (1975), Lewis (1978), Currie (1990), Sainsbury (2010), Davies (2015), Recanati (2000), 2018, and many others. According to this view, fictional and non-fictional uses are distinguished only in terms of force[10]. In other words, there is no syntactic or semantic difference between fictional and non-fictional discourse. As Searle (1975) wrote in an often quoted passage,

> There is no textual property, syntactical or semantic, that will identify a text as a work of fiction. What makes it a work of fiction is, so to speak, the illocutionary stance that the author takes toward it [...] (Searle, 1975, p. 325).

Similarly, Currie (1990) writes,

> If Doyle had been writing history instead of fiction when he wrote "It rained in London on January 1, 1895," he would have been making an assertion. The transition from history to fiction is marked, at least, by the loss of one kind of force: assertative force (Currie, 1990, p. 6–7).

Correspondingly, (23) has the same syntactic and semantic profile when used to make an assertion about the actual world and when used as part of a fictional story[11]. That is, it is true or false (at a world *w*) if and only if Salieri commissioned the *Requiem* (in *w*). Instead, the difference between fictional and non-fictional, assertoric utterances is a difference in force.

On the view I favor, the FID reports in (4)–(6) are fictional uses. We have already seen that they are not non-fictional uses, since they do not assert attributions of actual inner speech. Further, our cases are not metafictional uses. They are not assertions about fictions. For instance, (6) is not making an assertion like, in such-and-such fictional story, Darwin thinks, "Oh, Lyell warned me." If it did, it would be (actually) true or false depending on whether, in the relevant fiction, Darwin thinks, "Oh, Lyell warned me." This I take to be the wrong result.

In other words, just like the sentences in a novel, (4)–(6) introduce fictional scenarios in which it is true that the relevant inner speech events take place. Similarly, for instance, Byatt's fictional use of (7) makes it true in *The Children's Book* that on 19 June 1895 two boys stood in the Prince Consort Gallery and looked down on a third boy[12].

(7) Two boys stood in the Prince Consort Gallery, and looked down on a third. It was June 19th, 1895 (Byatt, 2009, p. 5).

Correspondingly, (6) makes it true in the fiction about Darwin it introduces that Darwin thought to himself, "Oh, Lyell warned me."

A positive motivation for taking our examples to involve fictional uses of FID is that they are overtly indistinguishable from assertoric uses of FID, such as (22). There is no overt linguistic material in our examples that is responsible for their difference from assertoric uses of FID, such as a modal, or the like. Given this, the fact that our uses of FID do not report actual inner speech is plausibly due to the pragmatics of the way they are used, rather than to their semantic or syntactic profile. In other words, they conform straightforwardly to the observation that fictional and non-fictional uses are distinguished only in terms of force.

To be sure, one can posit covert structure for our examples, such as an unpronounced operator, and thereby claim that there is a syntactic, and hence semantic, difference[13]. The fictional account does not do so. I take that to be a point in its favor, given the overt indistinguishability. But moreover, I take it to be a *prima facie* plausible suggestion that the difference between the way, say, (4) represents Livy as wondering about the role of the she-wolf in the myth and the way that inner speech is standardly reported by means of FID is a pragmatic difference, and not a difference in what is said, or truth-conditional content.

If we accept that the FID reports in these cases are used fictionally, we explain the first observation noted in the last section. That is, the occurrences of FID in our examples do not convey attributions of actual inner speech because they are used fictionally to introduce fictional scenarios in which the relevant thoughts or utterances take place.

3.2. Other Examples of Free Indirect Discourse in Non-Fiction

Before moving on to spelling out the fictional account of the occurrences of FID we have looked at, it is worth commenting further on other occurrences of FID in non-fiction.

Fludernik (1993, p. 88) observed that FID "occurs widely in non-literary texts." (emphasis removed) These are also cases of FID appearing in non-fictional discourse[14]. We noted above that at least some of these cases are assertoric uses, that is, occurrences of FID that attribute actual thoughts or utterances. This was the case for (22), repeated here:

(22) Reform of the Lords was a long-pursued mirage, Mr Powell told a House which started to fill up as news that he was on his feet spread.

[10] See Walton (1990, ch. 2) for opposition.

[11] By contrast, the orthodox approach to metafictional uses follows Lewis (1978) in seeing them as having the same content as the corresponding sentence prefixed with an operator like "In *Amadeus*..." For recent, different versions of this approach to metafictional discourse, see e.g., Predelli (2008), Recanati (2018), Stokke (2020).

[12] In this sense, fictional uses usually give rise to metafictional truths, although not always, because narrators can be unreliable. For example, given that (7) makes its content true in *The Children's Book*, one can subsequently use the same sentence, (7), to make a true, metafictional assertion about that fiction.

[13] This suggestion should not be confused with accounts of FID itself that posit covert operators, like that of Sharvit (2008).

[14] See also Zeman (2018, 181 fn. 6), and see Fludernik (1993) for more references.

If it was found that the Lords really curbed the Commons, MPs would not stand for it for long. (Transcription of British parliamentary debate from *Survey of English Usage*; cited in Fludernik, 1993, p. 88).

Here are two more examples from Fludernik (1993):

(24) He [Montagu] suffered agonies from her sexual rejection, which was known to their intimate friends; and when his political career was finished, what was there to live for? And yet some cynics might have said that he had achieved his ambition. The daughter of a famous aristocratic family had accepted him. Did not this prove that he was on equal terms with the rulers of the land? (*New York Review of Books*, 1991; cited in Fludernik, 1993, p. 89).

(25) Within a matter of weeks, according to a *Newsweek* poll, 77 percent of the American public had become aware of George Holliday's Rodney King video, had counted the kicks and the baton strikes and identified on the grainy, badly lit tape which officers were which, here was the indefatigable Wind, over there the stomper Briseno (*New York Review of Books*, 1991; cited in Fludernik, 1993, p. 89).

Are these fictional uses of FID within non-fiction? That is, are they non-assertoric uses, as in our examples, or are the ordinary, non-fictional, assertoric uses, like (22)? Let us consider each in turn.

In (24) FID is used to report thoughts, that is, inner speech. So, the question is whether (24) should be taken as asserting, for instance, that Montagu actually thought to himself, "Doesn't this prove that I am on equal terms with the rulers of the land?" Most likely, this is not the right reading of (24). Rather, (24) is most naturally understood as on a par with our own examples. That is, it presents a fictional report of Montagu's thoughts, just as, for instance, (6) presents a fictional report of Darwin's thoughts.

Finally consider (25). This case is arguably different from each of the two preceding examples. It is natural to think that FID in this case reports inner speech like, "there is the indefatigable Wind, over there the stomper Briseno." Yet there is no clear indication of the subject of the report. As such, (25) is an instance of what we might call *unidentified* FID[15]. Here is an example from Thomas Mann's *Buddenbrooks*:

(26) A tooth—Senator Buddenbrook had died of a toothache, that was the word around town. But, confound it all, people didn't die of that! He had been in pain, Herr Brecht had broken off the crown, and afterward he had simply collapsed on the street. Had anyone ever heard the like? (Mann, 1904, p. 666)

Here FID gives a report of someone's speech. But whose? We are presented with "the word around town." But this is not a report of a particular person's speech. Instead, it is natural to think that we are being presented with the speech of an arbitrary member of the group in question, in this case the relevant townsfolk. (25) is arguably parallel, except that it appears in a non-fictional context. I take the phenomenon of unidentified FID to be different from the kind of fictional uses of FID that we have looked at, and I will not attempt an analysis of such cases here.

So, we can note that, as exemplified by (22), FID can be used assertorically in non-fictional discourse. That is not surprising. Indeed we have presupposed as much here. That is, FID is a style of report which can be used to attribute thoughts (i.e., inner speech) and utterances both outside and inside fictional contexts, just like ID and DD. When FID occurs in non-fictional contexts, it standardly attributes actual thoughts or utterances, as in (22). By contrast, the cases we have examined attribute fictional thoughts or utterances to actual people, although they occur in non-fictional contexts,

3.3. Fictional Records

I have suggested that the non-assertoric uses of FID we have examined are fictional uses, that is, they are used with fictional force. To make this suggestion more concrete, it is convenient to implement it within the framework for understanding discourse and assertion familiar from the work of Stalnaker 1970, 1978, 1998, 2002, 2014. On this picture, discourse relies on a body of information, called the *common ground*, that is taken for granted for the purpose of the exchange. The common ground acts both as support for utterance interpretation and as storage for information communicated by the participants.

Central to this theory of communication is an understanding of assertion. To utter a sentence *S* with assertoric force is to propose that the propositional, or truth-conditional, content of *S*, given the context, become part of the common ground. By contrast, a non-assertoric utterance is one that falls short of making a proposal to increment common ground information with what it says. For instance, if uttered ironically, (27) is not an assertion, and correspondingly is not a proposal to make it common ground that *The Da Vinci Code* is a great novel.

(27) Oh yeah! *The Da Vinci Code* is a GREAT novel!

Given this, to say that fictional discourse is distinguished from non-fictional discourse by not being assertoric is to say that fictional utterances do not involve proposals to increment common ground information.

Instead, I suggest that utterances made with fictional force interact with alternative bodies of information comprising what is part of the story at a given time during its unfolding[16]. Fictional discourse features presuppositions, anaphora, indexicals, and other elements that rely on contextual information. As a simple illustration, consider this sentence from Doyle's *A Study in Scarlet*:

(28) Sherlock Holmes rose and lit his pipe (Doyle, 1887, p. 24).

(28) is the first appearance of Holmes's pipe in *A Study in Scarlet*, itself the first appearance of Holmes to the reading public. Accordingly, (28) has the effect of making audiences include

[15] Wood (2008, 24) calls this "unidentified free indirect style."

[16] Related accounts of fictional discourse in terms of alternative contextual information have been proposed by Bonomi and Zucchi (2003), Stokke (2018), 2020, Eckardt (2015, in press), Semeijn (2017), and Maier and Semeijn (in press).

Holmes's pipe in the story, as an instance of presupposition accommodation. Correspondingly, the pipe is available for anaphoric reference later in the text, as in the following hypothetical continuation:

(29) Sherlock Holmes rose and lit his pipe. A few minutes later he put it down in an ashtray.

Observations of this kind motivate a picture according to which, during the unfolding of a fictional story, a body of information delineating what is taken to be part of the story evolves. I will call this a *fictional record*. As illustrated by (28), and as is apparent from run-of-the-mill fictional discourse, fictional records play a role analogous to the role played by ordinary common ground information in everyday conversation.

Along these lines, we think of utterances made with fictional force as contributions to fictional records. While fictional utterances are non-assertoric in the sense that they are not directed at adding information to what is common ground, they nevertheless function to increase information that is part of fictional records. So when (28) is read by the audience, they update the fictional record with, at least, the information that Holmes has a pipe, that Holmes rose (got up), and that Holmes lit his pipe.

3.4. Adding Free Indirect Discourse Reports to Fictional Records

I suggest that when audiences engage with works like those from which we have drawn our examples, they take some of the sentences as updating the information they think of as the "official" information of the discourse (analogous to the ordinary common ground of a conversation), and others as updating fictional records. In our cases the latter being the uses of FID.

For instance, readers of (4) will be aware of what is, and has been, conveyed as official information of the book. Along the lines of what we said earlier, the official information might include things like that Livy wrote *Ab Urbe Condita*, that he was born in modern-day Padua, or that he was a friend of Augustus. At the same time, audiences are aware of fictional information conveyed by the text. In the case of (4), this fictional information comprises, at least, that Livy wondered to himself, "Is his wife the prostitute?"

In other words, at least one effect of the fictional uses of FID is to add to a fictional record that the relevant inner speech event took place. For instance, in the case of (5), roughly, there will be a fictional record that includes the information that Caesar thought to himself, "The barbarians will simply line up to surrender." At the same time, as we said before, more is conveyed by FID reports than just the occurrence of inner speech events. For instance, it will likewise be part of the fictional record that Caesar thought of the Britons as "barbarians," that he thought the armor and the eagle standards of the Roman army would intimidate them, that Rome had a special place in history, and so on.

Specifically, the passage in (5) has the effect, at least, of producing an information state in which, apart from the official information, there is a cache of fictional information. For ease of reference, call the fictional record to which the FID reports about Caesar in (5) are added A. We can then schematize A as follows:

A
A_1 Caesar thought to himself, "Faced with the glittering armor of the legions and the eagle standards, the barbarians will simply line up to surrender."
A_2 Caesar thought to himself, "The barbarians will understand that history always fights on the side of Rome."
A_3 Caesar thought that the Britons were barbarians.
A_4 Caesar thought that the Britons would be easily conquered.
A_5 Caesar thought that Rome had a special place in history.
...

One can think of fictional records, like ordinary common grounds, as sets of propositions. Further, one can think of propositions as sets of possible worlds, in the familiar fashion. Given this, it may be useful for some purposes to represent a fictional record itself as a set of worlds, corresponding to the standard notion of a *context set*, that is, as the intersection of all the propositions in a particular fictional record, which will represent the possibilities compatible with what is included in the relevant record. For instance, A might be represented as the set $A_1 \cap ... \cap A_n$.

However, for the purposes of this discussion, these further implementations will not play a role. The aim here is to give an account of the non-assertoric force of the relevant occurrences of FID, and further to suggest some factors in how they facilitate learning about actual historical events. I have argued that the former point can be explained by seeing these utterances as aimed at updating fictional records. In the next section I turn to the second point.

4. LEARNING FROM FICTIONAL REPORTS

4.1. Coming to Know and Coming to Believe

On the view outlined in the last section, for instance, the fictional uses of FID in (5) produce a fiction about Caesar in which (at least) $A_1 ... A_5$ are true. I suggest that, just as one can often learn things from fictions, in these cases, audiences can learn things from the fictions introduced by the occurrences of FID.

As we use the terms here, learning that p implies coming to know that p. Hence, trivially, audiences cannot come to learn something that is actually false from any fiction. Yet, of course, they may come to believe such things based on fictions, and thereby be misled. As a simple, hackneyed example, audiences to *A Study in Scarlet* might come to believe that 221B Baker Street existed in 1887 when the novel was published. Yet since this is false, trivially they cannot learn that, but rather they will acquire a false belief in this case.

Considering our case of A, let us assume that A_1–A_2 are actually false, while A_3–A_5 are actually true. That is, while the inner speech events reported by the former did not in fact occur,

the attributions of beliefs to Caesar in the latter are nevertheless accurate. In that case, we may take it that, given that other constraints on learning from fiction are satisfied, audiences may learn A_3–A_5 from A.

To be sure, there is a significant challenge in specifying what these other constraints are. Yet it is not among my aims here to take up that challenge[17]. A few remarks are in order, however. In particular, as is standard, we should distinguish between two broad ways of learning from fiction. Gendler (2000) describes this difference as follows[18]:

> *narrative as clearinghouse*: I export things from the story that you the storyteller have intentionally and consciously imported, adding them to my stock in the way that I add knowledge gained by testimony. [...]
> *narrative as factory*: I export things from the story whose truth becomes apparent as a result of thinking about the story itself. These I add to my stock the way I add knowledge gained by modeling (Gendler, 2000, p. 76).

Given this distinction, it is plausible to think that our cases fall under the second of these general ways of learning from fiction. In particular, the process by which audiences come to believe, and learn, A_3–A_5 from reading (5) is plausibly understood as based on these things becoming "apparent as a result of thinking about the story," that is, in our case, as the result of understanding the FID report. Presented with the fiction in which Caesar thought to himself, "The barbarians will simply line up to surrender," and so on, audiences are likely to learn, for instance, that he thought of the Britons as "barbarians" as a result of that truth about Caesar becoming apparent from the story.

This suggests that the process in question is far from automatic and, unquestionably, many factors are involved. For instance, the reason that an audience may come to believe A_3–A_5 from reading (5) most likely stems from factors such as the their awareness of the genre they are reading, other evidence that they have about Caesar, and of Rome in general, their assessment of the overall plausibility of A_3–A_5, and more. I will not attempt a further account of this here.

Instead, I want to focus on some ways in which learning from fiction has been shown to be cognitively beneficial. More colloquially, we are good at learning things from fiction, demonstrably better than we are at learning things from "dry" expositions of factual information. As I go on to explain, there are good reasons to think that the fictional uses of FID in our examples fit this general pattern.

4.2. Perspective and Historical Fiction

Stacie Friend (2007b, 41) has reported a number of studies in cognitive psychology showing that "the manipulation of the point of view from which we learn about events" involved in historical fiction "generates numerous epistemic advantages" with respect to acquiring information about the actual events depicted in the fiction[19].

Taking Gore Vidal's *Lincoln: A Novel* as an example, Friend highlights that the novel presents things to the audience as if they were eyewitnesses:

> *Lincoln* plunges us directly into the flow so that we "see" the president in action: we learn about Lincoln through the eyes and minds of people close to him, rather than from Vidal's real retrospective point of view (Friend, 2007b, p. 41).

As Friend observes, a key difference between this kind of historical fiction and conventional, non-fictional history is that it allows for direct representation of the thoughts and attitudes of historical figures[20]:

> Because the writer of a work of non-fiction could not possibly have such access to the minds of other people, histories and biographies standardly present the thoughts of real individuals as inferences from the evidence. And they provide information about their evidential sources. This is by contrast with Vidal's narration, which provides the reader with fictional, seemingly *direct* access to the thoughts of certain characters (Friend, 2007b, p. 38).

Friend summarizes a number of ways in which this kind of shift in perspective has been shown to facilitate learning and retention of information on the part of audiences. Here I want to highlight two of these. First,

> One advantage is that such eyewitness descriptions are more likely to be *concrete*, thereby generating more imagery; this in turn seems to significantly enhance memorability (Friend, 2007b, p. 41).

Second,

> A related epistemic advantage of *Lincoln* depends on the *reduction of exposition* afforded by Vidal's technique. [...] It turns out that narratives display an advantage over expositions in studies of reading comprehension. Expository texts, when they treat unfamiliar topics, prompt subjects to process information as so many separate items to be memorized [...]. By contrast, narratives prompt readers to focus on the situation the text is about (Friend, 2007b, p. 42).

As I argue below, these points equally apply to the fictional uses of FID we have examined.

4.3. Concreteness and Reduction of Exposition

Our cases can be described as examples of historical fiction, albeit in the concentrated form of free-standing FID reports. If what I have argued is on the right track, they constitute small-scale pockets of historical fiction within historical non-fiction.

[17] For discussion, see e.g., Lewis (1978), Currie (1990), Gendler (2000), Green (2010), and Currie (2020).

[18] See also Lewis (1978), Green (2010), Stock (2017), and Hazlett (2017). See García-Carpintero (2016) for an overview of recent work.

[19] For similar suggestions, see e.g., Lamarque (1997), Elgin (2007), Camp (2017). And see Friend (2007b) for further references.

[20] See Walton (1990, section 2.1) for similar observations.

In particular, they are instances of precisely the kind of fictional presentation of immediate access to the thoughts of historical figures that Friend describes.

For instance, by interrupting the non-fictional discourse of the book with the FID report in (6), "Oh, Lyell had warned him... Lyell had warned him," Wolfe gives us "fictional, seemingly *direct* access" to Darwin's thoughts. So, we should expect that at least some of the particular features of the ways in which we routinely learn things about actual historical events from historical fiction apply to our cases, too.

As we noted above, Friend reports that historical fiction has been seen to facilitate learning and retention of information by, among other things, being concrete and by reducing exposition. It is relatively clear that FID reports of inner speech are concrete in the relevant sense. In particular, both FID and DD contrast with ID in this respect. Rather than being presented with a description of the content of the relevant thoughts, we are presented with those thoughts themselves, so to speak. As such, it is plausible to think that FID is likely to generate imagery in the sense that has been shown to enhance memorability.

With respect to the second point we highlighted above, it is worth being clear that there are potentially two ways in which one can understand the idea of reduction of exposition—the reason being that there are two main ways of understanding the notion of exposition. First, one might think of exposition as representation of events, actions, scenes, and so on, that is, as opposed to representation of speech or thought. In this sense, exposition corresponds roughly to the traditional category of *diegesis*, as contrasted with *mimesis*. On this way of understanding exposition, it is trivial to say that FID reduces exposition, since FID is a device for representing speech or thought.

Second, however, one can understand exposition as a particular way of organizing information. This is how Friend uses the term in the quote above, where exposition is contrasted with "narrative." So, the claim is not that the mere replacement of *diegesis* with *mimesis* amounts to the kind of reduction of exposition that is cognitively beneficial for learning. Rather, it is the replacement of non-narrative presentation with a more narrative style.

As I explain below, it is arguable that FID is particularly conducive to this kind of reduction of exposition.

4.4. Narrative and Imitation

There are reasons to think that FID itself facilitates narrative presentation, whereas DD does not. Of course, DD routinely figures in narrative discourse. Yet it is not difficult to imagine a non-narrative, factual presentation of events that includes DD reports of, for example, statements by politicians, military officials, and so on. Much history is written in this way. Similarly, a government report on some issue might take that form. Yet it would be hard to imagine such a report including FID, even FID reporting speech. Indeed, the transcript of British parliamentary debate in (22) we cited earlier arguably has a narrative character, even though FID is used non-fictionally.

FID reports are typical of the kind of narrative style that counts as reduced exposition. Many studies of FID highlight this feature. In an early treatment, Cohn (1966, p. 98) called FID "narrated monologue," and noted that it "enables the author to recount the character's silent thoughts without a break in the narrative thread." More recently, Zeman (2018, 174) has argued that FID is "restricted to narrative discourse mode only." The central motivation for this conclusion being that FID involves two kinds of perspective, the speaker's and that of the subject of the report, where this duality is seen as the key characteristic of narrative discourse.

I suggest that, first, we should understand the relevant kind of reduction of exposition as a tendency toward narrative presentation, rather than merely providing FID or DD reports (*mimesis*) instead of pure description of events (*diegesis*). And second, FID in particular, as opposed to DD, facilitates this kind of reduction of exposition because FID itself is narrative device. As argued by Zeman (2018), and others, FID is a narrative device in that it essentially involves two perspectives, that of the narrator (or speaker) and the protagonist (or subject of the report).

Given this, we should ask to what extent the double perspective involved in FID might itself play a role in the acquisition of factual information. One suggestion is that FID allows the author to communicate attitudes toward the historical figures in question in a particularly effective way.

Currie (2010) has argued that FID is distinguished from DD by involving *imitation*, whereas DD is a way of *replicating* speech or thoughts:

> When Barkis says "Barkis is willin", I might report this by saying 'Barkis said he is willing', or 'Barkis said "Barkis is willin"', neither of which strikes one as particularly imitative of his odd turn of phrase, though the second certainly draws attention to it. If I say "Barkis was willin", I have injected a distinct element of imitation into the report. (Currie, 2010, p. 142)

On Currie's view, this involvement of imitation in FID is particularly due to its double perspective, that is, its blending of DD and ID:

> With FID [...] we have the sense that it is the narrator speaking, though speaking in a way which is highly constrained by the words, the tone, the style of the character whose speech is represented: there is something theatrical about FID as a mode of reporting which makes it difficult not to think of the speaker as imitating another. (loc. cit.)

For Currie, a central aspect of imitation, in this sense, is that "we have the sense of sharing with the author a way of experiencing and responding to those events, leading to a sense of guided attending on our part" (Currie, 2010, p. 106).

I argued that the use of FID in (5) allows readers to learn, for instance, that Caesar thought the Britons were uncivilized and would be easily conquered. So, one suggestion is that because of the imitative feature of FID, which stems directly from its involvement of two perspectives, readers not only learn that fact about Caesar but is allowed to share the author's understanding of it and attitudes toward it. Rather than just being told that Caesar thought the Britons were uncivilized and would be easily conquered, we are allowed to share the author's sense of Caesar's

particular imperious high-mindedness, his being "giddy with invincibility," as manifested by the representation of Caesar as thinking, "the barbarians will simply line up to surrender." Yet, at the same time, by using FID fictionally, and hence non-assertorically, the author is able to convey these attitudes toward Caesar without outright lying or making baseless claims about what Caesar's particular thoughts were.

Similar points can be said to apply to the other examples. For instance, the FID uses of direct questions in (4) convey, not just that Livy was interested in re-thinking or interpreting the myth, but that questions like how to think of the identity of the *lupa* character was particularly high on his mind, and that he was consciously pondering them. Similarly, the FID use of "Oh, Lyell had warned him..." in (6) may suggest that Darwin was regretful of not having considered the possibility earlier, and not having heeded Lyell's warning, and that such things were weighing on his conscience and persistently bothering him.

In other words, FID can be seen as a device for representing speech and thought that is specifically narrative in character. In particular, FID instantiates the double perspective that, as argued by Zeman (2018), is distinctive of narrative discourse. By doing so, FID involves imitation, which in turn allows audiences to learn about the author's understanding of, or attitudes toward, the people in question. More generally, it is plausible that the uses of FID in our examples are paradigmatic examples of the kind of reduction of exposition that Friend cites as one of the key cognitive benefits of historical fiction.

5. CONCLUSION

FID is used in both fictional and non-fictional discourse. We have examined some cases in which FID occurs within non-fictional discourse that are characterized by the fact that FID is used non-assertorically. The occurrences of FID in these cases do no assert attributions of inner speech, as FID standardly does.

I have argued that these occurrences of FID are used with fictional force. As such, they contribute to fictional records that are available to audiences alongside "official" discourse information. In turn, they introduce small-scale (historical) fictions into the otherwise non-fictional discourse. By doing so, they facilitate acquisition and retention of factual information about the relevant historical figures and events by the audience.

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

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Abrusán, M. (2020). The spectrum of perspective shift: Protagonist projection vs. free indirect discourse. *Linguist. Philos.* doi: 10.1007/s10988-020-09300-z

Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Boston, MA: Routledge; Kegan Paul.

Bary, C., and Maier, E. (2014). Unembedded indirect discourse. *Proceedings of Sinn und Bedeutung* 18, 77–94.

Beard, M. (2015). *SPQR: A History of Ancient Rome*. New York, NY: Liveright. doi: 10.1525/9780520958203

Bonomi, A., and Zucchi, S. (2003). A pragmatic framework for truth in fiction. *Dialectica* 57, 103–120. doi: 10.1111/j.1746-8361.2003.tb00259.x

Byatt, A. (2009). *The Children's Book*. New York, NY: Knopf.

Camp, E. (2017). Perspectives in imaginiative engagement with fiction. *Philos. Perspect*. 31:73–102. doi: 10.1111/phpe.12102

Cappelen, H., and Lepore, E. (2007). *Language Turned On Itself*. Oxford; New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199231195.001.0001

Cohn, D. (1966). Narrated monologue: definition of a fictional style. *Comp. Literat.* 18, 97–112. doi: 10.2307/1770156

Currie, G. (1990). *The Nature of Fiction*. Cambridge; New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511897498

Currie, G. (2010). *Narratives and Narrators: A Philosophy of Stories*. Oxford; New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199282609.001.0001

Currie, G. (2020). *Imagining and Knowing: The Shape of Fiction*. Oxford; New York, NY: Oxford University Press. doi: 10.1093/oso/9780199656615.001.0001

Davies, D. (2015). Fictive utterance and the fictionality of narratives and works. *Brit. J. Aesthet.* 55, 39–55. doi: 10.1093/aesthj/ayu061

Doron, E. (1991). Point of view as a factor of content. *Proc. SALT* 1, 51–64. doi: 10.3765/salt.v1i0.2997

Doyle, A. (1981 [1887]). "A study in scarlet," in *The Penguin complete Shelock Holmes* (London: Penguin), 15–87.

Eckardt, R. (2015). *The Semantics of Free Indirect Discourse: How Texts Allow Us to Mindread and Eavesdrop*. Leiden; Boston, MA: Brill. doi: 10.1163/9789004266735

Eckardt, R. (in press). "In search of the narrator," in *The Language of Fiction*, eds E. Maier and A. Stokke (Oxford; New York, NY: Oxford University Press).

Elgin, C. Z. (2007). "The laboratory of the mind," in *A Sense of the World: Essays on Fiction, Narrative and Knowledge*, eds J. Gibson, W. Huemer, and L. Pocci (London: Routledge), 43–54.

Fludernik, R. (1993). *The Fictions of Language and the Language of Fiction: The Linguistic Representation of Speech and Consciousness*. London: Routledge.

Friend, S. (2007a). Fictional characters. *Philos. Compass* 2, 141–156. doi: 10.1111/j.1747-9991.2007.00059.x

Friend, S. (2007b). "Narrating the truth (more or less)," in *Knowing Art: Essays in Aesthetics and Epistemology*, eds M. Kieran and D. M. Lopes (Dordrecht: Springer), 35–49. doi: 10.1007/978-1-4020-5265-1_3

García-Carpintero, M. (2016). Recent debates on learning from fiction. *Teorema* 35, 5–20.

Gendler, T. S. (2000). The puzzle of imaginative resistance. *J. Philos.* 97, 55–81. doi: 10.2307/2678446

Green, M. (2010). "How and what can we learn from fiction?" in *A Companion to the Philosophy of Literature*, eds G. Hagberg and W. Jost (Oxford: Wiley-Blackwell), 350–366. doi: 10.1002/9781444315592.ch18

Hazlett, A. (2017). "Imagination that amounts to knowledge," in *Art and Belief*, eds E. Sullivan-Bissett, H. Bradley, and P. Noordhof (Oxford; New York, NY: Oxford University Press), 119–134.

Huddleston, R., and Pullum, G. (2002). *The Cambridge Grammar of the English Language*. Cambridge; New York, NY: Cambridge University Press. doi: 10.1017/9781316423530

Kripke, S. A. (2011). "Vacuous names and fictional entities," in *Philosophical Troubles*, ed S. Kripke (Oxford; New York, NY: Oxford University Press), 52–74. doi: 10.1093/acprof:oso/9780199730155.003.0003

Lamarque, P. (1997). Learning from literature. *Dalhousie Rev*. 77, 7–21.

Langland-Hassan, P., and Vicente, A. (2018). "Introduction," in *Inner Speech: New Voices*, eds P. Langland-Hassan and A. Vicente (Oxford; New York, NY: Oxford University Press), 1–28. doi: 10.1093/oso/9780198796640.001.0001

Lewis, D. (1983 [1978]). Truth in fiction. In Philosophical Papers, volume 1, pages 261-280. Oxford University Press, Oxford and New York. doi: 10.1093/0195032047.003.0015

Maier, E. (2015). Quotation and unquotation in free indirect discourse. *Mind Lang*. 30, 345–373. doi: 10.1111/mila.12083

Maier, E., and Semeijn, M. (in press). "Extracting fictional truth from unreliable sources," in *The Language of Fiction*, eds E. Maier and A. Stokke (Oxford; New York, NY: Oxford University Press).

Mann, T. (1994 [1904]). *Buddenbrooks: The Decline of a Family*. New York, NY: Vintage International.

Predelli, S. (2005). *Contexts: Meaning, Truth and the Use of Language*. Oxford: Clarendon Press. doi: 10.1093/0199281734.001.0001

Predelli, S. (2008). Modal monsters and talk about fiction. *J. Philos. Logic* 37, 277–297. doi: 10.1007/s10992-007-9073-z

Recanati, F. (2000). *Oratio Obliqua, Oratio Recta: An Essay on Metarepresentation*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/5163.001.0001

Recanati, F. (2018). Fictional, metafictional, parafictional. *Proc. Aristotelian Soc*. 118, 25–54. doi: 10.1093/arisoc/aoy001

Sainsbury, M. (2010). *Fiction and Fictionalism*. London; New York, NY: Routledge. doi: 10.4324/9780203872567

Schama, S. (2009 [2000]). *A History of Britain: At the Edge of the World? 3000BC-AD1603*, Vol. 1. London: Random House.

Schlenker, P. (2004). Context of thought and context of utterance: a note on free indirect discourse and the historical present. *Mind Lang*. 19, 279–304. doi: 10.1111/j.1468-0017.2004.00259.x

Searle, J. (1975). The logical status of fictional discourse. *N. Literary History* 6, 319–332. doi: 10.2307/468422

Semeijn, M. (2017). "A stalnakerian analysis of metafictive statements," in *Proceedings of the 21st Amsterdam Colloquium*.

Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguist. Philos*. 31, 353–395. doi: 10.1007/s10988-008-9039-9

Stalnaker, R. (1999 [1970]). "Pragmatics," in *Context and Content* (Oxford; New York, NY: Oxford University Press), 31–46. doi: 10.1093/0198237073.003.0002

Stalnaker, R. (1999 [1978]). "Assertion," in *Context and Content* (Oxford; New York, NY: Oxford University Press), 78–95. doi: 10.1093/0198237073.003.0005

Stalnaker, R. (1999 [1998]). "On the representation of context," in *Context and Content* (Oxford; New York, NY: Oxford University Press), 96–114. doi: 10.1093/0198237073.003.0006

Stalnaker, R. (2002). Common ground. *Linguist. Philos*. 25, 701–721. doi: 10.1023/A:1020867916902

Stalnaker, R. (2014). *Context*. Oxford; New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199645169.001.0001

Stock, K. (2017). "Fiction, testimony, belief, and history," in *Art and Belief*, eds E. Sullivan-Bissett, H. Bradley, and P. Noordhof (Oxford; New York, NY: Oxford University Press), 19–41. doi: 10.1093/oso/9780198805403.003.0002

Stokke, A. (2018). *Lying and Insincerity*. Oxford; New York, NY: Oxford University Press. doi: 10.1093/oso/9780198825968.001.0001

Stokke, A. (2020). Fictional names and individual concepts. *Synthese*. doi: 10.1007/s11229-020-02550-1

Wolfe, T. (2016). *The Kingdom of Speech*. London: Jonathan Cape.

Wood, J. (2008). *How Fiction Works*. New York, NY Farrar, Straus and Giroux.

Zeman, S. (2018). "What is a Narration-and Why Does It Matter?" in *Linguistic Foundations of Narration in Spoken and Sign Languages*, eds A. Hübl and M. Steinbach (Amsterdam; Philadelphia, PA: John Benjaminis), 173–206. doi: 10.1075/la.247.08zem

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.

Copyright © 2021 Stokke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Communication

ORIGINAL RESEARCH
published: 15 February 2021
doi: 10.3389/fcomm.2020.611935

Reading Fictional Narratives to Improve Social and Moral Cognition: The Influence of Narrative Perspective, Transportation, and Identification

Lena Wimmer[1,2]*, *Stacie Friend*[3], *Gregory Currie*[4] and *Heather J. Ferguson*[1]

[1]School of Psychology, University of Kent, Canterbury, United Kingdom, [2]Department of Education, University of Freiburg, Freiburg im Breisgau, Germany, [3]Department of Philosophy, Birkbeck, University of London, London, United Kingdom, [4]Department of Philosophy, University of York, York, United Kingdom

There is a long tradition in philosophy and literary criticism of belief in the social and moral benefits of exposure to fiction, and recent empirical work has examined some of these claims. However, little of this research has addressed the textual features responsible for the hypothesized cognitive effects. We present two experiments examining whether readers' social and moral cognition are influenced by the *perspective* from which a narrative is told (voice and focalization), and whether potential effects of perspective are mediated by transportation into the story or by identification with the protagonist. Both experiments employed a between-subjects design in which participants read a short story, either in the first-person voice using internal focalization, third-person voice using internal focalization, or third-person voice using external focalization. Social and moral cognition was assessed using a battery of tasks. Experiment 1 (N = 258) failed to detect any effects of perspective or any mediating roles of transportation or identification. Implementing a more rigorous adaptation of the third-person story using external focalization, Experiment 2 (N = 262) largely replicated this pattern. Taken together, the evidence reported here suggests that perspective does not have a significant impact on the extent to which narratives modulate social and moral cognition, either directly or indirectly via transportation and identification.

Keywords: morality, focalization, perspective, identification, social cognition, fiction, narrative

OPEN ACCESS

Edited by:
Emar Maier,
University of Groningen, Netherlands

Reviewed by:
Roel M. Willems,
Radboud University Nijmegen, Netherlands
Hannah De Mulder,
Leiden University, Netherlands

***Correspondence:**
Lena Wimmer
lena.wimmer@ezw.uni-freiburg.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *29 September 2020*
Accepted: *16 December 2020*
Published: *15 February 2021*

Citation:
Wimmer L, Friend S, Currie G and Ferguson HJ (2021) Reading Fictional Narratives to Improve Social and Moral Cognition: The Influence of Narrative Perspective, Transportation, and Identification. Front. Commun. 5:611935. doi: 10.3389/fcomm.2020.611935

INTRODUCTION

Narrative fictions in various media and genres have long been regarded by humanistic scholars as educative, either through their embodiment of moral principles (Johnson, 1750) or their capacity to make us more sensitive to the needs and outlooks of others (Nussbaum, 1990; Nussbaum, 1995). Only very recently have such claims been subject to empirical test, where there has been a particular focus on the relation between narrative fiction and improvements in social cognition (overview: Dodell-Feder and Tamir, 2018). However, narrative fiction is a broad and heterogeneous category, raising the question of whether and why some texts might have more influence than others. In this paper we address this question by examining the role played by a specific textual feature, narrative perspective, on social and moral cognition.

Social cognition is defined as the perception, interpretation and use of social information, and comprises a set of abilities that we draw on daily to manage social relationships in a variety of situations (Fiske and Taylor, 2013). It includes affective empathy (i.e., tuning into other peoples' feelings or thoughts; Baron-Cohen and Wheelwright, 2004; Singer and Klimecki, 2014), Theory of Mind (ToM; understanding others' mental states; Wellman et al., 2001), and emotion recognition (i.e., the perception of others' emotional states; Brühne, 2005; Frith and Frith, 1999). The promotion of social cognition seems crucial not only in view of empirical evidence that understanding others facilitates everyday social interactions (e.g., Watson et al., 1999; Leppänen and Hietanen, 2001; Findlay et al., 2006), but also in light of evidence suggesting that deficits in both affective empathy and ToM are associated with mental health disorders, such as autism spectrum disorder (Baron-Cohen, 2000; Hobson, 2007) and schizophrenia (Lee, 2007; Bora et al., 2009). Furthermore, the cultivation of affective empathy and ToM are often regarded as essential for moral education (Dolby, 2012; Nussbaum, 1995; for a contrary view see Bialystok and Kukar, 2018). In recent years, researchers have become increasingly interested in testing whether social and moral cognition can be improved by reading fictional narratives (Calarco et al., 2017). In this paper we address this question, but we go further by testing whether effects on social and moral cognition are modulated by the narrative perspective in a fictional story.

Many empirical researchers favor the idea that reading narratives is associated with enhanced social cognition. Several lines of thought make that a plausible view. Stories are typically about characters and their social interactions; hence readers must apply their social cognition, including affective empathy and ToM, in order to understand narratives (e.g., Calarco et al., 2017; Salem et al., 2017; Deane et al., 2019; Mar, 2018a; Mar, 2018b). Furthermore, the social content of (fictional) narratives frequently broadens the scope of social information individuals are exposed to by describing experiences they would not have in real life, or by presenting events from novel perspectives (Calarco et al., 2017). In addition written narratives provide a safe environment to practice social cognition since, unlike in real life, readers can re-read passages several times in order to make sense of social situations, and misunderstandings do not result in adverse consequences for the reader or anyone else (Mar and Oatley, 2008). Tentative support for the proposal that reading stories promotes social cognition comes from two meta-analyses that have summarized effects of reading short fictional narratives on social cognition (Dodell-Feder and Tamir, 2018) and associations of lifetime exposure to fictional narratives with social cognition (Mumper and Gerrig, 2017). However, the aggregate effect sizes reported in both meta-analyses are small, with some experiments failing to detect any effects of reading short fictional stories at all (e.g., Panero et al., 2016; Samur et al., 2018). This inconsistent pattern of findings might be explained by the fact that different studies focus on different categories of text (e.g., fiction *vs* nonfiction, or literary *vs* popular fiction) and/or psychological mechanisms (e.g., transportation, imagery generation) presumed to be responsible for effects on social cognition (Deane et al., 2019). Texts in different categories differ along multiple dimensions, leaving it unclear which features might be responsible for any effects; and psychological mechanisms may also be triggered by a variety of text characteristics. To find out whether and why certain fictional narratives enhance social cognition, a targeted investigation of text features and their relationship to psychological mechanisms is needed.

Two psychological mechanisms that have been proposed to underlie narrative-based social cognitive benefits are identification and transportation (see Consoli, 2018). Identification refers to the process of taking a character's perspective in light of that character's beliefs, values, and goals (Van Krieken et al., 2017). Transportation[1] is "the experience of being imaginatively carried away into the story world" (Gerrig, 1993, as cited in Deane et al., 2019). Although transportation and identification are empirically distinguishable concepts (Tal-Or and Cohen, 2010; Tal-Or and Cohen, 2016), the processes seem to work synergistically, so that transportation facilitates identification and vice versa (Calarco et al., 2017). In particular, transportation is thought to reduce the psychological distance between readers and story characters, which in turn facilitates the reader's ability to take the character's perspective, share their emotions, and understand their (inter-)actions (Calarco et al., 2017; Consoli, 2018). Both identification and transportation appear to enhance social cognition. Identification is itself an exercise of mentalizing/ToM, while empirical evidence has supported a link between transportation into a narrative and enhanced social cognition beyond the story world (Johnson, 2012; Johnson, 2013; Bal and Veltkamp, 2013; Walkington et al., 2019). Therefore, we have good reason to think that transportation and identification mediate the relationship between reading narratives and social cognition.

Transportation and identification are themselves mediated, at least in part, by features of the text. Previous work that has focused on the role of textual determinants has highlighted the importance of "privileged access to the perceptions, evaluations, and goals of a character" as a central mechanism through which readers are transported into the protagonists' world and identify themselves with characters (van Krieken et al., 2017, p.4). Writers can vary this level of access by altering the perspective or point of view from which a narrative is told (van Krieken, 2018). Narrative perspective is a multi-dimensional construct (cf. Hühn et al., 2009). Two dimensions are especially relevant to how the inner life of a character is presented; Genette, 1980 called these *voice* and *focalization*. Voice concerns who tells the story, which in most narratives will be a first-person narrator or a third-person narrator, and is typically reflected in the grammatical use of certain pronouns. Focalization concerns the perspective or point of view of the narrator and specifies restrictions on access to information. For instance, when a story is narrated in the first person by a particular character, the narration is often restricted to what the character knows, including their thoughts and feelings. However, narrative voice and focalization are independent. Third-person narrators may tell a story from an unrestricted ("omniscient") point of view, with broad access; or a more restricted perspective, either through what Genette called *internal focalization* (where the narrator has access to what a particular character knows and experiences) or *external focalization* (where the narrator does not have access to the character's thoughts or experiences). The following passages, both in third-person voice, illustrate the

[1]Some researchers (e.g., Calarco et al., 2017) use the term absorption synonymously.

distinction between internal focalization (1) and external focalization (2):

(1) Sarah looked at her daughter. She knew she had been like this as a child, sometimes. But why did Amy *always* have to act this way? Couldn't she see how much it hurt?
(2) Sarah looked at her daughter. She frowned, and said, "I was like you sometimes when I was a child. But why do you *always* have to act this way? Can't you see how much it hurts?"

In (1), readers have direct access to the mental life of the protagonist (Sarah), whereas in (2) the scene is described from the point of view of an external observer without direct access to Sarah's mental life. Because focalization is what determines the extent to which the narrator offers insight into a character's inner life, it appears more relevant to transportation and identification than voice.

Although Genette's distinction between narrative voice and focalization has initiated an animated discussion among narratologists (see e.g., Fludernik, 2001; Margolin, 2009; Currie 2010), empirical investigations have almost exclusively studied manipulations of voice, assuming that this is sufficient for testing perspective (e.g., Van Lissa et al., 2016; Hartung et al., 2017b; Creer et al., 2019; Kim et al., 2019; Samur et al., 2020). So far, only one empirical study has investigated the effects of focalization or perspective as distinct from voice. Salem et al. (2017) compared first-person narration (internally focalized); third-person external focalization; and two forms of third-person internal focalization: psycho-narration, in which the narrator reports the thoughts of the protagonist (e.g., "She worried about her children"), and free indirect discourse (FID), which stays closer to the protagonist's subjectivity (e.g., "What would happen to her children?"). (Example (1) above contains both.). Results indicated that focalization had effects on perspective-taking independently of narrative voice, suggesting that focalization is indeed more important for processes relevant to identification and, hence, possibly social cognition in the real world.

However, targeted investigations are needed that manipulate voice and focalization separately and assess both proximal effects, i.e., processes of transportation and identification with characters, and distal outcomes, i.e., general social cognition. In sum, (internal vs. external) focalization is thought to exert a greater influence on transportation and identification than (first- *vs* third-person) voice. As a consequence of this influence, internally focalized narratives are expected to lead to greater benefits for social cognition than externally focalized narratives, regardless of narrative voice.[2]

A related question is whether manipulations of perspective might have an effect on moral cognition. Theorists in the humanities have traditionally argued that reading (fictional) stories has the capacity to generate moral improvement (e.g., Nussbaum, 1990; Nussbaum, 1995), but there is a paucity of experimental evidence to support this claim. The few existing investigations (Johnson et al., 2013; Koopman, 2015; Kidd and Castano, 2019) provide somewhat suggestive findings; however none of them has examined the ways in which textual features might influence moral cognition. Reading narratives might be assumed to benefit moral cognition via at least two routes.

The first route is via an increase of social cognition, as outlined above (Koopman, 2015). Nussbaum (2001) suggests that "tragic fictions promote extension of concern by linking the imagination powerfully to the adventures of the distant life in question" (p. 352). The idea is that through empathetic sharing of characters' mental states, readers enlarge their capacity for imaginative contact with the thoughts, feelings and intentions of others in real-life. This may have multiple effects: increasing readers' compassion, being more aware of the impact of their own actions on others' welfare, and, by providing them with better information about motives and intentions, allowing them to make better, more sophisticated moral judgements about the actions of others (Killen et al., 2011; Ugazio et al., 2014). Insofar as moral cognition is assumed to improve as a result of narrative-based improvements in social cognition, which are expected to rely on internal (*vs* external) focalization, internal focalization should also generate improvements in moral cognition. It is, however, agreed that improving social-cognitive abilities does not automatically lead to morally better outcomes; while empathy is often associated with prosocial virtues, it can be used for manipulation and deception (e.g. Bloom, 2017; Breithaupt, 2018; Bubandt and Willerslev, 2015; Vermeule, 2010; see also Sutton et al., 1999). Therefore, improvement of social cognition cannot be considered a sufficient condition of moral improvement – enhanced social cognition does not guarantee moral progress. Yet it may be that readers who are appropriately motivated can utilize the social knowledge gained from narratives to become morally better people. If so, the effect is expected to be modulated by focalization.

The second route is more direct, via observational learning (Johnson et al., 2013; Mumper and Gerrig, 2019; Black and Barnes, 2020a). According to this view, readers can learn morally positive attitudes and behaviors when a story character is rewarded for a morally positive action or punished for a morally negative action; and identification with the character can increase the likelihood of executing an observed behavior. Since identification is thought to depend on internal (*vs*. external) focalization, this is another way in which manipulating focalization can generate changes in moral cognition. If readers internalize such pattern of reward blindly, however, without gaining insight into universal principles of ethical conduct, they could be equally likely to learn morally negative attitudes and behaviors when a story character is rewarded for a morally

[2]Whilst we predict that the more direct access to characters' inner life bestowed by the internal perspective facilitates transportation, identification, and ultimately social cognition, other researchers have assumed that having additional information about mental states reduces (or does at least not enhance) readers' understanding of those states. However, these studies have focused on explicit information provided by a narrator. For instance, Kotovych et al., 2011 observed support for their hypothesis that character transparency, that is, the extent to which the character's actions and attitudes are clear and understandable, is reduced when information about a character is explicitly stated compared to when readers must infer such details. Furthermore, results of Peskin and Astington (2004) suggest that social cognitive abilities are promoted if readers have to work out the mental lives of fictional characters themselves, compared to when mental states are spelled out for the readers.

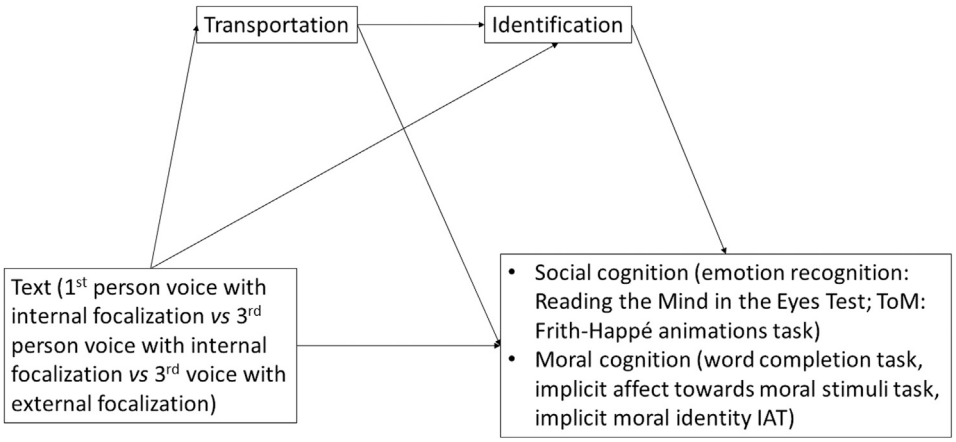


FIGURE 1 | Predicted relationships between the variables of interest and their operationalization.

negative action or punished for a morally positive action. Reading narratives would therefore lead to a "moral boundary erosion" (Black and Barnes, 2020a). Yet there is some reason to believe that reading narratives operates as a "moral laboratory" (Hakemulder, 2000), leading to increased moral sensitivity rather than the opposite. Typically, readers prefer narratives in which good behavior is presented as preferable to bad behavior (Raney, 2004). Also, readers have been found to report an inability to imaginatively engage with immoral fictional worlds, a phenomenon known as "imaginative resistance" (overview: Black and Barnes, 2020b). In another study by Vezzali et al. (2015), reading a passage from *Harry Potter* about prejudice improved attitudes toward immigrants for those who identified with Harry Potter, whereas there was no effect for those who identified more with Voldemort, the villain; this suggests that identification is more likely to improve than impair moral cognition. For these reasons, reading internally (*vs.* externally) focalized narratives is expected to lead not only to greater improvements in social cognition, but also in moral cognition, in both cases via increases of transportation and identification.

Here we report two experiments that investigated whether social and moral cognition are affected by the perspective from which a fictional narrative is told. We considered both voice (first- *vs* third-person) and point of view (internal *vs*. external focalization). We also looked at potential mediating effects of transportation into the story world and identification with the protagonist. In Experiment 1, we employed a between-subjects design in which participants read a complete short story by an established author either in its original version, i.e., a third-person narrative using internal focalization (specifically, FID); an adapted version telling the narrative in first-person voice using internal focalization; or an adapted version telling the narrative in third-person voice using external focalization, as pre-registered in the Open Science Framework. Social cognition was assessed in terms of emotion recognition, as indicated by the Reading the Mind in the Eyes Test-Revised (Baron-Cohen et al., 2001), and ToM, as indicated by the Frith-Happé animations task (Abell et al., 2000). Two basic self-concepts that have been associated with moral behavior are, respectively, communion and agency (Bakan, 1966). Communion is associated with allocentric behaviors, since it is related to cultivating social relationships and pro-social traits including cooperation, while agency is associated with egocentric behaviors, since it is linked with distancing the self from others and anti-social traits including assertiveness (Bakan, 1966; Bartz and Lydon, 2004). We assessed these two self-concepts implicitly using a word-fragment completion task, which indicated ease of access to the respective concepts (Bartz and Lydon, 2004). In addition, the Immediate Affect towards Moral Stimuli task reflected participants' affective reactions towards morally positive/negative stimuli, which have been associated with guilt feelings in a moral dilemma, and with emotional reactions to/rejection of an unfair offer (Hofmann and Baumert, 2010). Finally, an Implicit Association Test (IAT) measured participants' moral *vs.* immoral self-concept, which predicts moral actions such as honest behavior despite negative consequences (Perugini and Leone, 2009).

Experiment 2 used the same basic design and internal focalization text versions as Experiment 1, but the version with external focalization was edited even further to strengthen the focalization manipulation. **Figure 1** provides a schematic representation of the predicted relationships between the variables of interest and how these were operationalized. Note that although transportation and identification likely influence each other, the interrelation between the two variables is not at the core of the research questions addressed here, and thus we did not make specific hypotheses about how a combination of transportation and identification might mediate the relationship

TABLE 1 | Overview and justification of dependent variables, their operationalization, and their use in previous experiments on narrative-based benefits for social and moral cognition.

Outcome		Justification for investigating outcome	Operationalization	Justification for using operationalization	Use of operationalization in previous investigations in the field
Social cognition	Emotion recognition	Mar (2018a) predicts that stories containing accurate social content can teach lessons about human psychology including emotional expression	Reading the Mind in the Eyes Test (RMET; Baron-Cohen et al. (2001))	• Oakley et al. (2016) suggested that the RMET measures emotion recognition • This task has been applied widely in previous research, which facilitates comparison with existing findings	Yes e.g., Djikic et al. (2013), Kidd and Castano (2013), Kidd and Castano (2019), Black and Barnes (2015), Liu and Want (2015), Kidd et al. (2016), Panero et al. (2016), and Samur et al. (2018)
	Theory of Mind (ToM)	Mar (2018a) suggests that stories that provide access to protagonists' inner life exercises ToM	Frith-Happé animations task by White et al. (2011)	• This task draws on people's tendency to attribute human psychology to silent moving shapes • This task has been used as a reliable measure of ToM in adults (White et al. (2011)): High levels of ToM are suggested if participants correctly recognize mental interactions between animated shapes	No
Moral cognition		Theorists in the humanities have traditionally argued that reading (fictional) stories has the capacity to generate global moral improvement across a range of components of moral cognition (e.g., Nussbaum (1990), Nussbaum (1995))	Word completion task introduced by Bartz and Lydon (2004)	• This task provides an implicit way of assessing moral cognition that reduces susceptibility to social desirability • It indicates the ease of access to morally relevant concepts (Bartz and Lydon (2004)), namely communion (related to cultivating social relationships and pro-social traits) and agency (linked with distancing the self from others and anti-social traits)	No
			Implicit affect towards moral stimuli task (IAMS task; Hofmann and Baumert (2010))	• This task reflects affective reactions towards morally positive/negative stimuli • Performance in this task has been associated with guilt feelings in a moral dilemma, and with emotional reactions to/rejection of an unfair offer (Hofmann and Baumert (2010))	No
			Implicit moral identity IAT	• This task measures moral *vs.* immoral self-concept • Performance in this task predicts moral actions such as honest behavior despite negative consequences (Perugini and Leone (2009)) • This task has proven to be a better predictor of real-life action than measures of explicit attitudes (Perugini and Leone (2009))	No

between text reading and social/moral cognition. **Table 1** presents an overview and justification of dependent variables, their operationalization, and their use in previous experiments on narrative-based benefits for social and moral cognition.

The following hypotheses were tested in both experiments:

(1) Measures of social and moral cognition would be higher after reading a narrative told through internal focalization than external focalization.
(2) The voice (first- *vs.* third-person) in which a narrative is told would not *per se* influence social cognition. Therefore, social cognition would not differ after reading narratives told in different voices as long as focalization was held constant between stories.
(3) Reading narratives achieves benefits in social and moral cognition via transportation into the narrative and identification with characters. Thus, we expected effects of story reading on social and moral cognition to be mediated by transportation and identification with the story's protagonist.

EXPERIMENT 1

Materials and Methods

All methodological procedures were approved by the Research Ethics Committee of the School of Psychology at the University of Kent, United Kingdom, prior to commencement, and pre-registered on the Open Science Framework, https://osf.io/wpd2c. The experiment followed a between-subjects design involving one randomized factor with three levels of text (first-person voice with internal focalization *vs* third-person voice with internal focalization *vs* third-person voice with external focalization). First-person voice with external focalization was not realized since this would have required an unacceptable level of distortion to the original story, either by removing the content referring to the protagonist's mental life altogether, which would have made this story version much shorter than the remaining versions, or by replacing this type of content with something else, which again would have substantially reduced comparability with the three other story versions. Furthermore, narratives using first-person voice with external focalization can occur, but are extremely rare (Genette, 1983; Edmiston, 1989); thus, this textual condition does not seem of high practical relevance and would likely have come across as artificial to readers.

Participants

Participants were recruited through Prolific Academic and were paid £6.00. All participants had English as their primary language, and provided written informed consent before data collection. We used the software program G*Power to conduct an a-priori power analysis, aiming for 0.95 power to detect a medium effect size of f = 0.25 at the standard 0.05 alpha error probability. This resulted in a total *N* of 251. Allowing for a dropout rate of approx. 33% (as observed in previous studies, e.g., Kidd and Castano, 2019), we

aimed to recruit a total *N* of 340 to yield the target of *N* = 252 (84 per group). Power analysis further showed that a total sample size of *N* = 252 would have a power of 0.99 to detect mediation, given a medium size effect and 5% significance level. This sample size would have a power of 0.61 to detect a small effect.

Participants were excluded from analyses if they met at least one of the following criteria: 1) First language was not English; 2) reading time was below 360s[3]; 3) failed a check of text comprehension by indicating that the story was told by an anonymous narrator after reading the internal first-person version of the story, OR that the story was told by the mother, Marjorie (the protagonist) after reading the external third-person version of the story; 4) more than two mock authors were selected in the Author Recognition Test–Genres (Mar and Rain, 2015; cf. Wimmer et al., submitted)[4]; 5) failed an attention check item that was interspersed within a questionnaire (Transportation Scale; Green and Brock, 2000).

The target sample was reached after 347 volunteers. When exclusion criteria were applied, 258 participants remained in the final sample, 87 of whom had read the first-person narrative with internal focalization (54.0% female, mean age = 36.55, SD of age = 14.01), 87 of whom had read the third-person narrative with internal focalization (52.9% female, mean age = 33.69, SD of age = 10.41), and 84 of whom had read the third-person narrative with external focalization (56.0% female, mean age = 35.00, SD of age = 12.16). See **Figure 2** for a schematic of the flow of participants through the experiment.

Reading Stimuli

Three versions of Rose Tremain's "The Closing Door", a complete short story about a widow seeing off her 10-year-old daughter to boarding school, served as reading stimuli. This text was chosen for several reasons. First, its length, i.e., approximately 3000 words, was considered both long enough to evoke the effects under investigation and short enough to be read in its entirety during an experimental session. Second, the story is by a prize-winning literary author in a highly regarded collection, and previous research has suggested that literary stories might be especially good at affecting social cognition (Kidd and Castano, 2013). Third, it describes both 'outer events' (physical happenings) and 'inner views' (how the events are perceived by the protagonist), so that the content was amenable to manipulations of voice and focalization. In contrast, manipulating the voice and focalization of stories dealing exclusively with outer events would probably have a low impact on the reader, while changing a story about the inner life of a character to external focalization would obliterate the vast majority of the content. Finally, we anticipated that the farewell theme of the story was relatively familiar to all participants, meaning that they would in principle be able to identify with the protagonist, even if inter-individual variation was expected.

The original story represented the third-person narrative with internal focalization (FID). In addition to that, two adaptations were created by SF. In one version, the story is told by the protagonist in the first person with internal focalization. In the other version, the story is told by an anonymous third-person narrator using external focalization. Word count was comparable across the three versions of the story (first-person narrative with internal focalization: 3127 words; third-person narrative with internal focalization: 3058 words; third-person narrative with external focalization: 2962 words). **Table 2** exemplifies the use of perspective and focalization in the three texts; full stimuli are available on the pre-registration pages, https://osf.io/wpd2c.

Assessment Tasks

Lifetime Exposure to Print

The Author Recognition Test–Genres (Mar and Rain, 2015) provided an indicator of reading habits. Participants were asked to accurately recognize the names of 110 fiction and 50 non-fiction authors (targets) among names of 40 non-authors (foils). A fiction sub-score was calculated based on the number of correctly identified fiction authors (possible range: 0–110).

Trait Empathy

In the Empathy Quotient (Baron-Cohen and Wheelwright, 2004), 40-item version, respondents are asked to indicate the degree to which they agree with statements, such as "I can easily tell if someone else wants to enter a conversation", using a 4-point rating scale that ranges from strongly disagree to strongly agree. A sum score was calculated for each respondent, and higher scores indicating greater levels of empathy (possible range: 0–80). Internal consistency in the current sample was Cronbach's $\alpha = 0.91$.

Identification

Participants' identification with the protagonist was assessed using a scale adapted from Cohen (2001) that consisted of the following items, which were rated on a 7-point Likert scale ranging from strongly disagree to strongly agree, and used to calculate a sum score (possible range: 8–56):

- I was able to understand the events in the story in a manner similar to that in which the mother, Marjorie[5], understood them.
- I think I have a good understanding of the mother.

[3]The reading time cut-off was based on an average word reading speed of 239 ms, reported by Benjamin and Gaab (2012). This yields an average reading duration of 731 s for each text excerpt. Thus, the cut-off of 360 s is just below half of the time an average reader would need to read the text. This criterion has previously been applied by Wimmer, Currie, Friend, and Ferguson (submitted).

[4]Using guesses within Author Recognition Tests as an exclusion criterion is a common strategy in the field (cf. Kidd and Castano, 2013; Kidd and Castano, 2019; Kidd et al., 2016; Panero et al., 2016; Samur et al., 2018) as it can indicate a rather negligent participation style. Excluding participants with more than two guesses is more rigorous than measures applied in previous experiments (e.g., Kidd and Castano, 2019) since Moore and Gordon (2015) suggested a higher penalty for selecting mock authors than originally recommended. Six out of 270 participants (2.22%) were excluded due to this criterion (cf. **Figure 2**).

[5]The character's name, Marjorie, was mentioned in the first item only to ensure that respondents associated "the mother" with this character.

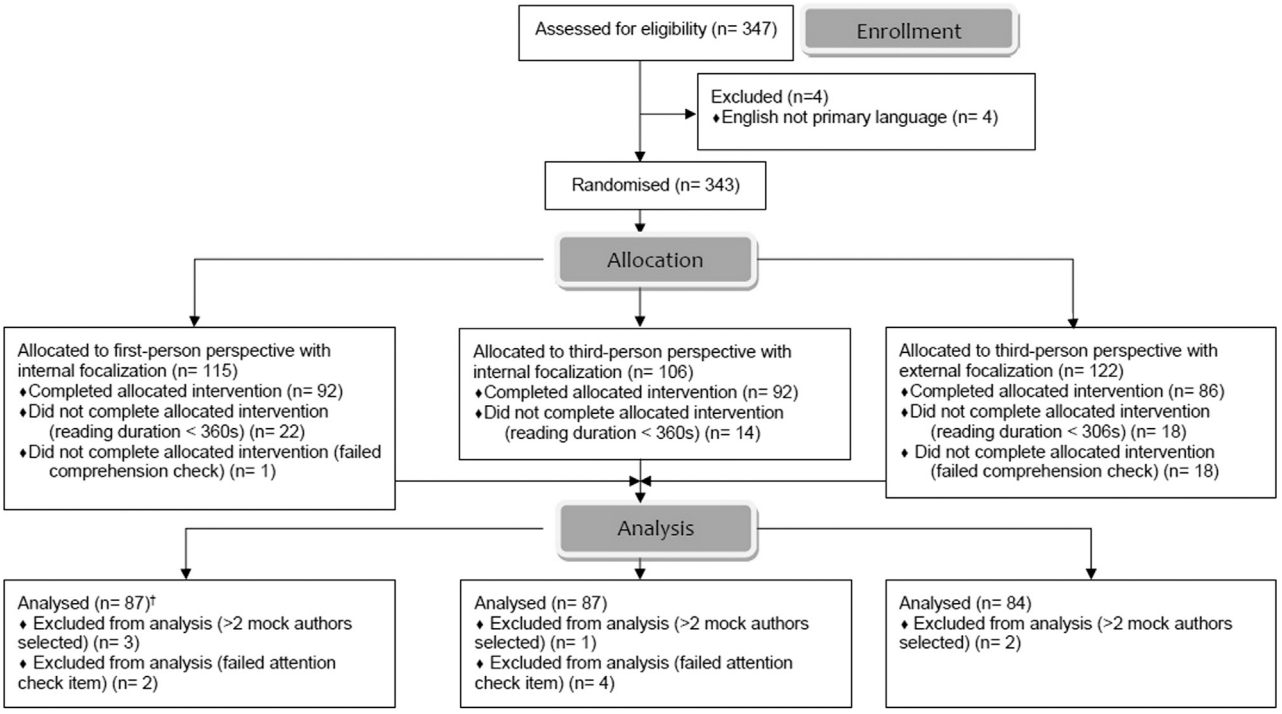


FIGURE 2 | Flow of participants through Experiment 1.

- I tend to understand the reasons why the mother did what she did.
- While reading the story, I could feel the emotions the mother portrayed.
- During reading, I felt I could really get inside the mother's head.
- At key moments in the story, I felt I knew exactly what the mother was going through.
- During reading, I wanted the mother to succeed in managing her emotions.
- When the mother succeeded I felt joy, but when she failed, I was sad.

In the present sample, internal consistency of this scale was Cronbach's $\alpha = 0.87$.

TABLE 2 | Excerpts of text stimuli of Experiment 1.

First-person narrative, internal focalization	third-person narrative, internal focalization	third-person narrative, external focalization
And it was coming nearer, nearer, the moment when I would have to unwind Patience's arms from round my waist and lead her forwards to the barrier. I tried to stand a bit more upright, but the weight of Patience clinging to me was implacable, as though I had been roped to the ground. And I thought, I am bent like an old person, bent down by the gravity of love [..]. I knew that this now risked to become what Tim would have called 'a scene', and that the other parents would pity me, or even despise me, for not crushing it the minute it started, so, with a strength that surprised me, I grabbed Patience by the fist that had struck me and turned her round to face the trains and the great vaulted station roof above them, still black from the years of war.	And it was coming nearer, nearer, the moment when Marjorie would have to unwind Patience's arms from round her waist and lead her forwards to the barrier. She tried to stand a bit more upright, but the weight of Patience clinging to her was implacable, as though she had been roped to the ground. And she thought, I am bent like an old person, bent down by the gravity of love [..]. Marjorie knew that this now risked to become what Tim would have called 'a scene', and that the other parents would pity her, or even despise her, for not crushing it the minute it started, so, with surprising strength, she grabbed Patience by the fist that had struck her and turned the sobbing child round to face the trains and the great vaulted station roof above them, still black from the years of war.	And the moment was coming nearer when Marjorie would have to find a way to unwind Patience's arms from round her waist and lead her forwards to the barrier. Marjorie was making a visible effort to stand a bit more upright, but the weight of Patience clinging to her seemed to make this impossible, pulling her back to the ground. She was bent like an old person, intensifying her air of misery [..]. This now risked to become 'a scene'. It is common knowledge that any mother or father who failed to crush a scene the minute it started would be pitied, or even despised, by other parents. Perhaps Marjorie understood this, because suddenly and with surprising strength, she grabbed Patience by the fist that had struck her and turned the sobbing child round to face the trains and the great vaulted station roof above them, still black from the years of war.

Familiarity with Subject Matter

Previous experience with the subject matter of the stories was assessed using the following items:

- Reflecting on your own life experiences, how familiar were the events in the story to you? (7-point rating scale ranging from *very unfamiliar* to *very familiar*)[6]
- Have you ever attended boarding school?
- Has anyone in your family attended boarding school? (if so, who?)
- As a child were you ever sent away from your family for a long period of time?
- Are you a parent? (if so, do you have a partner or single?)
- Have you ever experienced a separation from a loved one? (if yes, please explain)

Transportation

Transportation was operationalized using the 12-item scale developed by Green and Brock (2000). High levels of transportation are indicated by a high sum score (possible range: 12–84). In the present sample internal consistency was Cronbach's $\alpha = 0.73$.

Emotion Recognition

We measured emotion recognition using the Reading the Mind in the Eyes Test-Revised (Baron-Cohen et al., 2001). Better emotion recognition skills were indexed by a high relative frequency of correct responses (possible range: 0–1). Although the Reading the Mind in the Eyes Test-Revised was initially introduced as a measure of mentalizing, more recently it has been proposed that it may rather reflect emotion recognition (Oakley et al., 2016), which is the operationalization we adopt in the present experiment.

Implicit Morality

First, a word completion task (Bartz and Lydon, 2004) assessed self-concepts in terms of agency and communion. Six out of 13 word fragments (e.g., _ _ _ erior) could be completed with agency words (e.g., 'superior') or non-agency words (e.g., 'interior') and six word fragments (e.g., _ind) could be completed with communion words (e.g., 'kind') or non-communion words (e.g., 'mind'); one word fragment could be completed with either an agency or communion or non-agency-non-communion word, i.e., c_ _ p_ _ _ tive ('competitive', 'cooperative', 'comparative'). Responses were coded as either a "hit" or a "miss" for the target word. An agency score (possible range: 0–7) was calculated by summing the hits to the target words: superior, confident, active, competitive, boast, greedy, and hostile. A communion score (possible range: 0–7) was computed by summing the hits to target words: kind, warm, gentle, cooperative, whiny, nag, and dependent. Higher implicit morality is indicated by a low agency and a high communion score, respectively.

Second, in the Implicit Affect towards Moral Stimuli task, we followed the experimental procedure and stimuli applied by Hofmann and Baumert (2010). In each trial, participants were asked to categorize a Chinese pictograph as "pleasant" or "unpleasant" using two response keys on the keyboard. Shortly before the Chinese pictograph was presented, a moral prime (e.g., an elderly couple walking arm-in-arm, or a man directing a gun into the camera), or control picture (e.g., a lightening striking a mountain side) appeared for 100 ms. It is assumed that the affective reaction to the moral primes presented is misattributed to the Chinese pictograph, thus influencing the response. We used ten pictures of morally positive behaviors and ten pictures of morally negative behaviors as moral primes. As comparison pictures, we included ten non-moral pictures of positive valence, as well as ten non-moral pictures of negative valence. Following the procedures of Hofmann and Baumert (2010), responses exceeding a threshold of 2000 ms (3.28% of all responses) or falling below 350 ms (6.07% of all responses) were considered outliers and excluded from analyses. To achieve an indicator of immediate affect towards moral stimuli, the individual difference index of the Implicit Affect towards Moral Stimuli (possible range: 0–1) was calculated for each participant (cf. Hofmann and Baumert, 2010; proportion of "positive" judgements on trials in which a Chinese pictograph was preceded by a positive moral prime *minus* percentage of "positive" judgements on trials in which a Chinese character was preceded by a negative moral prime). To control for general, morally unrelated affect, the individual differences index of the IACS (immediate affect towards control stimuli) was computed (i.e., proportion of "positive" judgements on trials with positive morality-irrelevant primes *minus* the proportion of "positive" judgments on trials with negative morality-irrelevant primes; possible range: 0-1).

Third, we applied an Implicit Association Test (IAT) of implicit moral identity (implicit moral identity IAT), replicating the experimental procedure and stimuli in Perugini and Leone (2009), and following the standard IAT sequence (Greenwald et al., 1998). The target categories were "Moral" (represented by the stimulus words honest, sincere, faithful, modest, altruist) *vs* "Immoral" (represented by the stimulus words deceptive, arrogant, dishonest, cheater, pretentious), and the paired categories were "Me" (represented by the stimulus words I, me, myself, self, my) *vs* "Others" (represented by the stimulus words them, they, others, your, you). As in Perugini and Leone (2009), implicit moral identity was indexed by the D6 measure (Greenwald et al., 2003), which is calculated as the mean latency in the "immoral-me" block *minus* the mean latency in the "moral-me" block, divided by the individual standard deviation of latencies across "immoral-me" and "moral-me" blocks. Responses with latencies below 400 ms or above 10,000 ms were excluded from analysis, and latencies of errors were replaced by the block mean of correct-response latencies plus 600 ms. Higher scores express a stronger implicit moral self-concept (possible range: -2-2).

Theory of Mind

A revised version of the Frith-Happé animations task (White et al., 2011) was used as an indicator of ToM. Participants

[6]The 6-item battery was used for exploratory purposes; however only the first item was included in statistical analyses, so that the possible range for the familiarity score was 1–7.

TABLE 3 | Experiment 1: Descriptive statistics for each dependent measure in each experimental group, and ANCOVA results for the main effect of text.

Dependent measure		First-person narrative with internal focalization		third-person narrative with internal focalization		third-person narrative with external focalization		ANCOVA: main effect of text			
		n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	df	*F value*	*p value*	η^2_p
Word completion task	Agency score (possible range: 0–7)	86	3.28 (1.27)	85	3.25 (1.25)	80	3.33 (1.24)	2, 245	0.13	0.879	0.001
	Communion score (possible range: 0–7)	86	1.67 (1.24)	85	2.09 (1.49)	80	1.86 (1.27)	2, 245	2.22	0.111	0.018
Frith-Happé animations task	MCQ 1: Accuracy sum score (possible range: 0–4)	86	2.01 (1.09)	85	1.89 (1.06)	80	1.95 (1.07)	2, 245	0.33	0.720	0.003
	MCQ 2: Accuracy sum score (possible range: 0–8)	78	3.17 (1.88)	76	2.97 (1.87)	74	2.96 (1.70)	2, 222	0.22	0.800	0.002
RMET: Relative frequency of correct responses (possible range: 0–1)		85	0.70 (0.14)	85	0.70 (0.17)	79	0.70 (0.15)	2, 243	0.00	0.997	0.000
Implicit affect towards moral stimuli score (possible range: 0–1)		72	0.13 (0.20)	73	0.14 (0.21)	70	0.18 (0.25)	2, 208	0.00	0.998	0.000
Implicit moral identity IAT: D6 (possible range: -2-2)		70	0.66 (0.35)	73	0.70 (0.35)	70	0.71 (0.30)	2, 207	0.28	0.755	0.003
Transportation Scale (possible range: 12–84)		86	58.13 (8.76)	85	56.34 (9.23)	80	56.59 (9.40)	2, 245	1.69	0.188	0.014
Identification Scale (possible range: 8–56)		86	46.56 (6.65)	85	45.35 (7.23)	80	44.89 (7.35)	2, 245	2.08	0.127	0.017
Author Recognition Test – Genres: Fiction sub score (possible range: 0–110)		87	17.94 (16.18)	87	16.33 (15.85)	84	16.44 (15.90)				
Empathy Quotient (possible range: 0–80)		87	39.11 (15.04)	87	41.13 (12.46)	84	39.92 (12.87)				

Note. *MCQ, multiple choice question; RMET, reading the mind in the eyes test–revised; IAT, implicit association test.*

watched four video clips depicting animated triangles and interpreted the interactions between these shapes. Before presentation of the clips, participants were told that each video would contain either: *no interaction*, defined as no obvious interaction between the triangles, with movement appearing random; or *physical interaction*, defined as 'interaction between the triangles in which actions are directed toward each other in order to achieve specific goals'; or *mental interaction*, defined as 'an interaction between the triangles involving the manipulation of the emotions and thoughts of one triangle by the other'. Immediately after watching each video, participants categorized the type of interaction as either no interaction, physical interaction, or mental interaction by responding to a multiple choice question. Choosing "mental interaction" was coded "1", the remaining choices were coded "0", so that the sum score integrating responses to all four video clips had a potential range of 0 to 4. When mental interaction was chosen correctly, two further multiple-choice questions probed details about the feelings of each of the shapes, with one out of five response options being correct. Correct answers were coded "1", errors were coded "0", resulting in a four-video sum score with a possible range of 0 to 8.

Procedure

Participants completed all tasks on a computer, starting with a Qualtrics survey. After giving their informed consent to participate, respondents completed the Author Recognition Test–Genres and the Empathy Quotient. Next, participants were randomly allocated to one of the three reading conditions (first-person voice with internal focalization *vs* third-person voice with internal focalization *vs* third-person voice with external focalization). Immediately after reading, participants answered a comprehension question, "Who do you think is narrating the story you just read?", by selecting one of four response options: the mother (Marjorie); an anonymous narrator; other (please state); don't know/do not wish to answer. Subsequently, participants completed the Identification and Transportation Scales as well as the familiarity items. Next, they performed the word completion task, Frith-Happé animations task, and Reading the Mind in the Eyes Test-Revised. They were then redirected to the Inquisit platform to complete the Implicit Affect towards Moral Stimuli task and implicit moral identity IAT. In contrast to Qualtrics, Inquisit enables measurement of reaction times on millisecond level (De Clercq et al., 2003). Finally, participants were redirected to a Qualtrics survey to receive written debriefing. The entire experiment took 70 min to complete, on average.

Data Analysis

All analyses were pre-registered, and the full datasets are available on the Open Science Framework web pages, https://osf.io/6gsqn/.

We adopted the standard significance level of $p < .05$ for all inferential tests. Each task was analyzed separately, with ANCOVAs that include text (first-person internal *vs* third-person internal *vs* third-person external) as between-subjects

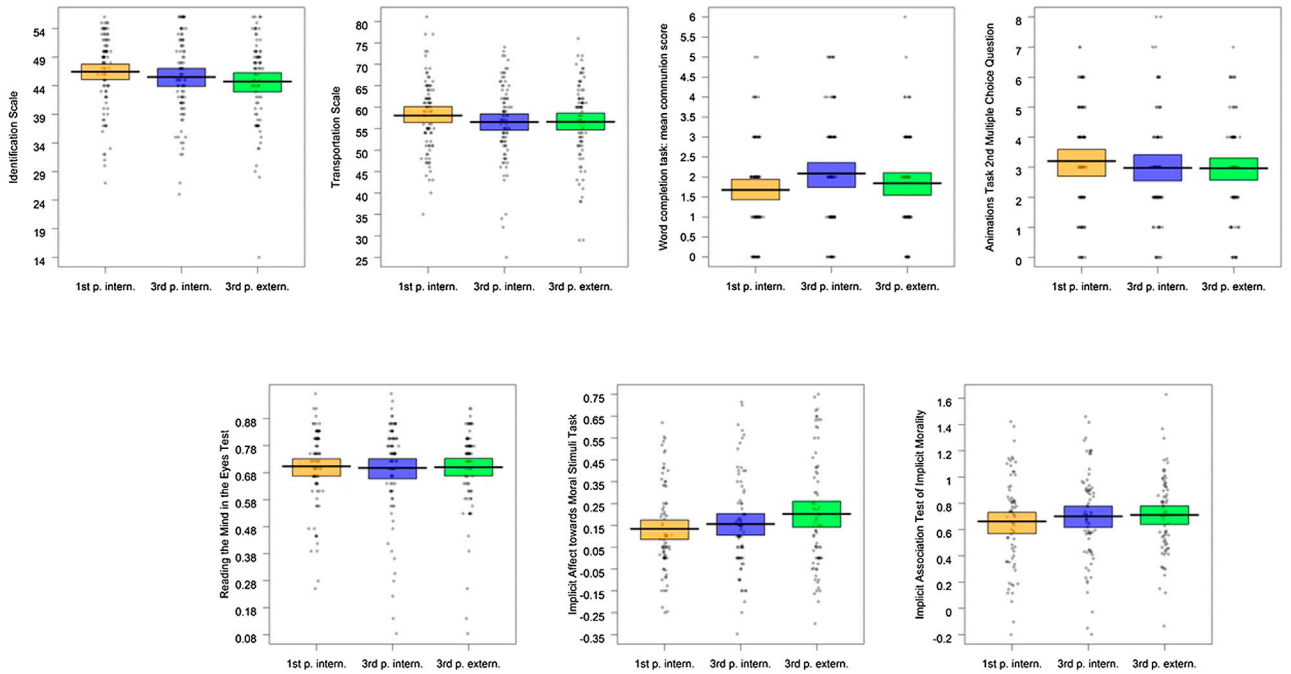


FIGURE 3 | Pirate plots for main outcome variables of Experiment 1.

predictor and the following variables as covariates: fiction sub-score of the Author Recognition Test–Genres, education level (operationalized as highest qualification), trait empathy (operationalized as Empathy Quotient sum score). These variables were included as covariates since education, expertise (i.e., lifetime exposure to narrative fiction as reflected by the Author Recognition Test–Genres fiction sub-score), and trait empathy are assumed to impact on transportation and identification (Consoli, 2018). For the Implicit Affect towards Moral Stimuli task, the IACS score was included as an additional covariate to control for morally irrelevant affect.

Mediation analyses were calculated using PROCESS macro (Model 6 with 2 mediators) in SPSS. One mediation was carried out for each indicator of social/moral cognition, transportation and identification were considered mediators, and text was the predictor. The mediation models did not include covariates since, firstly, this would have reduced statistical power; and secondly, inclusion of covariates was not vital for testing our hypothesis regarding the mediational role of transportation and identification.

Eighty-two missing items values - 81 from the Empathy Quotient, 1 from the Identification Scale – were imputed with the individual scale mean on the remaining items.

Results

Main Analyses

For each dependent measure, descriptive statistics and inferential statistics of the main effect of text are summarized in **Table 3**, and the key effects are plotted in **Figure 3**. Replicating previous research with the Implicit Affect towards Moral Stimuli task, participants overall were significantly more likely to judge a pictograph as pleasant after a positive moral prime than after a negative moral prime (M = 0.64 *vs* 0.46; t (299) = 12.12, $p < 0.001$, d = 0.68), and were significantly more likely to judge a pictograph as pleasant after a positive morality-irrelevant control stimulus than after a negative morality-irrelevant control stimulus (M = 0.70 *vs.* 0.45; t (296) = 16.28, $p < 0.001$, d = 0.95). This supports the underlying assumption of the Implicit Affect towards Moral Stimuli task that ratings are biased towards prime valence. Overall accuracy on the Reading the Mind in the Eyes Test-Revised and animations tasks was good (M = 70% and 1.94, respectively), and the positive mean D6 value (M = 0.69) in the moral IAT is consistent with previous research showing a preference for an implicit moral self-concept. Taken together, these data confirm that across the tasks, accuracy was comparable with previous studies, which suggests similar levels of compliance with instructions.

None of the ANCOVAs revealed a significant effect of text ($ps > 0.10$). Looking at the impact of covariates, Empathy Quotient sum score had a significant effect on the identification scale, F (1, 245) = 38.82, $p < 0.0001$, $\eta_p^2 = 0.137$, on the transportation scale, F (1, 245) = 34.74, $p < 0.0001$, $\eta_p^2 = 0.124$, on the D6 score in the implicit moral identity IAT, F (1, 207) = 4.31, $p = 0.039$, $\eta_p^2 = 0.020$, and on the agency score in the word completion task, F (1, 245) = 5.63, $p = 0.018$, $\eta_p^2 = 0.022$. Bivariate correlations demonstrated that Empathy Quotient was positively related to each of these outcomes ($0.141 < r < 0.347$, $ps < 0.038$).

Furthermore, the fiction sub-score of the Author Recognition Test–Genres had a significant effect on the communion score in the word completion task, $F(1, 245) = 4.98$, $p = 0.027$, $\eta_p^2 = 0.020$, and the IACS score had a significant effect on the Implicit Affect towards Moral Stimuli score, $F(1, 208) = 113.75$, $p < 0.0001$, $\eta_p^2 = 0.354$. Both covariates were positively correlated with the respective outcome ($rs = 0.138$ and 0.605, respectively, $ps < 0.025$). There were no further significant main effects ($ps > 0.07$).

Mediations were calculated to test whether a potential relationship between reading narratives and improved social cognition was mediated by identification with characters and transportation into the story. Because all 95% confidence intervals of the indirect effect of story reading on any indicator of social/moral cognition through identification and transportation contained zero (95% CI of partially standardized relative indirect effect that came closest to excluding zero: [−0.0328, 0.090]), there was no evidence that the effect of reading stories on social cognition was mediated by identification or transportation.

Exploratory Analyses

Exploratory analyses examined whether the predominant lack of textual effects was the result of confounds with third variables, in particular familiarity with subject matter. According to a one-way ANOVA, participants in the three groups did not report different levels of familiarity with the subject matter of the text, indicated by respondents' answer to "Reflecting on your own life experiences, how familiar were the events in the story to you?", $p = 0.445$. However, familiarity was significantly positively correlated with transportation, $r = 0.30$, $p < 0.0001$ (this is in line with a previous finding by Green, 2004), and identification, $r = 0.22$, $p < 0.0001$ (and not correlated with remaining outcomes, $ps > 0.20$), showing that participants who were more familiar with the subject felt more transported into events in the text and identified to a greater extent with the character.

Discussion

Experiment 1 investigated whether the perspective from which a story is told, both in terms of narrative voice and focalization, affects readers' social and moral cognition immediately after reading. It also tested whether such distal effects are mediated by proximal outcomes of reading a narrative, in particular transportation into the story world and identification with the protagonist.

Hypothesis 1 predicted that performance on social and moral cognition tasks would be better after reading an internally focalized narrative than after reading an externally focalized narrative. This assumption was not supported by the present data, as none of the indicators of social or moral cognition differed between the three reading groups (first-person voice with internal focalization *vs* third-person voice with internal focalization *vs* third-person voice with external focalization).

Hypothesis 2 postulated that the narrative voice in which a story is told would not influence social or moral cognition. Since none of the outcomes differed between the three groups, this prediction was confirmed.

Hypothesis 3 assumed that effects of perspective on social cognition and morality would be mediated by transportation and identification with the story's protagonist. As none of the mediations revealed indirect effects of transportation or identification, this hypothesis was not supported.

Interestingly, the most proximal outcomes, i.e., transportation and identification, were affected by rather stable characteristics, namely trait empathy and familiarity with subject matter, as partly predicted by Consoli (2018), hypothesis 6. Possibly, such trait-level qualities determine more strongly how we are affected by narratives than textual features such as perspective and focalization. In that case it would not be surprising that no effects of the texts on more distal outcomes, such as ToM and emotion recognition, were observed.

However, the predominant lack of group differences could also indicate that the three story versions under investigation were not sufficiently diverse to yield measurable effects, especially regarding the difference between internal and external focalization. In line with this assumption, the participants in the external focalization group had far more difficulty recognizing that their version of the story was told by an anonymous narrator than participants in the first-person group had in recognizing that that their version was told by the mother (18 *vs*. 1 participants failed the comprehension check in each condition, respectively; see also **Figure 2** and Participants section above). Apparently, a substantial number of participants perceived the third-person version with external focalization to be at least partly told by the mother, even though the third-person voice clearly indicated that the protagonist was not the narrator. Taken together, the external focalization version may not have been sufficiently external, which could have driven the lack of group differences in both proximal effects (i.e., transportation, identification) and distal outcomes (i.e., social and moral cognition). To test this possibility, in Experiment 2 we further edited the externalized version of the third-person narrative to reinforce the external focalization. We then compared this new version with the two internal focalization versions used in Experiment 1, testing for the same effects as in that experiment.

EXPERIMENT 2

Materials and Methods

The Research Ethics Committee of the School of Psychology at the University of Kent, United Kingdom, approved the study prior to its start. Unless otherwise mentioned, methodological procedures were the same as in Experiment 1.

Participants

The target sample was reached after 334 volunteers. This time, only nine participants who read the external third-person version of the story indicated that the story was told by the mother (i.e., participant failed check of text comprehension; see **Figure 4**), suggesting that the measures taken did further externalize this story version. When exclusion criteria were applied, 262 participants remained in the final sample, 86 of whom had read the first-person narrative with internal focalization (60.50% female, mean age = 33.63, *SD* of age = 13.58), 86 who had read the third-person narrative with internal focalization (65.1% female, mean age = 31.47, *SD* of age = 12.24), and

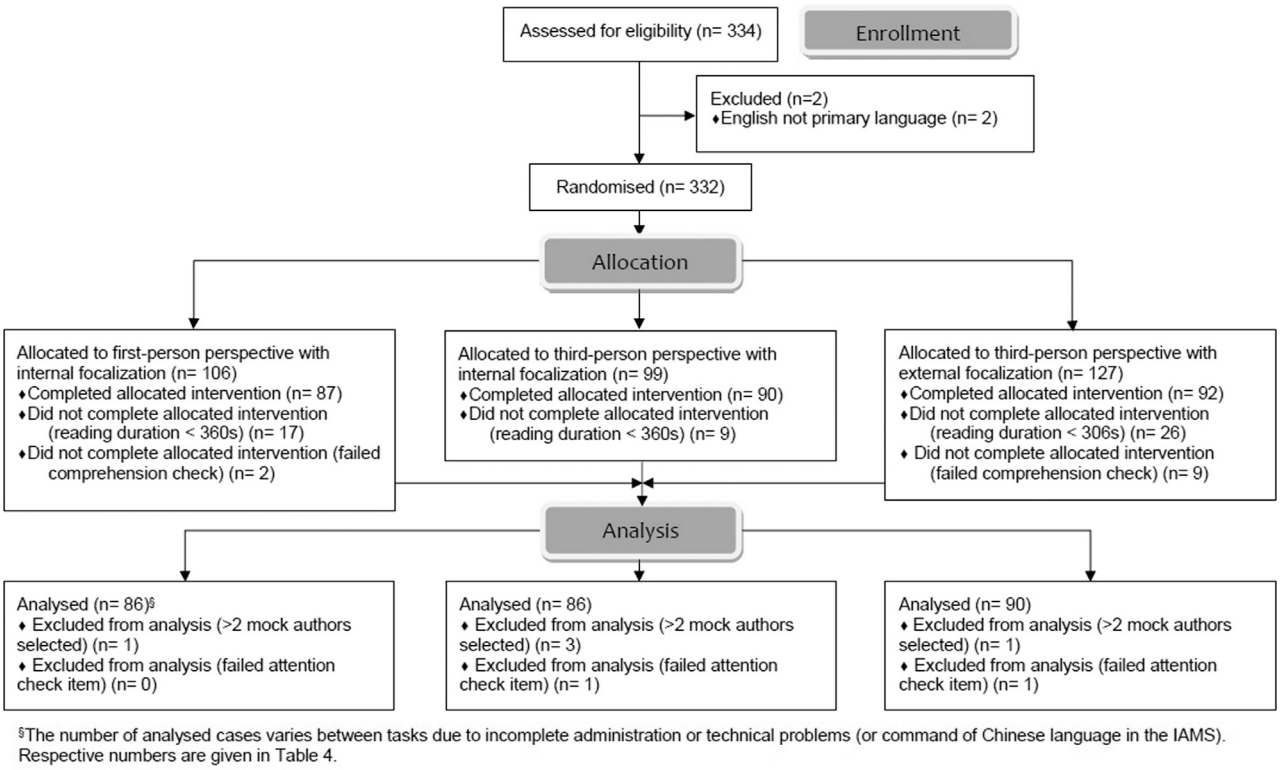


FIGURE 4 | Flow of participants through Experiment 2.

TABLE 4 | Exemplary comparison of the text stimuli used in Experiments 1 and 2.

third-person narrative, external focalization, used in experiment 1	third-person narrative, external focalization, used in experiment 2
As the bus approached Sloane Square, the women stubbed out their cigarettes and got up and came swaying along towards Marjorie, smelling of expensive perfume and of all the smoke they had inhaled. The great blue eyes of Bette stared at Marjorie for a moment. Later that evening she would whisper to Wallis, 'Wasn't that the woman with the caterwauling child at the station? What was she doing on our bus?' They went down the stairs and walked away. They did not look up. Marjorie jumped off the bus just as it was pulling away. She stumbled, but didn't fall.	As the bus approached Sloane Square, the women stubbed out their cigarettes and got up and came swaying along the aisle, smelling of expensive perfume and of all the smoke they had inhaled. The great blue eyes of Bette stared at Patience's mother for a moment. Later that evening she would whisper to Wallis, 'Wasn't that the woman with the caterwauling child at the station? What was she doing on our bus?' They went down the stairs and walked away. They did not look up. Patience's mother jumped off the bus just as it was pulling away. She stumbled, but didn't fall.

90 who had read the third-person narrative with external focalization (60.0% female, mean age = 32.86, *SD* of age = 13.34). See **Figure 4** for a schematic of the flow of participants through the experiment.

Reading Stimuli

Again, three versions of Rose Tremain's short story "The Closing Door" served as reading stimuli. In addition to the original third-person narrative with internal focalization and the first-person version, both of which were used in Experiment 1, the third-person version using external focalization used in Experiment 1 was further externalized by, e.g., referring to the protagonist, not as 'Marjorie', but as 'Patience's mother' or 'the mother', and removing descriptions of other characters that mentioned Marjorie's perspective on them. The new version now consisted of 2979 words. **Table 4** provides an exemplary comparison of the two third-person versions with external focalization used in Experiment 1 and Experiment 2; full stimuli are available on the Open Science Framework, https://osf.io/ef6a5/.

Assessment Tasks

The same measures as in Experiment 1 were applied. The following internal consistencies were observed in the sample of Experiment 2: Empathy Quotient: Cronbach's α = 0.88, Transportation Scale: Cronbach's α = 0.76, Identification Scale: Cronbach's α = 0.87.

Results

Main Analyses

For each dependent measure, descriptive statistics and inferential statistics of the main effect of text are summarized in **Table 5**, and the key effects are plotted in **Figure 5**. Fifty-

TABLE 5 | Experiment 2: Descriptive statistics for each dependent measure in each experimental group, and ANCOVA results for the main effect of text.

Dependent measure		First-person narrative with internal focalization		third-person narrative with internal focalization		third-person narrative with external focalization		ANCOVA: main effect of text			
		n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	df	*F value*	*p value*	η_p^2
Word completion task	Agency score (possible range: 0–7)	84	3.40 (1.37)	84	3.67 (1.32)	89	3.28 (1.41)	2, 251	1.86	0.158	0.015
	Communion score (possible range: 0–7)	84	1.86 (1.31)	84	1.54 (1.08)	84	1.65 (1.21)	2, 251	1.57	0.211	0.012
Frith-Happé animations task	MCQ 1: Accuracy sum score (possible range: 0–4)	84	1.67 (1.06)	84	2.05 (1.02)	89	2.00 (1.13)	2, 251	2.89	0.058	0.022
	MCQ 2: Accuracy sum score (possible range: 0–8)	72	2.49 (1.46)	78	3.01 (1.64)	79	3.09 (1.60)	2, 223	3.31	0.038	0.029
RMET: Relative frequency of correct responses (possible range: 0–1)		84	0.69 (0.17)	84	0.69 (0.20)	89	0.69 (0.18)	2, 251	0.01	0.986	0.000
Implicit affect towards moral stimuli score (possible range: 0–1)		74	0.20 (0.25)	67	0.15 (0.19)	77	0.20 (0.24)	2, 211	2.09	0.126	0.019
Implicit moral identity IAT: D6 (possible range: -2-2)		68	0.59 (0.38)	65	0.60 (0.35)	73	0.68 (0.30)	2, 200	1.41	0.247	0.014
Transportation Scale (possible range: 12–84)		84	56.81 (7.48)	84	58.24 (9.25)	89	56.92 (10.46)	2, 251	0.13	0.881	0.001
Identification Scale (possible range: 8–56)		84	46.56 (6.00)	83	47.27 (6.04)	89	45.61 (7.58)	2, 250	0.57	0.569	0.005
Author Recognition Test – Genres: Fiction sub score (possible range: 0–110)		86	15.45 (15.81)	86	15.10 (16.40)	90	16.94 (16.84)				
Empathy Quotient (possible range: 0–80)		85	43.55 (10.92)	85	46.03 (11.07)	90	43.33 (12.98)				

Note. *MCQ, multiple choice question; RMET, reading the mind in the eyes test–revised; IAT, implicit association test.*

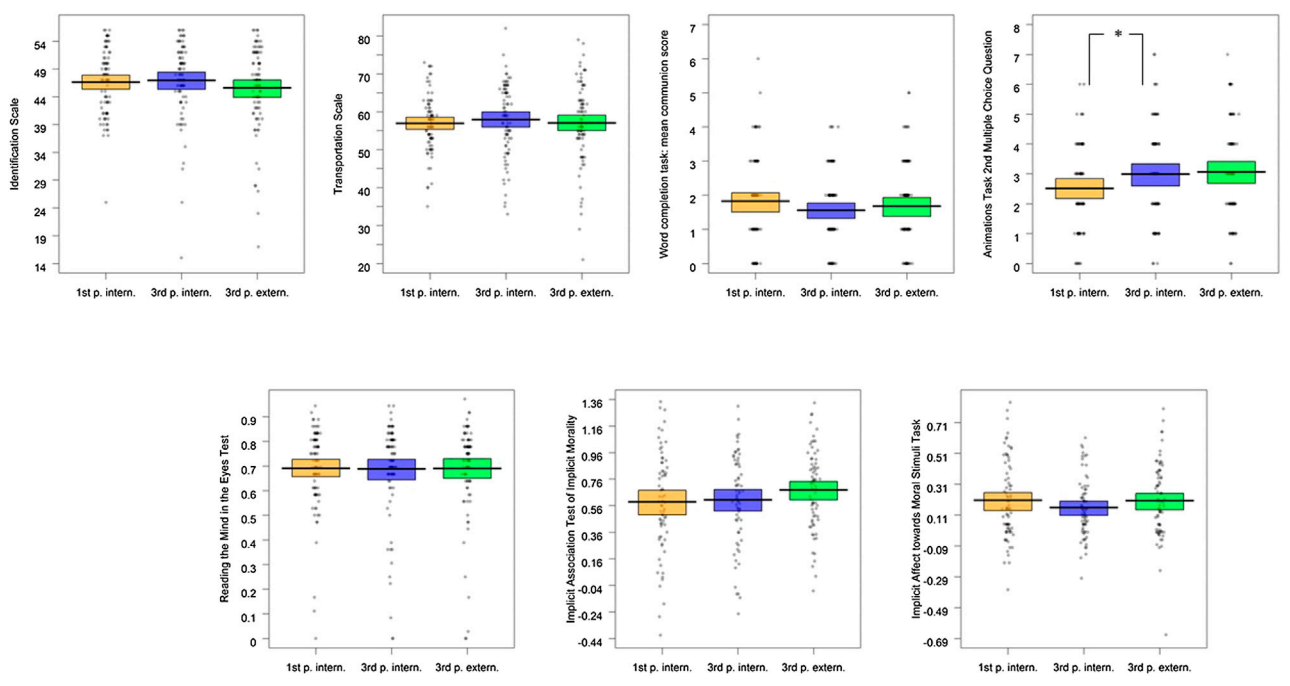


FIGURE 5 | Pirate plots for main outcomes variables of Experiment 2.

seven missing items values - 53 from the Empathy Quotient, 2 from the Identification Scale, 2 from the Transportation Scale - were imputed with the individual scale mean on the remaining items. Missing values for 3 further participants - 2 from the Empathy Quotient, 1 from the Identification Scale - could not be imputed because more than 20% of scale items were missing. The respective scales of these participants were omitted from final analyses.

Replicating previous research with the Implicit Affect towards Moral Stimuli task, participants overall were significantly more likely to judge a pictograph as pleasant after a positive moral prime than after a negative moral prime (M = 0.70 *vs.* 0.44; t (225) = 13.11, $p < 0.001$, d = 0.84), and were significantly more likely to judge a pictograph as pleasant after a positive morality-irrelevant control stimulus than after a negative morality-irrelevant control stimulus (M = 0.63 *vs.* 0.44; t (225) = 12.52, $p < 0.001$, d = 0.83). This supports the underlying assumption of the Implicit Affect towards Moral Stimuli task that ratings are biased towards prime valence. Overall accuracy on the Reading the Mind in the Eyes Test-Revised and animations tasks was good (M = 69% and 1.89, respectively), and the positive mean D6 value (M = 0.62) in the moral IAT is consistent with previous research showing a preference for an implicit moral self-concept.

Only one of the ANCOVAs showed a significant effect of text (all others, ps > 0.05). Specifically, the second set of multiple choice questions in the animations task, F (2, 223) = 3.31, p = 0.038, η_p^2 = 0.029, revealed that the internal third-person version outperformed the first-person version, contrast estimate = −0.544, SE = 0.253, 95%CI [−1.044, −0.045], p = 0.033, but did not differ from the external third-person version, contrast estimate = −0.043, SE = 0.248, 95%CI [−0.532, 0.446], p = 0.863. This pattern suggests that reading a third-person narrative may have had a greater impact on ToM compared to a first-person narrative, regardless of focalization.

As to the impact of covariates, Empathy Quotient sum score had a significant effect on the identification scale, F (1, 250) = 39.89, $p < 0.0001$, η_p^2 = 0.138, on the transportation scale, F (1, 251) = 26.11, $p < 0.0001$, η_p^2 = 0.126, and on the first multiple choice question in the animations task, F (1, 251) = 4.17, p = 0.042, η_p^2 = 0.016. Bivariate correlations demonstrated that Empathy Quotient was positively related to each of these outcomes ($0.150 < r < 0.387$, ps < 0.016). Furthermore, the fiction sub score of the Author Recognition Test–Genres had a significant effect on the percentage of correct responses in the Reading the Mind in the Eyes Test-Revised, F (1, 251) = 4.01, p = 0.046, η_p^2 = 0.016, and on the second set of multiple choice questions in the animations task, F (1, 223) = 8.80, p = 0.003, η_p^2 = 0.038. The Author Recognition Test–Genres fiction sub score was positively correlated with both outcomes, Reading the Mind in the Eyes Test-Revised: r = 0.065, p = 0.065; animations task: r = 0.203, p = 0.002, however, correlations reached significance for the animations task only. The IACS score had a significant effect on the implicit affect towards moral stimuli score, F (1, 211) = 172.76, $p < 0.0001$, η_p^2 = 0.450, and was significantly positively correlated with this outcome, r = 0.668, $p < 0.0001$. There were no further significant main effects (ps > 0.05).

Mediations were calculated to test whether a potential relationship between reading stories and improved social cognition was mediated by identification with characters and transportation into the story. All 95% confidence intervals of the indirect effect of story reading on any indicator of social cognition through identification and transportation contained zero (95% CI of partially standardized relative indirect effect that came closest to excluding zero: [−0.0145, 0.0093]), thus the result supported Experiment 1 in showing no evidence that the effect of reading stories on social cognition is mediated by identification and transportation.

Exploratory Analyses

Exploratory analyses examined whether the predominant lack of textual effects was the result of confounds with confound variables, especially familiarity with the subject matter (as in Experiment 1). According to a one-way ANOVA, participants in the three groups did not differ in this respect, p = 0.444. However, familiarity was significantly positively correlated with transportation, r = 0.30, $p < 0.0001$, and the Author Recognition Test–Genres fiction sub score, r = 0.14, p = 0.023 (and not correlated with remaining outcomes, ps > 0.15).

Discussion

Experiment 2 tested the same research questions as Experiment 1, namely whether the perspective from which a written narrative is told, both in terms of narrative voice and focalization, impacts on recipients' levels of general ToM, emotion recognition, and morality immediately after reading; and whether such distal effects are mediated by proximal outcomes of reading a narrative, particularly transportation and identification. Compared with Experiment 1, Experiment 2 used a stronger manipulation of focalization for the version of the story with third-person narration and external focalization. The remaining methods replicated those used in Experiment 1.

Hypothesis 1 (social and moral cognition would be higher after reading a narrative told through internal focalization than external focalization) was again not confirmed. Similar to Experiment 1, most indicators of social and moral cognition did not differ between the three reading groups (first-person voice with internal focalization *vs* third-person voice with internal focalization *vs* third-person voice with external focalization), and the only significant group difference (i.e., in the animations task) was that the first-person narrative with internal focalization scored lower than both third-person versions.

As in Experiment 1, hypothesis 2 (the narrative voice in which a story is told would not affect social or moral cognition) was mainly corroborated, since most dependent measures did not differ between the three groups. However, the only significant group difference was in contrast to hypothesis 2: the first-person with internal focalization group scored lower on the animations task measuring ToM than the third-person with internal focalization group. Thus, although this effect of voice did not emerge in Experiment 1, raising the possibility that it

reflects a false positive, it could suggest that reading a third-person narrative may have had a greater impact on ToM compared to a first-person narrative, regardless of focalization.

Paralleling the implication of Experiment 1, hypothesis 3 (effects of perspective on social and moral cognition would be mediated by transportation and identification with the story's protagonist) was rejected because none of the mediations detected indirect effects of transportation or identification. In addition, transportation was positively linked with trait empathy and familiarity with subject matter, and identification (but not familiarity with subject matter as found in Experiment 1) was associated with trait empathy.

In summary, despite implementing an even more externalized version of the external focalization version, which should have facilitated the detection of focalization-based effects (if there were any), Experiment 2 failed to observe any benefits of internal over external focalization. The finding that participants in the third-person with external focalization group in Experiment 2 were more likely to recognize that the story was told by an anonymous narrator than in Experiment 1 (9/101 = 8.91% of participants in Experiment 2 *vs.* 18/104 = 17.31% of participants in Experiment 1 failed the comprehension check) provides a manipulation check. It demonstrates that the measures taken to further externalize the external focalization version served their purpose. Therefore, Experiment 2 renders it unlikely that the lack of focalization-based effects in Experiment 1 was due to a lack of contrast between the three story versions.

GENERAL DISCUSSION

A recent wave of research has suggested that reading narratives offers a means to improve our social cognition, a set of skills crucial for everyday interpersonal functioning. These effects are often associated by theorists in other disciplines with moral improvement. The present research investigated whether potential narrative-related benefits for both social and moral cognition depend on the perspective/voice from which narratives are told, and the degree to which readers are transported into the story world and identify themselves with story characters. The two experiments reported here do not suggest that perspective has a significant impact on the extent to which narratives modulate social and moral cognition, either directly or as an indirect effect of transportation or identification. Indeed, exploratory analyses using the combined sample from Experiments 1 and 2 (N = 520) also failed to reveal significant differences between the three texts on our measures of social or moral cognition, showing that the null effects cannot be explained by insufficient power to detect a small effect size.

Whilst previous empirical work has investigated narrative perspective almost exclusively in terms of narrative voice, despite Genette's (1980) seminal distinction between narrative voice and focalization, the present research manipulated voice and focalization separately to disentangle the contribution of both aspects. We found no effects of focalization on either proximal effects (transportation/identification) or distal outcomes (social/moral cognition). The null result for identification is inconsistent with the only pre-existing experiment on focalization by Salem et al. (2017), where story versions using internal focalization, i.e., first-person and psycho-narration, were associated with an enhanced tendency to take the protagonists' perspective (they did not measure transportation/social cognition/moral cognition). The divergence may be explained by several differences between the study by Salem et al. (2017) and the present experiments. For instance, the text stimuli used by Salem et al. (2017) were short excerpts whose length was approximately 15% of the word count of the complete short stories used in the present experiments. Secondly, Salem et al. (2017) deployed three measures of perspective-taking: relatedness (extent to which readers can relate to the protagonist), spatial perspective-taking, and identification (psychological perspective-taking). Only the third, identification, overlapped with our study, where it was measured in a similar though not identical way. But the effects of internal focalization were almost entirely on the other two measures, relatedness and spatial point of view, rather than on identification. Future investigations using a variety of text stimuli and broad assessments of identification are needed to identify the conditions under which narrative perspective influences transportation into the story world and identification with characters.

In line with our prediction, perspective in terms of narrative voice did not consistently influence ToM and morality. Only one dependent measure was affected by narrative voice in one experiment (and did not replicate across the two experiments), though the one finding suggested that the first-person narrative using internal focalization was *less effective* in enhancing ToM to the third-person version using internal focalization. Given the inconsistency on affected outcomes, our results do not support a robust effect of narrative voice on social or moral cognition.

Finally, the lack of mediation by transportation or identification could be explained by the non-existent effect of focalization on more distal outcomes referring to social and moral cognition. The consequent lack of variation in these outcomes may have made the detection of mediating effects more difficult. Nevertheless, the present lack of perspective-based effects on identification contradicts the notion that linguistic cues about a character's point of view facilitate identification with this character, as predicted by the Linguistic Cues Framework (van Krieken et al., 2017). According to this model, stories using internal focalization should lead to greater identification than external focalization stories, since the former provide more information about the character's mental life than the latter. Hence, the current pattern of findings encourages a revision of the Linguistic Cues Framework to achieve a more exhaustive account of the factors eliciting identification with narrative characters.

The results of our exploratory analyses seem to suggest that trait-level variables, in particular dispositional empathy, have a greater impact on identification than textual characteristics, in particular narrative perspective. In other words, how much readers identify with a particular story character seems to depend mainly on readers' pre-existing level of trait empathy;

the current results do not show that variations of narrative perspective have the power to override the influence of trait empathy, at least if these variations concern a short reading assignment. Individuals with a strong dispositional tendency to empathize with others seems to apply this tendency also in the case of reading a short story, whereas individuals with weak dispositional empathy apparently also do not tend to empathize with characters of a short story; the perspective from which a given story is told does not seem to modulate such a disposition. This implies that theories of character identification should include relevant reader traits. It should however be borne in mind that the effect of trait empathy observed in the present research was an incidental finding not related to hypothesis testing. Clearly, targeted investigations are needed to clarify the role of reader traits *vs* text properties within processes of character identification.

This seems particularly desirable in view of similar recent experiments which observed a stronger impact of rather stable reader traits on experience and processing of narratives than state-level variables, i.e., narrative voice of reading stimuli and mental imagery instructions during reading, respectively (Hartung et al., 2017a; Mak et al., 2020). The available evidence suggests that the effects of reading short narratives is, to a relatively large degree, determined by pre-existing reader traits, and less so, if at all, by state variables including textual variations. Mischel's (1977) concept of situation strength may provide an explanation for this pattern, if we consider reading short stories as "weak situations". Mischel's (1977) introduced the concept of situation strength to explain whether behavior is shaped by stable personality traits (applied to the current experiments: trait empathy) or temporary situational influences (applied to the current experiments: reading a short story told from a certain narrative perspective). According to Mischel, situational impact is crucial in strong situations, which "lead everyone to construe the particular events the same way, induce uniform expectancies regarding the most appropriate response pattern, provide adequate incentives for the performance of that response pattern and require skills that everyone has to the same extent" (p. 347). In contrast, behavior is assumed to be determined by personality traits in weak situations, which "are not uniformly encoded, do not generate uniform expectancies concerning desired behavior, do not offer sufficient incentives for its performance, or fail to provide the learning conditions required for successful genesis of behavior" (p. 347). Future work could explore this explanatory approach, for instance by investigating whether the impact of state variables, e.g., textual properties, increases with increasing length of reading assignments (i.e., when situational impact is strengthened).

Taken together, the present results do not support the proposal that either narrative voice or perspective/focalization underlies the benefits, if any, of reading narratives for social and moral cognition. Since we applied the Reading the Mind in the Eyes test, the indicator of social cognition most frequently used in previous experiments observing narrative-based benefits for social cognition (cf. **Table 1**), we can rule out that the null effects trace back to the use of incomparable measures. There is also no evidence in our study to suggest that transportation or identification mediate the relationship between reading narratives and ToM/moral cognition. However, these findings may be limited by the stimulus texts and measurement instruments applied.

The pre-existing evidence for immediate effects of reading a story on social cognition has been relatively weak (Dodell-Feder and Tamir, 2018). However, even if such effects exist, they are unlikely to be produced by all stories (e.g. Mar, 2018). So it is possible that the story used in the current experiments was not suitable to evoke the effects under investigation. Future research should consider a broader spectrum of textual stimuli and dimensions of identification, which may provide a more complete picture of the conditions under which narratives foster our social cognition. Still, one would assume that if text features such as voice and focalization have an effect, they should do so in more than one type of story.

CONCLUSION

Two experiments consistently showed that narrative perspective, either in terms of narrative voice or focalization, did not influences readers' social or moral cognition, either directly, or indirectly via transportation into the story world or identification with the protagonist. Narrative perspective also did not affect transportation or identification. This suggests that the short-term benefits of reading fictional narratives found by other studies may not have been reliant on perspective. Furthermore, the present findings cast doubt on models that propose a central role of narrative perspective for identification with a story character; instead, the current pattern points to the possibility that reader traits, especially empathy, could determine identification to a greater extent than narrative perspective. These possibilities should be tested in future research. In addition, we encourage studies that assess multiple dimensions of identification and employ a broad range of stimulus texts to facilitate generalization of claims.

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: Open Science Framework, https://osf.io/6gsqn/ (Experiment 1), https://osf.io/vh6re/ (Experiment 2).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee of the School of Psychology at the University of Kent, United Kingdom. The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors designed the study, read and approved the submitted version. LW contributed to data collection, performed the statistical analysis, and wrote the first draft of the manuscript. SF, CG, and HF commented and made edits on the draft manuscript.

FUNDING

This work was supported by the Leverhulme Trust under Grant RPG- 2017-365.

REFERENCES

Abell, F., Happé, F., and Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognit. Dev.* 15 (1), 1–16. doi:10.1016/s0885-2014(00)00014-9

Bakan, D. (1966). *The duality of human existence: an essay on psychology and religion*. Chicago, IL: Rand McNally.

Bal, P. M., and Veltkamp, M. (2013). How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation. *PLoS One* 8 (1), e55341. doi:10.1371/journal.pone.0055341

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry* 42 (2), 241–251. doi:10.1111/1469-7610.00715

Baron-Cohen, S., and Wheelwright, S. (2004). The Empathy Quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* 34 (2), 163–175. doi:10.1023/b:jadd.0000022607.19833.00

Baron-Cohen, S. (2000). "Theory of mind and autism: a review," in *International review of research in mental retardation* (Cambridge, MA: Academic Press), Vol. 23, 169–184.

Bartz, J. A., and Lydon, J. E. (2004). Close relationships and the working self-concept: implicit and explicit effects of priming attachment on agency and communion. *Pers. Soc. Psychol. Bull.* 30 (11), 1389–1401. doi:10.1177/0146167204264245

Benjamin, C. F., and Gaab, N. (2012). What's the story? The tale of reading fluency told at speed. *Hum. Brain Mapp.* 33 (11), 2572–2585. doi:10.1002/hbm.21384

Bialystok, L., and Kukar, P. (2018). Authenticity and empathy in education. *Theor. Res. Educ.* 16 (1), 23–39. doi:10.1177/1477878517746647

Black, J. E., and Barnes, J. L. (2020a). Fiction and morality: investigating the associations between reading exposure, empathy, morality, and moral judgment. *Psychol. Pop. Med.* [Epub ahead of print]. doi:10.1037/ppm0000281

Black, J. E., and Barnes, J. L. (2020b). Morality and the imagination: real-world moral beliefs interfere with imagining fictional content. *Phil. Psychol.* 33 (7), 1018–1044. doi:10.1080/09515089.2020.1775799

Black, J. E., and Barnes, J. L. (2015). The effects of reading material on social and non-social cognition. *Poetics* 52, 32–43. doi:10.1016/j.poetic.2015.07.001

Bloom, P. (2017). Against empathy: the case for rational compassion. Random House.

Bora, E., Yucel, M., and Pantelis, C. (2009). Theory of mind impairment in schizophrenia: meta-analysis. *Schizophr. Res.* 109 (1–3), 1–9. doi:10.1016/j.schres.2008.12.020

Bubandt, N., and Willerslev, R. (2015). The dark side of empathy: mimesis, deception, and the magic of alterity. Comparative Studies in Society and History. 57 (1), 5–34.

Breithaupt, F. (2018). The bad things we do because of empathy. *Interdiscipl. Sci. Rev.* 43 (2), 166–174.

Brüne, M. (2005). Emotion recognition, "theory of mind," and social behavior in schizophrenia. *Psychiatr. Res.* 133 (2-3), 135–147. doi:10.1016/j.psychres.2004.10.007

Calarco, N., Fong, K., Rain, M., and Mar, R. A. (2017). "Absorption in narrative fiction and its possible impact on social abilities," in *The handbook of narrative absorption*. Editors F. Hakemulder, M. M. Kuijpers, and K. ESH Tan (Amsterdam: Balint and MM Doicaru), 293–313.

Cohen, J. (2001). Defining identification: a theoretical look at the identification of audiences with media characters. *Mass Commun. Soc.* 4 (3), 245–264. doi:10.1207/s15327825mcs0403_01

Consoli, G. (2018). Preliminary steps towards a cognitive theory of fiction and its effects. *J. Cultural Cognit. Sci.* 2 (1-2), 85–100. doi:10.1007/s41809-018-0019-5

Creer, S. D., Cook, A. E., and O'Brien, E. J. (2019). Can readers fully adopt the perspective of the protagonist?. *Q. J. Exp. Psychol.* 73 (5), 664–675. doi:10.1177/1747021819891407

Currie, G. (2010). *Narratives and narrators: a philosophy of stories*. Oxford, UK: Oxford University Press.

De Clercq, A., Crombez, G., Buysse, A., and Roeyers, H. (2003). A simple and sensitive method to measure timing accuracy. *Behav. Res. Methods Instrum. Comput.* 35 (1), 109–115. doi:10.3758/bf03195502

Deane, P., Somasundaran, S., Lawless, R. R., Persky, H., and Appel, C. (2019). The key practice, building and sharing stories and social understandings: the intrinsic value of narrative. *ETS Res. Rep. Ser.* 2019 (1), 1–78. doi:10.1002/ets2.12266

Djikic, M., Oatley, K., and Moldoveanu, M. C. (2013). Reading other minds. *Ssol.* 3 (1), 28–47. doi:10.1075/ssol.3.1.06dji

Dodell-Feder, D., and Tamir, D. I. (2018). Fiction reading has a small positive impact on social cognition: a meta-analysis. *J. Exp. Psychol. Gen.* 147 (11), 1713–1727. doi:10.1037/xge0000395

Dolby, N. (2012). *Rethinking multicultural education for the next generation: the new empathy and social justice*. New York, NY: Routledge.

Edmiston, W. F. (1989). Focalization and the first-person narrator: a revision of the theory. *Poetics Today* 10 (4), 729–744. doi:10.2307/1772808

Findlay, L. C., Girardi, A., and Coplan, R. J. (2006). Links between empathy, social behavior, and social understanding in early childhood. *Early Child. Res. Q.* 21 (3), 347–359. doi:10.1016/j.ecresq.2006.07.009

Fiske, S. T., and Taylor, S. E. (2013). *Social cognition: from brains to culture*. London, UK: SAGE.

Fludernik, M. (2001). New wine in old bottles? Voice, focalization and new writing. *New Lit. Hist.* 32 (3), 619–638. doi:10.1353/nlh.2001.0033

Frith, C. D., and Frith, U. (1999). Interacting minds--a biological basis. *Science* 286 (5445), 1692–1695. doi:10.1126/science.286.5445.1692

Genette, G. (1980). *Narrative discourse*. Ithaca, NY: Cornell University Press.

Genette, G. (1983). *Nouveau discours du recit*. Paris, France: Seuil.

Gerrig, R. J. (1993). Experiencing narrative worlds: on the psychological activities of reading. 1st Edn. Routledge. doi:10.4324/9780429500633

Green, M. C., and Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *J. Pers. Soc. Psychol.* 79 (5), 701–721. doi:10.1037//0022-3514.79.5.701

Green, M. C. (2004). Transportation into narrative worlds: the role of prior knowledge and perceived realism. *Discourse Process* 38 (2), 247–266. doi:10.1207/s15326950dp3802_5

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74 (6), 1464. doi:10.1037//0022-3514.74.6.1464

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 (2), 197–216. doi:10.1037/0022-3514.85.2.197

Hakemulder, J. (2000). *The moral laboratory: experiments examining the effects of reading literature on social perception and moral self-concept*. Amsterdam, Netherlands: John Benjamins.

Hartung, F., Hagoort, P., and Willems, R. M. (2017a). Readers select a comprehension mode independent of pronoun: evidence from fMRI during narrative comprehension. *Brain Lang.* 170, 29–38. doi:10.1016/j.bandl.2017.03.007

Hartung, F., Withers, P., Hagoort, P., and Willems, R. M. (2017b). When fiction is just as real as fact: No differences in reading behavior between stories believed to be based on true or fictional events. *Front. Psychol.* 8, 1618. doi:10.3389/fpsyg.2017.01618

Hobson, P. (2007). "Empathy and autism," in *Empathy in mental illness*. Editors T. F. D. Farrow and P. W. R. Woodruff (Cambridge, UK: Cambridge University Press), 126–142.

Hofmann, W., and Baumert, A. (2010). Immediate affect as a basis for intuitive moral judgement: an adaptation of the affect misattribution procedure. *Cognit. Emot.* 24 (3), 522–535. doi:10.1080/02699930902847193

P. Hühn, W. Schmid, and J. Schönert (Editors) (2009). *Point of view, perspective, and focalization: modeling mediation in narrative.* Berlin: De Gruyter.

Johnson, D. R., Cushman, G. K., Borden, L. A., and McCune, M. S. (2013). Potentiating empathic growth: generating imagery while reading fiction increases empathy and prosocial behavior. *Psychol. Aesth. Creativity Arts* 7 (3), 306–312. doi:10.1037/a0033261

Johnson, D. R. (2012). Transportation into a story increases empathy, prosocial behavior, and perceptual bias toward fearful expressions. *Pers. Indiv. Differ.* 52 (2), 150–155. doi:10.1016/j.paid.2011.10.005

Johnson, D. R. (2013). Transportation into literary fiction reduces prejudice against and increases empathy for Arab-Muslims. *Ssol.* 3 (1), 77–92. doi:10.1075/ssol.3.1.08joh

Johnson, S. (1750). *The Rambler*. London: Harrison and Co. 31.

Kidd, D. C., and Castano, E. (2013). Reading literary fiction improves theory of mind. *Science* 342 (6156), 377–380. doi:10.1126/science.1239918

Kidd, D., and Castano, E. (2019). Reading literary fiction and theory of mind: three preregistered replications and extensions of Kidd and Castano (2013). *Soc. Psychol. Pers. Sci.* 10 (4), 522–531. doi:10.1177/1948550618775410

Kidd, D., Ongis, M., and Castano, E. (2016). On literary fiction and its effects on theory of mind. *Ssol*. 6 (1), 42–58. doi:10.1075/ssol.6.1.04kid

Killen, M., Lynn Mulvey, K., Richardson, C., Jampol, N., and Woodward, A. (2011). The accidental transgressor: morally-relevant theory of mind. *Cognition* 119 (2), 197–215. doi:10.1016/j.cognition.2011.01.006

Kim, N., Kim, H. K., Wojcieszak, M., Igartua, J. J., and Lim, C. M. (2019). The presence of the protagonist: explaining narrative perspective effects through social presence. *Med. Psychol.* 23 (6), 891–914. doi:10.1080/15213269.2019.1665548

Koopman, E. M. (2015). Empathic reactions after reading: the role of genre, personal factors and affective responses. *Poetics* 50, 62–79. doi:10.1016/j.poetic.2015.02.008

Kotovych, M., Dixon, P., Bortolussi, M., and Holden, M. (2011). Textual determinants of a component of literary identification. *Ssol.* 1 (2), 260–291. doi:10.1075/ssol.1.2.05kot

Lee, K. (2007). "Empathy deficits in schizophrenia," in *Empathy in mental illness*. Editors T. F. D. Farrow and P. W. R. Woodruff (Cambridge, UK: Cambridge University Press), 17–32.

Leppänen, J. M., and Hietanen, J. K. (2001). Emotion recognition and social adjustment in school-aged girls and boys. *Scand. J. Psychol.* 42 (5), 429–435. doi:10.1111/1467-9450.00255

Liu, A., and Want, S. (2015). Literary fiction did not improve affective ToM. Available at: www.psychfiledrawer.org/replication.php?attempt=MjI1 (Accessed October 1, 2018).

Mak, M., Vries, C. D., and Willems, R. M. (2020). The influence of mental imagery instructions and personality characteristics on reading experiences. *Collabra Psychol.* 6 (1), 43. doi:10.1525/collabra.281

Mar, R. A., and Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspect. Psychol. Sci.* 3 (3), 173–192. doi:10.1111/j.1745-6924.2008.00073.x

Mar, R. A., and Rain, M. (2015). Narrative fiction and expository nonfiction differentially predict verbal ability. *Sci. Stud. Read.* 19, 419–433. doi:10.1080/10888438.2015.1069296

Mar, R. A. (2018a). Evaluating whether stories can promote social cognition: introducing the social processes and content entrained by narrative (SPaCEN) framework. *Discourse Process* 55 (5-6), 454–479. doi:10.1080/0163853x.2018.1448209

Mar, R. A. (2018b). Stories and the promotion of social cognition. *Curr. Dir. Psychol. Sci.* 27 (4), 257–262. doi:10.1177/0963721417749654

Margolin, U. (2009). "Focalization: where do we go from here?" in Point of view, perspective, and focalization. Modeling mediation in narrative. Editors P. Hühn et al. Berlin: de Gruyter, 48–58.

Mischel, W. (1977). "The interaction of person and situation," in *Personality at the cross-roads: current issues in interactional psychology*. Editors D. Magnusson and N. S. Endler (Hillsdale, NJ: Lawrence Erlbaum), 333–352.

Moore, M., and Gordon, P. C. (2015). Reading ability and print exposure: item response theory analysis of the author recognition test. *Behav. Res. Methods* 47 (4), 1095–1109. doi:10.3758/s13428-014-0534-3

Mumper, M. L., and Gerrig, R. J. (2017). Leisure reading and social cognition: a meta-analysis. *Psychol. Aesth. Creativity Arts.* 11 (1), 109–120. doi:10.1037/aca0000089

Mumper, M. L., and Gerrig, R. J. (2019). How does leisure reading affect social cognitive abilities?. *Poetics Today* 40 (3), 453–473. doi:10.1215/03335372-7558080

Nussbaum, M. C. (1990). *Love's knowledge*. Oxford, UK: Oxford University Press.

Nussbaum, M. C. (1995). *Poetic justice: the literary imagination and public life*. Boston, MA: Beacon Press.

Nussbaum, M. C. (2001). *Upheavals of thought. The intelligence of emotions*. Cambridge, UK: Cambridge University Press.

Oakley, B. F., Brewer, R., Bird, G., and Catmur, C. (2016). Theory of mind is not theory of emotion: a cautionary note on the Reading the Mind in the Eyes Test. *J. Abnorm. Psychol.* 125 (6), 818–823. doi:10.1037/abn0000182

Panero, M. E., Weisberg, D. S., Black, J., Goldstein, T. R., Barnes, J. L., Brownell, H., et al. (2016). Does reading a single passage of literary fiction really improve theory of mind? An attempt at replication. *J. Pers. Soc. Psychol.* 111 (5), e46. doi:10.1037/pspa0000064

Perugini, M., and Leone, L. (2009). Implicit self-concept and moral action. *J. Res. Pers.* 43 (5), 747–754. doi:10.1016/j.jrp.2009.03.015

Peskin, J., and Astington, J. W. (2004). The effects of adding metacognitive language to story texts. *Cognit. Dev.* 19 (2), 253–273. doi:10.1016/j.cogdev.2004.01.003

Raney, A. A. (2004). Expanding disposition theory: reconsidering character liking, moral evaluations, and enjoyment. *Commun. Theor.* 14 (4), 348–369. doi:10.1111/j.1468-2885.2004.tb00319.x

Salem, S., Weskott, T., and Holler, A. (2017). Does narrative perspective influence readers' perspective-taking? An empirical study on free indirect discourse, psycho-narration and first-person narration. *Glossa.* 2 (1), 61. doi:10.5334/gjgl.225

Samur, D., Tops, M., and Koole, S. L. (2018). Does a single session of reading literary fiction prime enhanced mentalising performance? Four replication experiments of Kidd and Castano (2013). *Cognit. Emot.* 32 (1), 130–144. doi:10.1080/02699931.2017.1279591

Samur, D., Tops, M., Slapšinskaitė, R., and Koole, S. L. (2020). Getting lost in a story: how narrative engagement emerges from narrative perspective and individual differences in alexithymia. *Cognit. Emot.* 1–13. doi:10.1080/02699931.2020.1732876

Singer, T., and Klimecki, O. M. (2014). Empathy and compassion. *Curr. Biol.* 24 (18), R875–R878. doi:10.1016/j.cub.2014.06.054

Sutton, J., Smith, P. K., and Swettenham, J. (1999). Social cognition and bullying: social inadequacy or skilled manipulation?. *Br. J. Dev. Psychol.* 17 (3), 435–450. doi:10.1348/026151099165384

Tal-Or, N., and Cohen, J. (2010). Understanding audience involvement: conceptualizing and manipulating identification and transportation. *Poetics* 38 (4), 402–418. doi:10.1016/j.poetic.2010.05.004

Tal-Or, N., and Cohen, J. (2016). Unpacking engagement: convergence and divergence in transportation and identification. *Ann. Int. Commun. Assoc.* 40 (1), 33–66. doi:10.1080/23808985.2015.11735255

Ugazio, G., Majdandžić, J., and Lamm, C. (2014). "Are empathy and morality linked? Insights from moral psychology, social and decision neuroscience, and philosophy," in *Empathy and morality*. Editor H. L. Maibom (Oxford, UK: Oxford University Press), 155–171.

Van Krieken, K., Hoeken, H., and Sanders, J. (2017). Evoking and measuring identification with narrative characters - a linguistic cues framework. *Front. Psychol.* 8, 1190. doi:10.3389/fpsyg.2017.01190

Van Krieken, K. v. (2018). How reading narratives can improve our fitness to survive. Narrative Inquiry 28 (1), 139–160. doi:10.1075/ni.17049.kri

Van Lissa, C., Caracciolo, M., van Duuren, T., and van Leuveren, B. (2016). Difficult empathy: the effect of narrative perspective on readers' engagement with a first-person narrator. *Diegesis.* 5 (1), 43–63.

Vermeule, B. (2010). Why do we care about literary characters?. Johns Hopkins University Press.

Vezzali, L., Stathi, S., Giovannini, D., Capozza, D., and Trifiletti, E. (2015). The greatest magic of Harry Potter: reducing prejudice. *J. Appl. Soc. Psychol.* 45 (2), 105–121.

Walkington, Z., Wigman, S. A., and Bowles, D. (2019). *The impact of narratives and transportation on empathic responding*: Poetics. doi:10.1016/j.poetic.2019.101425

Watson, A. C., Nixon, C. L., Wilson, A., and Capage, L. (1999). Social interaction skills and theory of mind in young children. *Dev. Psychol.* 35 (2), 386. doi:10.1037/0012-1649.35.2.386

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* 72 (3), 655–684. doi:10.1111/1467-8624.00304

White, S. J., Coniston, D., Rogers, R., and Frith, U. (2011). Developing the Frith-Happé animations: a quick and objective test of Theory of Mind for adults with autism. *Autism Res.* 4 (2), 149–154. doi:10.1002/aur.174

Wimmer, L., Currie, G., Friend, S., and Ferguson, H. J. *The effects of reading narrative fiction on social cognition: a multi-method approach.*

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

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 04 March 2021
doi: 10.3389/fpsyg.2021.634930

Storytelling on Oral Grounds: Viewpoint Alignment and Perspective Taking in Narrative Discourse

Kobie van Krieken and José Sanders*

Centre for Language Studies, Radboud University Nijmegen, Nijmegen, Netherlands

In this paper, we seek to explain the power of perspective taking in narrative discourse by turning to research on the oral foundations of storytelling in human communication and language. We argue that narratives function through a central process of alignment between the viewpoints of narrator, hearer/reader, and character and develop an analytical framework that is capable of generating general claims about the processes and outcomes of narrative discourse while flexibly accounting for the great linguistic variability both across and within stories. The central propositions of this viewpoint alignment framework are that the distance between the viewpoints of participants in the narrative construal – narrator, character, reader – is dynamic and regulated by linguistic choices as well as contextual factors. Fundamentally, viewpoint alignment is grounded in oral narrative interaction and, from this conversation, transferred to the written narrative situation, varying between demonstration and invasion of the narrative subjects and guiding readers' route of processing the narrative (*experiential* versus *reflective*). Our claim is that variations in viewpoint alignment are functional to the communicative context and intended outcomes of narratives. This is illustrated with the analysis of a corporate journalistic narrative that comprises both interactional and non-interactional aspects of storytelling. The concept of viewpoint alignment further explains the oral fundaments of narrative discourse in conversational storytelling and poses new questions on the relation between the dynamic processing of stories on the one hand and their static outcomes on the other.

Keywords: identification, narrative, perspective taking, storytelling, viewpoint alignment, interaction, conversation

OPEN ACCESS

Edited by:
Jorrig Vogels,
University of Groningen, Netherlands

Reviewed by:
Arie Verhagen,
Leiden University Centre for Linguistics (LUCL), Netherlands
William S. Horton,
Northwestern University, United States

****Correspondence:***
Kobie van Krieken
k.vankrieken@let.ru.nl

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

Received: *29 November 2020*
Accepted: *25 January 2021*
Published: *04 March 2021*

Citation:
van Krieken K and Sanders J (2021) Storytelling on Oral Grounds: Viewpoint Alignment and Perspective Taking in Narrative Discourse. Front. Psychol. 12:634930. doi: 10.3389/fpsyg.2021.634930

INTRODUCTION

A large part of human communication is inherently narrative, that is, representing specific events of specific persons in a specific spatiotemporal setting. Narrative styles and structures are ubiquitous in political communication (Shenhav, 2005; Polletta, 2009), journalism (Van Krieken, 2019), social media (Papacharissi and de Fatima Oliveira, 2012; Georgakopoulou, 2014), education (Devine et al., 2014), informal and professional interpersonal conversations (Avdi, 2008; Rühlemann, 2013), health communication (Wilkin and Ball-Rokeach, 2006), organizational discourse (Barker and Gower, 2010), and so on. In each of these communication subfields, research is carried out on both the processing and effects of narrative communication. A central finding is that narratives influence beliefs, attitudes, intentions, and behaviors more strongly than non-narratives (Braddock and Dillard, 2016). This impact can be explained by processes of "transportation" into the narrative

world – the phenomenological experience of temporarily leaving the actual word behind and "traveling" to the story world (Gerrig, 1993) – and identification with narrative characters (see Van Laer et al., 2014, for a meta-analysis).

However, fundamental questions about what determines transportation, identification, and effective narrative communication have yet to be answered. Review studies on narratives in health communication and educational texts found numerous differences in the experimental narratives used as stimuli in effect studies, making it difficult to pinpoint the exact determinants of narrative persuasion (De Graaf et al., 2016; Sangers et al., 2020). Apart from great variation in length, modality, content, and context, which all inevitably influence the audience's representation of a narrative, there are clues hinting at the importance of linguistic viewpoint phenomena. De Graaf et al.'s (2016) review study indicated that health narratives written from a first-person viewpoint tend to result in stronger identification and to be more effective than narratives written from a third-person viewpoint. However, a second-person viewpoint arouses more emotions and makes readers adopt the narrative character's perspective more strongly than a first-person viewpoint does (Brunyé et al., 2009, 2011).

A different study found that variation in referential viewpoint affects identification: narrators who use pronominal references for narrative characters signal a proximate viewpoint, which can result in stronger identification than nominal references, by which narrators signal a more distant viewpoint (Van Krieken and Sanders, 2017). In addition, yet another study showed that the viewpoint marking by verbs of perception (e.g., "to see," "to observe") guides narrative readers into interpreting the story more strongly from the viewpoint of the character (rather than the narrator) than verbs without viewpoint marking qualities (Van Krieken, 2018). Finally, previous research found that readers are more strongly inclined to adopt a character's perceptual perspective in present tense stories compared to past tense stories (Macrae, 2016): this indicates that the use of verb tense can guide readers into sharing viewpoints with the narrative character. Taken together, these studies signal that various linguistic manifestations of viewpoint – at least grammatical person, referential expressions, verbs of perception, and verb tense – appear to play a role in the cognitive representation and effect of narratives in their audiences.

The present paper builds on these signals and aims to transcend disciplinary boundaries by arguing that generally, narratives function through a central process of viewpoint alignment that can be traced back to oral storytelling. In conversations, narratives alternate between communicative participants who tell stories that function as entertaining, illustrative, and/or educational messages for the other(s) (Labov and Waletzky, 1967). The rhetorical point of stories is met by combining the representation of experience with reflection on experience; both involve viewpoint sharing, but in different ways. Essentially, we argue that the degree to which the viewpoints of participants in a given narrative context are aligned prompts different routes of processing – experiential versus reflective routes – and, by consequence, results in different outcomes. The theoretical model we develop in this paper contributes to our understanding of the language of perspective as grounded in interactive situations, and of its role in the narrative processes and effects which are grounded in oral storytelling patterns.

In the next section, we will first describe how viewpoint alignment is essential to conversational storytelling and how viewpoint in narrative discourse is linguistically varied throughout the story as it is brought by the narrator. Subsequently, we elaborate on functions of viewpoint alignment in terms of (a) the narrative styles, evoking acts, events, and sensations in readers; (b) the typical plot elements of the narrative; (c) the temporal pace of the narrative; (d) the routes of narrative processing; and (e) the different relations between character and reader therein. We argue that these functions of viewpoint alignment together show how narrative discourse is conceptually built on the oral speech situation, varying between demonstration and invasion of the narrative subjects and representing these subjects' discourse from outside and within the narrative events.

In the third section, this framework will be illustrated by analyzing a narrative example; the analysis shows how variation in the viewpoint alignment serves different expected outcomes of the narrative in readers and relates these aims to traditional rhetorical functions of storytelling. In the final section, anchors for future experimental studies will be discussed that may empirically test the effect of viewpoint alignment on narrative processes and impact.

VIEWPOINT ALIGNMENT FRAMEWORK

Joint Attention in Conversational Storytelling

Our point of departure is the consideration that human communication is fundamentally a matter of joint attention (Tomasello, 1995). Joint attention implies that two subjects align their viewpoints to jointly observe an "entity that they know they share" (Tomasello, 2008, p. 344), such as a third subject or an object, and in fact, the notion of viewpoint even presupposes the (actual or imagined) presence of this jointly observed entity (Tomasello, 2008). Joint attention skills develop early in childhood, in the first year of life, when infants start to direct and follow the visual attention of adults and, somewhat later, when they start participating in joint attention activities while acquiring their first words. Underlying the development of these skills is a child's understanding that people are *intentional agents* who purposefully direct their own attention and that of the child to certain aspects of their mutual surroundings but not to other (Tomasello et al., 1993, p. 498). Human communication is similarly intentional in that we communicate to teach and learn, to share experiences, to construct identities, and to inform, convince, move, and entertain one another. From this perspective, viewing one another as intentional agents can be considered a prerequisite of successful human communication.

In establishing joint attention, people share viewing directions toward an object/subject or situation/event in their surroundings, cued by each other's gestures and gaze directions (e.g., Mundy and Newell, 2007). This principle also applies to the narrative

communication mode, with the crucial difference that in storytelling, the object/subject or situation/event of attention is generally not directly observable in the here-and-now communicative context but has to be imagined in a specific context (Van Krieken et al., 2017). Exactly this is what is challenging about viewing directions in narrative. In an oral conversational story, for example, a person may direct the other person's attention to a specific setting as follows: "*I was at this beach party last week, and there was a huge bouncy castle with two kids jumping on it. They were actually jumping quite high. Then all of a sudden. . .*" The speaker may facilitate the establishment of joint attention and imagination with co-speech gestures, for example to indicate the size of the bouncy castle and the height of the jumps (Okada et al., 2013). Children acquire such gestures while narrating (Demir et al., 2015), but in written narratives, language is the only modality through which joint attention between the narrator and reader can be established. Either way, sharing viewing directions in narrative communication is primarily a cognitive act of viewing with "the mind's eye" (cf. Green and Brock, 2002), guided by linguistic cues. The event-indexing model argues that this imaginary viewing requires the mental construction of a "microworld of what is conveyed in the story" and that this microworld consists of five dimensions that readers keep track of: time, space, protagonist, causality, and intentionality (Zwaan et al., 1995, p. 292). Upon construction of this mental representation of the story world, readers can vicariously experience the story events, becoming "immersed experiencers" (Zwaan, 2004, pp. 35–37). Immersive experiences have also been described as simulations, and mentally running simulations of story events and situations has been argued to enhance readers' understanding of social life as well as their social cognitive skills (Mar and Oatley, 2008; Mar, 2018).

Once a cognitive representation of a story is established, readers can imaginatively view the story events from different angles. In this paper, we argue that the nature of this imaginary viewing can be both described and analyzed with the notion of *viewpoint alignment*. Viewpoint, or perspective, is deeply rooted in cognition as well as language and drives our bodily experiences of the world surrounding us and, as such, has strong explanatory force in understanding how human communication works. Yet, a theoretical framework that explicitly acknowledges the centrality of viewpoint in communication processes – and narrative processes in particular – and that explains these processes as a function of linguistic viewpoint phenomena remains to be developed. In the following, we propose that viewpoint alignment is a useful way of translating theoretical accounts of perspective that are developed in cognitive linguistics to the domain of narrative communication.

Establishing and Varying Viewpoint in Narrative

The notion of viewpoint has received much attention in cognitive linguistics, a research domain aimed at illuminating the relation between the linguistic and the cognitive representation of objects, events, and situations. A central assumption is that this representation is dependent upon the viewpoint or vantage point, i.e., the (physical or metaphorical) point from which an object/subject or situation/event is represented (Langacker, 1987). For example, the viewpoint in an utterance like *Go to the shopping mall* is fundamentally different from the viewpoint in an alternative utterance like *Come to the shopping mall* (see Fillmore, 1966). In the first utterance, the speaker's vantage point is positioned outside the shopping mall while in the second utterance it is positioned inside the shopping mall, implying that the speaker here-and-now is outside the shopping mall in the former versus inside in the latter. Similarly, the viewpoint of a sentence is typically located with the subject of that sentence rather than the object (Kuno, 1987). In the sentence *Amber called Matt*, the viewpoint is located with *Amber*, but in the passive equivalent *Matt was called by Amber*, the viewpoint is located with *Matt*. Note that different cues can have opposite effects, creating a complex viewpoint construal such as in *Her father called Amber*. Here, the grammatical viewpoint is juxtaposed to the referential situation: *her father* as description is semantically dependent on *Amber*. The possessive construction nuances the grammatical dominancy, profiles *Amber* as the main subject (semantically), and establishes the narrative viewpoint with her (Langacker, 1995). In short, the linguistic expression of an event affects hearers' or readers' cognitive construal of that event (Verhagen, 2007), and it also affects their comprehension of events in their discourse context (Black et al., 1979).

In narrative communication, the distance between the viewpoints of speaker/hearer and narrative characters is variable such that there may be more or less alignment. The scale can be considered to run from no viewpoint alignment at all at the one end to full viewpoint alignment at the other end, with different degrees of partial alignment in between. For instance, the narrative sentences *Amber went into the shopping mall* and *Amber came into the shopping mall* have different virtual vantage points, consequently including a viewing orientation outside the locker room versus an inside viewing orientation. In the cognitive representation of the virtual event, hearers are invited to join the narrative speaker (henceforth: narrator) in "viewing" along with Amber in the first sentence – to position themselves alongside Amber, going *into* a space of which there is no inside "view" provided – and from the front in the second – to "see" her entering *from* a space of which there is no inside "view" into a space in which hearers are imaginatively positioned themselves. As such, the first sentence (with *went*) invites the hearer to imagine that the narrator represents Amber's viewing from up close, while the second (with *came*) presupposes a greater (virtual) distance; in fact, the narrator can invite hearers to represent, regulate, and alternate narrative characters' viewpoints, using specific linguistic cues, even up to the point that represents no alignment. This is the case when narrator and character have different views on the narrative objects and events, like in an utterance such as *Amber had gone; she was not in the shopping mall.*

Typically, and opposite to a communicative speech situation here-and-now, where narrator and hearer may view a situation from different angles, the narrative communicative situation presupposes the conceptual alignment between the viewpoints of narrator and hearer/reader because the narrator is treated

as a conversational participant (Bortolussi and Dixon, 2003): as such, the narrator provides the cognitive "filter" through which the hearer gains access to the narrative. As a result, the hearer shares the narrator's view on the narrative objects and events as well as the narrator's view on the character, while the character's viewpoint and hence the character's view on the narrative objects and events is inaccessible to the hearer. To the extreme, viewpoint alignment is achieved in case of a first-person-present-tense narrative construal: an "I" narration, or a direct quotation within a narrative. In nearly all narratives, such quotes will occur. Fundamental of all direct quotation in narrative discourse is the complete alignment of viewpoints between the narrator and the quoted character because in essence, quotes presuppose an embedded speech situation within the narrative (Sanders and Van Krieken, 2019). As such, they depict or represent an interaction and firmly ground the story in an – be it past or imaginative – oral conversational situation.

Functions of Variation in Viewpoint Alignment

In most cases, the distance between the viewpoints of communicative participants in narrative discourse is variable and regulated by linguistic choices and contextual factors. As shown by the examples above, the process of viewpoint alignment is regulated by a range of linguistic phenomena. These include (but are not limited to) choices in grammatical construction, verb tense, expressions of perception and cognition, and choice of referential expressions (Van Krieken et al., 2017).[1] These phenomena are indicative of variations in (a) narrative style, (b) narrative plot structure, (c) temporal pace, (d) narrative processing, and (e) narrator–character relation, which are each grounded in the interactive speech situation and which guide the readers' representation of the narrative toward the intended outcomes of the narrative.

Narrative Styles: Demonstration and Invasion

The use of specific linguistic viewpoint markers results in a particular narrative style, corresponding with the particular viewpoint alignment position. A useful distinction is that between a *demonstrative* and an *invasive* style (Sanders, 2017). The demonstrative style that typically entails viewpoint nonalignment is characterized by an external view on a character and lacks internal viewpoint indicators. Instead, the character's viewpoint is expressed by, for example, descriptions of the character's acts and other events observable "from the outside" (cf. *landscape of action*, Bruner, 1990). An example is *Amber went to the shopping mall and spent the afternoon in the hardware store. In the evening, she had dinner with her father while discussing alternative ways to refurbish the bathroom.*

By contrast, the invasive style that typically entails viewpoint alignment is characterized by linguistic elements that provide access to the inner world of a narrative character (cf. *landscape of consciousness*, Bruner, 1990), such as verbs and adjectives referring to the character's mind or emotions, thought reports, and stream of consciousness (Cohn, 1978; Van Krieken et al., 2017). An example is *Amber went to the shopping mall. Upon entering, she spotted a hardware store. It was difficult to find the hammer and the screwdriver she needed. That evening, she was proud to show her father the utensils she bought over dinner. Refurbishing the bathroom was her next goal.*

Quotes have different functions depending on the narrative's style and can be distinguished in documenting versus dramatizing quotations (Van Krieken et al., 2016). Documenting direct quotes are typical for the demonstrative style: they *evaluate* events and situations within the story world from a narrative-external character viewpoint. An example is *Amber went to the shopping mall and spent quite some time. "I needed a hammer and a screwdriver, but it was not easy to find them", she told her father*. Such quotes align the viewpoint of the narrator with the character's viewpoint at a point in time *after* the narrative events, enhancing the reflective processing.

Dramatizing direct quotes, by contrast, are typical for the invasive style: they *mimic* events and situations within the story world from a narrative-internal character viewpoint (Van Krieken et al., 2016). An example is *Amber went to the shopping mall. Upon entering, she spotted a hardware store and started looking for the hammer and screwdriver she needed. "If they are not in this corner, where are they?" she wondered. "This is ridiculous. Why do they hide these things?" she spoke out loud.* Such quotes align the viewpoint of the narrator with the character's viewpoint *during* the narrative events, enhancing the experiential processing.

Narrative Plot Structure: Climbing and Descending

Narratives may have a general demonstrative or invasive style, but more typically, narratives show a dynamic alternation between both styles in adjustment to the story structure and the progress of narrative time. Both in fictional and non-fictional narratives, the alignment between the viewpoints of narrator and character(s) is not fixed at all but may be aligned closer to and further away from one another from one narrative sentence to the next (Van Krieken et al., 2016). This alternation is enabled by the distinctive plot structure of narratives, which is inherently *temporal*. Time can be considered "the organizing axis" of storytelling (Neiger and Tenenboim-Weinblatt, 2016, p. 2). In producing and processing stories, narrative communicators move along this axis, thus traveling through narrative time (Sanders and Van Krieken, 2019).

Research on oral storytelling has clarified how different temporal frames serve different functions. Labov and Waletzky (1967) identified their basic story structure consisting of six elements: an *orientation* that paints the setting of the story in terms of time, place, and participants; *complicating actions* leading up to a *critical event* (the peak of the story), followed by a *resolution* (how the story ended) and *coda* (return from the story world to the here-and-now of the conversational setting). Finally, *evaluations*, often in the form of quotations, can be used throughout a story to express personal experiences and emotions related to the story events. Essentially, this pattern is

[1] In oral and (audio)visual narratives, the degree of viewpoint alignment is also regulated by extralinguistic phenomena, such as gesture and eye gaze (in spoken narratives; Parrill, 2010; Sweetser and Stec, 2016), camera positions (in filmic narratives; Alber, 2017), and pictographic cues (in comics; Borkent, 2017).

chronological in nature, but variations in order may increase or release tension (Knobloch et al., 2004).

Temporal Pace: Acceleration and Deceleration

The progression of time in natural storytelling is subject to patterning of climbing and descending narrative peaks (Fleischman, 1990; Fludernik, 1991). Typically, temporal pace will decelerate upon the approach of narrative peaks and accelerate toward the end of plot structures, corresponding with viewpoint alignment variation. Deceleration and acceleration can be understood in terms of narrative duration: how is narrative time related to historical (story) time? Genette (1971) (p. 30) distinguishes between historical *récit*, which reduces the narrative duration compared to historical time (typical for plot elements such as complicating actions and resolution); dramatic *scene*, especially dialog, where narrative duration equals historical time (typical for plot elements such as critical event – peak – and evaluations); *stasis*, where narrative duration exceeds historical time (typical for plot elements such as complicating actions and evaluations); and *ellipsis*, which indicates historical time not covered in the narrative (typical for plot elements such as orientation and coda). Note that viewpoint alignment between narrator and character varies with these temporal frames: in case of *scene*, alignment is maximized, while it is reduced in *statis*, minimized in *récit,* and absent in *ellipsis*.

Narrative Processing: Reflective and Experiential

The way a story is structured in terms of plot structure and the corresponding temporal frames is functional to the intended processing of the story as well as the story's intended outcomes (Sanders and Van Krieken, 2018) and is guided by the story's style of narration. Crucial is the degree of alignment between the viewpoint of the character whose story is narrated in the story world, and the viewpoint of the hearers/readers (henceforth: readers) who imaginatively shift their viewpoint to the story world, guided by the speaker's narration. When viewpoints are not aligned, such as in the demonstrative style "from the outside," readers are led into a *reflective route*. Narratives, or parts thereof, that do not align the narrator's viewpoint with the character's viewpoint, offer readers a non-restricted view on the story world in which they are free to take a position at any point in narrative time, looking forward or backward, hence with room for own their viewpoint and for reflection; typically, a third-person past-tense narrative construal with a focus on story actions creates room for this reflective representation. Such a reflective route is expected to primarily evoke thoughts, evaluations, and considerations in readers.

In case of alignment between the viewpoints of narrator and character, such as in the invasive style "from the inside," they move along the same path, and readers are invited to co-experience the story events as they unfold. Thus, the hearer's "view" on the narrative is restricted, like a tunnel view: what is yet to be narrated can only be accessed and processed from the point of view of the character whose viewpoint is represented. Narrator and character thus share the same conception of past, present, and future as they jointly travel through the story world. A narrative construal with a focus on character sensations creates a limited representation in which the narrative time is experienced at the pace of the character's experience. Consequently, the readers' view of the narrative events, guided by the narrator, is restricted, and readers are thereby led into an *experiential route*. This route can be expected to primarily evoke sensations, feelings, and imaginations in readers (e.g., Zarantonello et al., 2013). Such experiential routing may be most functional around central peaks in storytelling, while reflective routing may be found in the lower peripheral areas of stories, and both routes are reflected in the relation between narrative characters and readers. Note that while these distinctions are conceptually sharp, actual narratives exhibit various combinations of the styles. In that sense, the distinctions might perhaps be said to represent opposite endpoints of a continuum.

Character and Reader: Spectating and Identifying

When viewpoints are not aligned, the nature of the mediated relation between the narrative character and the reader is mostly observatory: guided by the narrator, readers meet the character as spectators (see Oatley, 1999, p. 445). Spectatorship does not refer to the adoption of another person's viewpoint but to a state in which readers relate their own viewpoint to the viewpoint of another person through observation. As Oatley (1999) (p. 445) puts it: "The reader becomes an unobserved observer in scenes of the lives of characters in the story world." In the reflective style, narrative time may be stretched or, by contrast, compressed, such that the character's experience of time is not guiding the representation, but a *reflection on* this experience in a slower or higher pace.

By contrast, when viewpoints are aligned, the nature of the mediated relation between the narrative character and the reader is one of *identification*. Rooted in psychological and developmental research, the original concept of identification refers to a process in which one temporarily sets aside his own viewpoint and imagines experiencing reality through someone else's viewpoint (Freud, 1989 [1940]). In its original sense, then, identification is similar to viewpoint alignment. Throughout the years, the meaning of identification has been stretched – most notably by communication scholars – to include the audience's adoption of a character's emotions, attitudes, goals, and identity (e.g., Brown, 2015). This can be seen as the *result* of narrow viewpoint alignment: because the reader is "forced" to take on the viewpoint of the character, there is little to no room to *not* adopt the character's emotions and attitudes.

VIEWPOINT ALIGNMENT ANALYSIS

In the following, we will demonstrate the premises of dynamic viewpoint alignment outlined above by applying this concept to a corporate journalistic narrative that has aspects of both spoken and written narrative and that exemplifies both demonstrative and invasive styles. In doing so, we show how a single story can evoke a processing route that is partly reflective and partly experiential and also how the framework can account for viewpoint alignment processes evoking various outcomes.

We use a sample text, which is a narrative from a Dutch corporate context, the fire brigade, called *Fire & Fire Brigade*[2]. For the complete text (the original in Dutch and an English translation), please see the **Appendix**. Other aspects of this text are discussed in Van Krieken and Sanders (forthcoming).

Interactive Story Context and Communicative Goals

Corporate journalistic media, such as *Fire & Fire Brigade*, are published by organizations to strengthen the commitment of staff to their organization and the quality of work in the organization by stimulating mutual communication (Van Ruler, 2005). Such magazines also serve, not infrequently, as a *public relations* means of communication for other stakeholders, such as the staff's social context, but also governmental services, financiers, clients, and other relations, giving a clear picture of "what goes on" in the organization. It means that inherently, a *multiple audience design* is involved in corporate magazines; even if a corporate magazine is only distributed internally, it will always take into account any other potential reader groups.

The story central to this analysis is an edition of a recurring column, entitled *Under the Helmet*. The title expresses the idea that the column is about events of the uniformed fire brigade staff and also about the way these events are experienced. Such a column thus has two functions: firstly, to share the, at times, profound experiences of a part of the organization that fulfills a core function (in this case: fighting fires and saving people from dangerous situations) with other parts of the organization (in this case: office functions and other facilitating divisions of the fire brigade), and secondly, to show how such profound experiences are shared and processed. The storytelling itself also plays an exemplary role: talking about these experiences is permitted. The column deals with uniformed "operations" in which an active unit of the fire brigade is deployed because of an emergency situation; in this case, a young child that had become trapped in a fairground ride. The fire brigade's task in such a case is to free the child as quickly and as unharmed as possible. Because of this urgency, in the setting of a public place with a great many bystanders, this apparently was a profound experience for the fire fighters involved. Sharing such an experience legitimizes the fire brigade as an organization: *this is what we do, this is what we are here for*. All employees and volunteers in the organization contribute from their own position to the performance of this core duty, and a story like this can strengthen their belief in the usefulness of their work. Of course, other reader groups such as the home front may also share in this legitimizing function. Importantly, such stories are implicitly educational about the moral frameworks within which this core duty is performed: *this is how we do this, and it's the right way*. As such, the story is aiming at several interconnected communicative goals: informing (telling what the work entails); instructing (showing how to do the work); and emotionally and morally engaging (emotional engagement objectivemaking the work experiences tangible, and their importance evident).

Viewpoint Alignment and Experience of Time

In this short story, the journalistic narrator firmly places the perspective with a single character, the experiencing subject, who is referred to in the third person: Klaas Wim Jansen[3], first commanding officer of a unit that was involved in the operation concerned. The story is told almost completely chronologically by the narrator, who does not take part in the story and whose viewpoint alternatingly is not aligned, partially aligned, and completely aligned with the central character in this story. Parts in which Jansen is allowed to tell his story in direct quotes as an embedded narrator alternate with parts in which the journalistic narrator narrates about Jansen as a character in the main story; put in journalistic terms: Jansen is the source, who tells his story to a corporate journalist, who reports it on behalf of the organization. Jansen's oral story is central to the text, and while it concerns his experiences, it is essentially characterized by its chronological structure (Labov and Waletzky, 1967). According to the Structural Affect Theory, this structure evokes more suspense than non-chronological story structures (Brewer and Lichtenstein, 1982). Empirical evidence shows that chronological structures do indeed generate suspense in readers (Knobloch et al., 2004) and also prompt stronger emotional reactions and enhance their experience of being transported into the story world (De Graaf and Hustinx, 2011). It can therefore be expected that readers of the fire brigade story experience a strong sense of transportation and suspense as the events unfold in the order in which they originally occurred to Jansen; thus, readers are invited to share Jansen's viewpoint and emotions.

Linguistic devices such as verbs of perception enable the reader to see and experience the events through his eyes, for example: *When he hears en route that a child is trapped under a fairground attraction, it goes quiet in the fire engine.* Or: *At first, he doesn't see the child.* Particularly in the present tense, verbs of perception as well as verbs of cognition are claimed to stimulate and guide readers' identification with narrative characters by establishing a direct link with their minds (Damsteegt, 2005; Sanders, 2010), in this specific narrative: with Jansen. The narrator alternates such internal viewpoint representations with direct quotations, but not with stream of consciousness. In other words, it is an invasive story – readers imaginatively look through the eyes of the character – but by choosing to do this in the form of direct quotations, the story also has a demonstrative nature: everything is said and demonstrated clearly. Such alternation between invasion and demonstration can be considered characteristic of journalistic stories, which aim at engaging the audience while emphasizing the truthfulness and factuality of what is narrated (e.g., Van Krieken et al., 2016).

In all but one place in the story, the present tense is used by the narrator, which gives a film-like effect which serves this story well,

[2]https://issuu.com/sduuitgeversbrandbrandweer/docs/bb201710, p. 31 (accessed January 24, 2021).

[3]For reasons of privacy, we use a pseudonym for the officer's real name.

since it is largely about a single exciting scene: the events appear immediately before the readers' eyes, in a pace that is increasingly congruent to the lived events. The present tense has been found to facilitate identification (Macrae, 2016). Because the readers take in the experiences in the order in which Jansen had them, for the largest parts of the narrative, viewpoints are aligned – but not completely and not continuously. Both the past tense used in Jansen's direct quotations, AND the 3rd person perspective in the narrator's parts indicate some distance between Jansen's viewpoint and the reader's. In the direct quotes the character is "*I*" but not *present*, and in the narrated utterances he is *present* but not "*I*."

Viewpoint Alignment and Direct Quotation

The direct quotations in this story do not only picture persons and their experiences in a lively way but also evaluate that which took place. In this story, the quotations express how grave, important, and profound these professional experiences were in the perception of those involved, as if it were something to learn from for future use (Boyd, 2009). Throughout the story, direct quotations are applied with two effects: building tension and transmitting emotions. The tension is visible at the beginning of the story: "*We responded and drove to the fairground, without me having any idea of what had happened.*" With this quote, the narrator invites readers to share the embedded narrator/character's ignorance of the situation. "'*I sympathize with you guys, lots of luck', added the operator. Operators never say that. Afterwards, it turned out he'd seen images from a surveillance camera in the vicinity.*" Readers are thus stimulated by the narrator to share the viewpoint of Jansen in wondering what the operator knows that he does not yet know at this point in the story. Including the perspective of another character in the story, through direct quotations by the main character who himself quotes yet another person, helps building tension and attention. Thus, the chronology is rhetorically disturbed by combining a flash-forward and flashback, and a partial discrepancy in viewpoint alignment is established between the viewpoint of the narrator, the embedded narrator/character, and the reader. Quotations again are used to make the impact clear: *"I could see from the look on the traffic controller's face (...) that the incident was a serious one."*

All direct quotes are narrative-external, and the majority propels the narrative time in a reduced way such that the historical time exceeds the duration in the narrative (*récit*), for example in *"We responded and drove to the fairground (...)"*. This phrase represents Jansen as being quoted at a later point in time, looking back at acts and events at the narrative timeline. This is a construal typical for journalistic narratives, where it serves to legitimize the credibility of the narrative reconstruction on the basis of an interaction between journalist and source (Van Krieken et al., 2016). By contrast, narrative-internal quotes are evaluations and observation made at the time of the narrative that put the time to a halt (*scene*). In the sample narrative, few internal quotes are used and these are all embedded in external quotes, creating a multilayered temporal pace structure in which narrative time decelerates within an accelerated frame. This corresponds to a dynamic alternation between low and high viewpoint alignment. It is worth noticing that indirect representations are used when other characters than the embedded narrator are involved ("*Together with [...] he decides that*" in the voice of the main narrator, "*Neither of us knew who she was*," in the embedded narrator's voice). In other words, only the main protagonist Jansen is represented with his own voice, which indicates that viewpoint alignment is shifting between these embedded (experiential, participating) and the external (reflective, journalistic) viewpoints.

Particularly interesting is the insertion of embedded narrative-internal quotes that express what *did not* happen but *could have happened*, not uncommon as a type of evaluation of the plot (Labov, 1997). The reflection of unreal versions of the plot puts the narrative time to a halt (*scene*) and enhances the critical point's meaning. An example is provided by the underlined part of the quote in the following excerpt, which emphasizes the tension and risk: *"It wasn't until we removed the covers from underneath the ride, that we could see properly how his little arm was stuck. This is never going to work, I thought. It looked really complicated"*. This present-tense quotation, embedded in an external quotation in the past tense, *shows*, rather than *tells*, how Jansen felt at the moment. Jansen introduces himself in the story as a thinking character, much like the inner thought in the story's opening scene ("*'Oh shit', was my first thought*"). An additional function of such quotations is that readers can recognize themselves, due to the language use of the characters, who are representatives of the primary target grouptarget group, in the customary word choices and register of the quotations used. Notably, such quotes do not presuppose that the quoted material was uttered or thought in exact the same words. Clark and Gerrig (1990) explain that quotations are *non-serious* actions, that is, not actually occurring as such at the moment the quotation is uttered, but demonstrations of the particular narrative moment through selective depiction rather than description (Clark and Gerrig, 1990, p. 764). Here, these quotes are uttered at some point during the interaction of the source/character Jansen with the journalistic narrator.

Direct quotation in present tense, guiding toward complete viewpoint alignment, is typical for the story's peak, in this case, the child's liberation. The narrative prepares for the peak by quoting Jansen's observations and emotions: *"When it was shouted downwards from above that the arm was loose, the relief was great. You really heave a sigh then."* These utterances are remarkable in that they combine three linguistic viewpoint techniques: a physical demonstration, a de-activation of viewpoint, and a generic statement. In the first place, the character himself explains his position in the physical space at the exact moment of the liberation: people shouting down [implicitly: to where I was at that moment], from above. Such merging of viewpoints, representing the character's physical position at the peak of the story, enables readers to share the experience of this crucial event with the character from the inside, as it were, and imagine it in an embodied way (Labov and Waletzky, 1967). This effect is established by the impersonal passive construction (*when it was shouted*) and nominalization (*the relief was great*), which, in

combination, deactivate the potential viewpoints of other agents in the scene, inviting the reader to keep identifying with this character rather than with others (note that, stronger than in English, Dutch passive signals that the causer's point of view should not be taken; see Cornelis, 1996, p. 247). To underscore the generalizability of the experienced feelings, the character evaluates the relief in a general expression: *"You really heave a sigh then."* The use of the present tense and the generic "you" combine to give the statement a broader validity and express the fact that not only was the relief great but also it was an understandable and justified emotion, and that it is not surprising that anyone who goes through such an experience takes it personally (see also Van Krieken et al., 2015). Moreover, readers have been shown to internalize emotions more strongly when a second-person pronoun (rather than a first person pronoun) is used (Brunyé et al., 2011). Hence, an intense co-experience of Jansen's relief is facilitated by the use of "you." These combined techniques of impersonalization and generic "you" is in fact used several times in the story, as the following excerpt shows: "*Then there's a commotion a bit further away [...]. You're just taken aback for a moment [...]. Soon after that, all attention went back to the operation [...].*" As such, this viewpoint deactivation seems to be indicative for the text's narrative strategy as a whole.

Viewpoint Alignment and Plot Structure

The analysis shows that the degree of viewpoint alignment may change during the story's journey, such that narrative events in a given temporal frame may be processed differently than events in other frames. Fundamentally, the viewpoint alignment appears to be closer toward the core of the narrative, resulting in an experiential time conceptualization, whereas viewpoint alignment is less close at the beginning and end of the narrative, resulting in a distanced, or reflective, time conceptualization.

Accordingly, it appears that different temporal frames serve different functions. The *tellability* of the current story is determined by the uniqueness and impact of the events related: a child that is stuck in a machine, the tension that accompanies the rescue, and the vulnerable nature of the young victim give the story high attention value and offer many opportunities of empathizing with the events and persons involved. The story plot has a clearly recognizable basic structure, the chronological list of events starting with the *orientation*: the indication of the location and nature of a P1 report: Priority 1 (urgency). The events follow each other in quick succession: the unit deploys, is shown the way to the scene of the incident, arrives there, and assesses the situation: a child is trapped in a fairground ride. This problem is quite clearly the *complication*. With the help of various other units, the situation is analyzed, a plan of action is determined, and the child is freed, one step at a time. The *critical event* is when the fire brigade gets the child's arm free. This moment of freeing merges into the *resolution* of the story: the child is transferred to the ambulance. The *resolution* for the fire brigade units involved is a different one: they go off duty, which means they are no longer on call, to evaluate what has happened; they subsequently go back on duty and complete their shift. The story ends with a *coda*: retelling the experience at home.

Viewpoint Alignment and Narrative Theme

In this story, *looking* is a recurring theme, and crucial is looking at, and looking through the eyes of, the trapped child: *"(...) the look in his eyes spoke volumes. His eyes were screaming for help. Now and then he closed them (his eyes). When we started talking to him, he opened them and looked at us with that expression."* This quotation occurs at a very compacted point in the time, in which the tension and emotion of the story come together without them having to be pointed out as such. Helpers and those helped look at each other repeatedly, and lengthily, and through this looking, there is a temporary but deeply felt contact, which expresses the great concern the professionals feel for the victim. In this particular part of the narrative, the narrator both invites and enables readers to look along – to literally align their viewpoint – with the main character and his colleagues, their attention drawn toward to the calm eyes of the child that form the focal point of the story. Such archetypal themes are universally human and make the story universally recognizable and profound (Campbell, 1949; Sanders and Van Krieken, 2019).

Notably, s*torytelling* itself is a recurring theme motive in the narrative analyzed in this article. First, during the evaluation at the fairground: *and you want to share your story*. Then: *The first two TS trucks and the HV go off duty, and all units return to the fire station for the evaluation. Everyone tells their story there again. When I got home, I told my story again.* The value being conferred here is that talking about the impressions and emotions that are unavoidable in such profound experiences is acceptable and recommendable. As such, the story, itself the *product* of storytelling in the organization, promotes storytelling as a *process*. Probably, in a traditionally masculine organization such as the fire brigade, talking about emotions arising from professional experiences is increasingly common. A corporate journalistic narrative like the current underlines this development as being good and desirable. The tellability of such stories is therefore also defined by their emotional and moral impact: telling people about that is worthwhile, much like conversational stories are being exchanged between people who want to share experiences in order to persuade each other of particular views on reality (Boyd, 2009).

Viewpoint Alignment and Rhetorical Outcome

To conclude, our analysis has shown that demonstrative and invasive storytelling styles are combined and alternated according to the narrative's plot structure and temporal frames, guiding readers closer to and then further away from characters in the story and finally bringing them to a moral conclusion. Thus, viewpoint alignment appears functional to different rhetorical outcomes of the story (Sanders and Van Krieken, 2018), with *phronesis* – i.e., "a form of moral sense making of the self that advances one's practical wisdom and prudence" (Sanders and Van Krieken, 2018, p. 6) reflecting nonalignment, and *catharsis* – "deeply emotional experiences" (Sanders and Van Krieken, 2018, p. 13) – reflecting full alignment. More specifically, the main function of the experiential route is *pathos*, i.e.,

TABLE 1 | Characteristics of different degrees of viewpoint alignment.

	Degree of viewpoint alignment			
	− Spatiotemporal *distance* between viewpoints; object of conceptualization is viewed from *different* positions.	+/− Spatiotemporal *nearness* between viewpoints; object of conceptualization is viewed from a *similar* position.	+ Spatiotemporal *collapse* of viewpoints; object of conceptualization is viewed from one and the same position *outside* the narrative.	++ Spatiotemporal *collapse* of viewpoints; object of conceptualization is viewed from one and the same position *within* the narrative.
Characteristics:				
(a) Narrative style (Bruner, 1990; Van Krieken et al., 2016; Sanders, 2017)	Demonstrative narration of observed events (acts and situations)	Invasive narration of physical and psychological events (sensations, imaginations)	Documenting quotation (external direct quotes)	Dramatizing quotation (internal direct quotes)
(b) Narrative plot structure (Labov and Waletzky, 1967; Fludernik, 1991)	Orientation, coda	Complicating actions, resolution	Complicating actions, evaluations	Critical event (peak), evaluations
(c) Temporal pace (Genette, 1971)	Still/*Ellipsis*	Slower/*Récit*	Faster/*Stasis*	Realistic/*Scene*
(d) Processing route (Zarantonello et al., 2013)	Reflective	Experiential	Reflective	Experiential
(e) Mediated relation between character and reader (Oatley, 1999; Van Krieken et al., 2016)	Spectatorship/distant	Identification/close	Legitimizing/distant	Merging/close
(f) Expected rhetorical outcome (Kearney, 2007; Sanders and Van Krieken, 2018)	*Phronesis*	*Pathos*	*Eleos*	*Catharsis*

the reader's empathizing with positive and negative feelings of positive expectation and agony (Kearney, 2007). These experiences are legitimized by evaluations and quotes that are fueled by *eleo*s or fear (Kearney, 2007), ensuring that these experiences are grounded and justified. On the peak of the narrative, physical experiences and positions are narrated in a merging style, functional to the story's climax or *catharsis* in the reader's experience of intense emotional relief (Koopman, 2013). The communicative goals of the narrative are advanced by these rhetorical functions: phronesis will help in informing about which complications the work entails, while pathos and eleos assist in instructing by showing how cautiously the work is done; and most importantly, catharsis serves to engage emotionally and morally with the work by making its experiences tangible, and their importance evident.

CONCLUSION AND DISCUSSION

In this paper, we developed a model of viewpoint alignment to explain for the processes and outcomes of narratives that can be traced back to the oral origins of storytelling structure. In **Table 1**, an overview is generated of the categories we distinguished to explain the variation in viewpoint alignment. The four columns represent the two main modes of narration – demonstrating versus and invading the narrative subjects – and the two modes of direct quotation therein – representing the narrative subjects' discourse from outside and within the narrative events – which, in various combinations and alterations, will dynamically guide the reader in reflective and experiential routes of narrative processing.

As conceptual model, viewpoint alignment is intended to complement influential theories of narrative communication, such as the Entertainment-Education and Elaboration Likelihood Model (Slater and Rouner, 2002), the Transportation-Imagery Model (Green and Brock, 2002), and the Mental Models approach (Busselle and Bilandzic, 2009). Even though it may not be explicitly acknowledged in these theories, the notion of viewpoint alignment is capable of explaining and predicting the mechanisms they describe. For instance, fundamental to the Transportation-Imagery Model and Mental Model approaches is the projection of the reader's viewpoint into the story world as a prerequisite for their engagement with the story characters and events and their sense of "getting lost" in the story. In a different way, the "central" processing route as distinguished in the Entertainment-Education and Elaboration Likelihood Model implies more room for one's own viewpoint in the narrative processing compared to the "peripheral route." Importantly, our model explains how such narrative processes are guided by linguistic structures that regulate the degree of viewpoint alignment.

Our model is furthermore distinct from previously developed theories in its assumptions about the nature of narrative processes. First, the Transportation-Imagery Model describes narrative processing as a process different from central and peripheral processes, but in our approach, experiential and reflective routes of (narrative) communication can be seen as two *types* of narrative processing routes. Second, the viewpoint

alignment approach furthermore does not assume that there are differences in *strength* or *depth* of processing of the different routes (contrary to the assumptions of the Entertainment-Education Elaboration Likelihood Model). Experiential and reflective routes could both be equally intense. Crucially, both types are not mutually exclusive but can complement one another, as we have posited in our analysis of the corporate narrative. In our view, this is an essential assumption in understanding the range of outcomes a narrative can have (e.g., on the reader's beliefs, attitudes, knowledge, and behavior) and how such static outcomes are the result of a dynamic process.

Questions about the relation between the dynamic processing of narrative discourse and the static outcomes could be answered in future experimental studies. The design of such studies can be informed by the narrative characteristics summarized in **Table 1**. For example, by manipulating the characteristics of demonstrative narrating, invasive narrating, and direct quotations, studies could test their effect on readers' degree of identification with the narrative character and the potential subsequent persuasive impact of the story. The use of online measures to assess readers' identification with the character, such as fMRI or galvanic skin responses (Van Krieken et al., 2017), could help to establish the degree to which this process varies during reading and to pinpoint the precise linguistic elements that increase (or decrease) identification. Studies in this direction are essential to test the expectations of our viewpoint alignment framework and to advance our understanding of the function and impact of perspective and perspective-taking in the language of stories.

At a conceptual level, the model developed in this paper supports the notion that time and perspective are intricately connected in the representation of narratives (see Van Krieken et al., 2019). Our model fleshes this idea further out by proposing that this connection, too, is dynamic by nature and develops over the course of a narrative: narrative time can be presented in a condensed manner in a given part of the story but in an expanded way in another part, which is intrinsically related to a change in the degree of viewpoint alignment. This conceptualization is firmly rooted in oral storytelling patterns and helps to understand how narratives can fulfill multiple communicative goals at the same time, which is often the case for nonliterary narratives such as corporate stories. Understanding narrative processing as a dynamic alignment of viewpoints that alternately invite readers to experience or reflect on the story events, explains how a single story can inform, instruct, move, persuade, and empower readers all at once.

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

KK and JS contributed equally to the manuscript and contributed to introduction, framework, analysis, conclusion, and discussion. Both authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by a Veni grant from the Netherlands Organization for Scientific Research (NWO), awarded to KK (project number 275-89-038). This grant funds 3 years of independent research.

ACKNOWLEDGMENTS

We would like to thank the two reviewers of our manuscript for their constructive and insightful comments and suggestions. We also thank Renske van Enschot and Joyce Karreman for their comments on an earlier version of a part of the analysis presented here. In addition, we thank the publisher, SDU, and the writer, Jildou Visser, of the original text from Brand & Brandweer for their graceful permission to reproduce and analyse this text, previously published in Brand&Brandweer 2017-10, page 31/item 'onder de helm'.

REFERENCES

Alber, J. (2017). "The representation of character interiority in film: cinematic versions of psychonarration, free indirect discourse and direct thought," in *Emerging Vectors of Narratology*, eds P. K. Hansen, J. Pier, P. Roussin, and W. Schmid (Berlin: Mouton de Gruyter), 265–284.

Avdi, E. (2008). Analysing talk in the talking cure: conversation, discourse, and narrative analysis of psychoanalytic psychotherapy. *Eur. Psychother.* 8, 69–87.

Barker, R. T., and Gower, K. (2010). Strategic application of storytelling in organizations: toward effective communication in a diverse world. *Int. J. Bus. Commun.* 47, 295–312. doi: 10.1177/0021943610369782

Black, J. B., Turner, T. J., and Bower, G. H. (1979). Point of view in narrative comprehension, memory, and production. *J. Verb. Learn. Verb. Behav.* 18, 187–198. doi: 10.1016/s0022-5371(79)90118-x

Borkent, M. (2017). Mediated characters: multimodal viewpoint constructions in comics. *Cogn. Ling.* 28, 539–564. doi: 10.1515/cog-2016-0098

Bortolussi, M., and Dixon, P. (2003). *Psychonarratology: Foundations for the Empirical Study of Literary Response*. Cambridge, MA: Cambridge University Press.

Boyd, B. (2009). *On the Origin of Stories: Evolution, Cognition, and Fiction*. Harvard: Harvard University Press.

Braddock, K., and Dillard, J. P. (2016). Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. *Commun. Monogr.* 83, 446–467. doi: 10.1080/03637751.2015.1128555

Brewer, W. F., and Lichtenstein, E. H. (1982). Stories are to entertain: a structural-affect theory of stories. *J. Pragmat.* 6, 473–486. doi: 10.1016/0378-2166(82)90021-2

Brown, W. J. (2015). Examining four processes of audience involvement with media personae: transportation, parasocial interaction, identification, and worship. *Commun. Theory* 25, 259–283. doi: 10.1111/comt.12053

Bruner, J. (1990). *Actual Worlds, Possible Minds*. Harvard: Harvard University Press.

Brunyé, T., Ditman, T., Mahoney, C., Augustyn, J., and Taylor, H. (2009). When you and I share perspectives: pronouns modulate perspective taking during narrative comprehension. *Psychol. Sci.* 20, 27–32. doi: 10.1111/j.1467-9280.2008.02249.x

Brunyé, T., Ditman, T., Mahoney, C., and Taylor, H. (2011). Better you than I: perspectives and emotion simulation during narrative comprehension. *J. Cogn. Psychol.* 23, 659–666. doi: 10.1080/20445911.2011.559160

Busselle, R., and Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychol.* 12, 321–347. doi: 10.1080/15213260903287259

Campbell, J. (1949). *The Hero With a Thousand Faces*. Novato, CA: New World Library.

Clark, H., and Gerrig, R. (1990). Quotations as demonstrations. *Language* 66, 764–805. doi: 10.2307/414729

Cohn, D. (1978). *Transparent Minds: Narrative Modes for Presenting Consciousness in Fiction*. Princeton, NJ: Princeton University Press.

Cornelis, L. (1996). English and Dutch: the passive difference. *Lang. Sci.* 18, 247–264. doi: 10.1016/0388-0001(96)00018-6

Damsteegt, T. (2005). The present tense and internal focalization of awareness. *Poet. Today* 26, 39–78. doi: 10.1215/03335372-26-1-39

De Graaf, A., and Hustinx, L. (2011). The effect of story structure on emotion, transportation, and persuasion. *Inform. Design J.* 19, 142–154. doi: 10.1075/idj.19.2.05gra

De Graaf, A., Sanders, J., and Hoeken, H. (2016). Characteristics of narrative interventions and health effects: a review of the content, form, and context of narratives in health-related narrative persuasion research. *Rev. Commun. Res.* 4, 88–131.

Demir, ÖE., Levine, S. C., and Goldin-Meadow, S. (2015). A tale of two hands: children's early gesture use in narrative production predicts later narrative structure in speech. *J. Child Lang.* 42, 662–681. doi: 10.1017/s0305000914000415

Devine, J. R., Quinn, T., and Aguilar, P. (2014). Teaching and transforming through stories: an exploration of macro- and micro-narratives as teaching tools. *Ref. Librar.* 55, 273–288. doi: 10.1080/02763877.2014.939537

Fillmore, C. J. (1966). Deictic categories in the semantics of 'come'. *Found. Lang.* 2, 219–227.

Fleischman, S. (1990). *Tense and Narrativity: From Medieval Performance to Modern Fiction*. Austin, TX: University of Texas Press.

Fludernik, M. (1991). The historical present tense yet again: tense switching and narrative dynamics in oral and quasi-oral storytelling. *Text Interdiscipl. J. Study Discourse* 11, 365–398.

Freud, S. (1989). *An Outline of Psycho-Analysis*. New York, NY: W.W. Norton & Company Inc.

Genette, G. (1971). "Time and narrative in *Á la recherche du temps perdu*," in *Narrative Dynamics: Essays on Time, Plot, Closure, and Frames*, ed. B. Richardson (Columbus, OH: Ohio State University Press), 25–34.

Georgakopoulou, A. (2014). Small stories transposition and social media: a micro-perspective on the 'Greek crisis'. *Discourse Soc.* 25, 519–539. doi: 10.1177/0957926514536963

Gerrig, R. J. (1993). *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. Yale: Yale University Press.

Green, M. C., and Brock, T. C. (2002). "In the mind's eye: transportation-imagery model of narrative persuasion," in *Narrative Impact: Social and Cognitive Foundations*, eds M. C. Green, J. J. Strange, and T. C. Brock (Mahwah, NJ: Erlbaum), 315–341.

Kearney, R. (2007). Narrating pain: the power of catharsis. *Paragraph* 30, 51–66. doi: 10.3366/prg.2007.0013

Knobloch, S., Patzig, G., Mende, A.-M., and Hastall, M. (2004). Affective news: effects of discourse structure in narratives on suspense, curiosity, and enjoyment while reading news and novels. *Commun. Res.* 31, 259–287. doi: 10.1177/0093650203261517

Koopman, E. (2013). The attraction of tragic narrative: catharsis and other motives. *Sci. Study Literat.* 3, 178–208. doi: 10.1075/ssol.3.2.03koo

Kuno, S. (1987). *Functional Syntax: Anaphora, Discourse and Empathy*. Chicago: University of Chicago Press.

Labov, W. (1997). Some further steps in narrative analysis. *J. Narrat. Life Hist.* 7, 395–415. doi: 10.1075/jnlh.7.49som

Labov, W., and Waletzky, J. (1967). "Narrative analysis: oral versions of personal experience," in *Essays on the Verbal and Visual Arts*, ed. J. Helm (Seattle, WA: University of Washington Press), 12–44.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar,: Theoretical Prerequisites*, Vol. 1. Stanford, CA: Stanford University Press.

Langacker, R. W. (1995). "Possession and possessive constructions," in *Trends in Linguistics: Language and the Cognitive Construal of the World*, eds J. R. Taylor and R. E. MacLaury (Berlin: Mouton de Gruyter), 51–80.

Macrae, A. (2016). You and I, past and present. *Diegesis* 5, 64–80.

Mar, R. (2018). Stories and the promotion of social cognition. *Curr. Direct. Psychol. Sci.* 27, 257–262. doi: 10.1177/0963721417749654

Mar, R., and Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspect. Psychol. Sci.* 3, 173–192. doi: 10.1111/j.1745-6924.2008.00073.x

Mundy, P., and Newell, L. (2007). Attention, joint attention, and social cognition. *Curr. Direct. Psychol. Sci.* 16, 269–274. doi: 10.1111/j.1467-8721.2007.00518.x

Neiger, M., and Tenenboim-Weinblatt, K. (2016). Understanding journalism through a nuanced deconstruction of temporal layers in news narratives. *J. Commun.* 66, 139–160. doi: 10.1111/jcom.12202

Oatley, K. (1999). Meeting of minds: dialogue, sympathy, and identification, in reading fiction. *Poetics* 26, 439–454. doi: 10.1016/s0304-422x(99)00011-x

Okada, S., Bono, M., Takanashi, K., Sumi, Y., and Nitta, K. (2013). "Context-based conversational and gesture classification in narrative interaction," in *Proceedings of the 15th ACM on International conference on multimodal interaction*, Yokohama, 303–310.

Papacharissi, Z., and de Fatima Oliveira, M. (2012). Affective news and networked publics: the rhythms of news storytelling on #Egypt. *J. Commun.* 62, 266–282.

Parrill, F. (2010). Viewpoint in speech-gesture integration: linguistic structure, discourse structure, and event structure. *Lang. Cogn. Process.* 25, 650–668. doi: 10.1080/01690960903424248

Polletta, F. (2009). *It Was Like a Fever: Storytelling in Protest and Politics*. Chicago, IL: The University of Chicago Press.

Rühlemann, C. (2013). *Narrative in English Conversation: A Corpus Analysis of Storytelling*. Cambridge: Cambridge University Press.

Sanders, J. (2010). Intertwined voices: journalists' modes of representing source information in journalistic subgenres. *Engl. Text Construct.* 3, 226–249. doi: 10.1075/etc.3.2.06san

Sanders, J. (2017). *We Hebben een Verhaal Nodig [We Need a Story*]. Inaugural lecture. Nijmegen: Radboud University. Available online at: https://repository.ubn.ru.nl/bitstream/handle/2066/178626/178626pub.pdf

Sanders, J., and Van Krieken, K. (2018). Exploring narrative structure and hero enactment in brand stories. *Front. Psychol.* 9:1645. doi: 10.3389/fpsyg.2018.01645

Sanders, J., and Van Krieken, K. (2019). Traveling through narrative time: how tense and temporal deixis guide the representation of time and viewpoint in news narratives. *Cogn. Ling.* 30, 281–304. doi: 10.1515/cog-2018-0041

Sangers, N. L., Evers-Vermeul, J., Sanders, T. J., and Hoeken, H. (2020). Vivid elements in Dutch educational texts. *Narrat. Inq.* 30, 185–209. doi: 10.1075/ni.18090.san

Shenhav, S. R. (2005). Concise narratives: a structural analysis of political discourse. *Discourse Stud.* 7, 315–335. doi: 10.1177/1461445605052189

Slater, M., and Rouner, D. (2002). Entertainment–education and elaboration likelihood: understanding the processing of narrative persuasion. *Commun. Theory* 12, 173–191. doi: 10.1093/ct/12.2.173

Sweetser, E., and Stec, K. (2016). "Maintaining multiple viewpoints with gaze," in *Viewpoint and the Fabric of Meaning*, eds B. Dancygier, W.-L. Lu, and A. Verhagen (Berlin: Mouton de Gruyter), 237–258.

Tomasello, M. (1995). "Joint attention as social cognition," in *Joint Attention: Its Origins and Role in Development*, eds C. Moore and P. J. Dunham (New York, NY: Psychology Press), 103–130.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Tomasello, M., Kruger, A. C., and Ratner, H. H. (1993). Cultural learning. *Behav. Brain Sci.* 16, 495–511.

Van Krieken, K. (2018). Ambiguous perspective in narrative discourse: effects of viewpoint markers and verb tense on readers' interpretation of represented perceptions. *Discourse Process.* 55, 771–786. doi: 10.1080/0163853x.2017.1381540

Van Krieken, K. (2019). "Literary, long-form or narrative journalism," in *The International Encyclopedia of Journalism Studies*, eds T. P. Vos and F. Hanusch (Hoboken, NJ: Wiley-Blackwell), 1–7. doi: 10.1002/9781118841570.iejs0135

Van Krieken, K., Hoeken, H., and Sanders, J. (2017). Evoking and measuring identification with narrative characters: a linguistic cues framework. *Front. Psychol.* 8:1190. doi: 10.3389/fpsyg.2017.01190

Van Krieken, K., and Sanders, J. (2017). Engaging doctors and depressed patients: effects of referential viewpoint and role similarity in health narratives. *Int. J. Communi.* 11, 4988–5006.

Van Krieken, K., Sanders, J., and Hoeken, H. (2015). Viewpoint representation in journalistic crime narratives: an analysis of grammatical roles and referential expressions. *J. Pragmat.* 88, 220–230. doi: 10.1016/j.pragma.2014.07.012

Van Krieken, K., Sanders, J., and Hoeken, H. (2016). "Blended viewpoints, mediated witnesses: a cognitive linguistic approach to news narratives," in *Viewpoint and the Fabric of Meaning: Form and Use of Viewpoint Tools Across Languages and Modalities*, eds B. Dancygier, W.-L. Lu, and A. Verhagen (Berlin: Mouton de Gruyter), 145–168.

Van Krieken, K., Sanders, J., and Sweetser, E. (2019). Linguistic and cognitive representation of time and viewpoint in narrative discourse. *Cogn. Ling.* 30, 243–252. doi: 10.1515/cog-2018-0107

Van Krieken, K., and Sanders, J. (forthcoming). "Linguistic perspective in written discourse," in *Creating a More Transparent Internet: The Perspective Web*, eds P. Vossen and A. Fokkens (Cambridge: Cambridge University Press).

Van Laer, T., De Ruyter, K., Visconti, L. M., and Wetzels, M. (2014). The extended transportation-imagery model: a meta-analysis of the antecedents and consequences of consumers' narrative transportation. *J. Consum. Res.* 40, 797–817. doi: 10.1086/673383

Van Ruler, B. (2005) *Wat is goede bedrijfsjournalistiek? [What is good corporate journalism?]* Den Haag: Logeion.

Verhagen, A. (2007). "Construal and perspectivization," in *The Oxford Handbook of Cognitive Linguistics*, eds D. Geeraerts and H. Cuyckens (Oxford: Oxford University Press), 48–81.

Wilkin, H. A., and Ball-Rokeach, S. (2006). Reaching at risk groups: the importance of health storytelling in Los Angeles Latino media. *Journalism* 7, 299–320. doi: 10.1177/1464884906065513

Zarantonello, L., Jedidi, K., and Schmitt, B. H. (2013). Functional and experiential routes to persuasion: an analysis of advertising in emerging versus developed markets. *Int. J. Res. Market.* 30, 46–56. doi: 10.1016/j.ijresmar.2012.09.001

Zwaan, R. A. (2004). "The immersed experiencer: toward an embodied theory of language comprehension," in *The Psychology of Learning and Motivation*, ed. B. H. Ross (Amsterdam: Elsevier), 35–62. doi: 10.1016/s0079-7421(03)44002-4

Zwaan, R. A., Langston, M. C., and Graesser, A. C. (1995). 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

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 van Krieken and Sanders. 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.

APPENDIX

Text from Dutch corporate magazine Brand & Brandweer [Fire & Fire Brigade], in particular a column in the series Onder de helm [Under the Helmet], entitled Ik leef met jullie mee, veel sterkte! [I Sympathize with You Guys, Lots of Luck]. The text was written by Jildou Visser and published in October 2017.

Original Dutch version:

P1 letsel ijssalon Intermezzo Heuvelring Tilburg. Voor die melding wordt eerste bevelvoerder Klaas Wim Jansen op 22 juli gealarmeerd. 'We rukten uit naar de kermis, maar ik had nog geen idee wat er aan de hand was.' Als hij aanrijdend hoort dat een kind bekneld ligt onder een kermisattractie, wordt het stil in de TS. Jansen: 'Ow shit, was het eerste dat door mijn hoofd schoot. "Ik leef met jullie mee, veel sterkte jongens", zei de centralist nog. Dat zegt een centralist nooit. Achteraf bleek dat hij beelden van een bewakingscamera in de buurt had gezien.'

De brandweerlieden worden bij de kermis opgewacht door een verkeersregelaar. 'Aan de blik op het gezicht van de verkeersregelaar en de manier waarop hij ons zwaaiend de weg wees, zag ik dat het een ernstig incident was. Vervolgens ging het zo snel. Vijftig meter verder waren we er al. Er waren veel mensen, schreeuwend, in paniek.' In eerste instantie ziet Jansen het kind niet. Door omstanders wordt hij erop gewezen dat het slachtoffertje op ongeveer twee meter hoogte ligt, bekneld onder een karretje van de kermisattractie. 'Dat zag er niet goed uit.' De attractie voor kleine kinderen bestaat uit aaneengeschakelde karretjes die langzaam omhoog en omlaag rijden. 'Het jongetje was uit een karretje gevallen en eronder terechtgekomen. Omstanders tilden het wagentje al op, zodat het niet meer met het volle gewicht op zijn buik lag. Hij was bij kennis en erg rustig. Hij heeft tijdens de hele inzet niet één keer geschreeuwd, maar die blik in zijn ogen zei alles. Zijn ogen schreeuwden om hulp. Af en toe sloot hij z'n ogen. Als we tegen hem begonnen te praten, deed hij z'n ogen weer even open en dan keek hij ons aan met die blik.'

Op dat moment weten de brandweerlieden nog niet hoe het jongetje vastzit. De bemanning van de eerste TS stabiliseert het karretje, haalt het daarna los en verwijdert het. Pas dan zien de manschappen dat de arm van de jongen klem zit door de sleuf met het mechanisme waarmee de karretjes werden voortbewogen. 'Toen we de kappen aan de onderkant van de attractie weghaalden, konden we pas echt goed zien hoe het armpje vastzat. Dit komt nooit meer goed, dacht ik. Het zag er erg ingewikkeld uit', aldus Jansen. Samen met de bevelvoerders van de tweede TS en het HV besluit hij dat ze de reling met de verlichting doorknippen om beter zicht te krijgen. Op dat moment keert een omstander zich tegen de brandweer. 'Een vrouw liep naar een van mijn manschappen toe toen hij de reling door wilde knippen. Hij keek mij aan. We wisten beiden niet wie het was. Het had een bezorgde ouder kunnen zijn, maar toen ze aan mijn manschap begon te trekken, wisten we genoeg. De politie greep gelukkig snel in', vertelt Jansen. 'Dan ontstaat even verderop een opstootje als een tweede persoon zich tegen de politie keert. Je bent even heel verbaasd, omdat je probeert een kind te redden en dan zo wordt tegengewerkt. Snel daarna ging de focus weer volledig op de inzet en op het kind terwijl de politie het vechtpartijtje verderop afhandelde.'

Nadat de reling kapot is geknipt, buigen de brandweerlieden de stalen plaat om zodat ze beter kunnen zien hoe de constructie in elkaar zit. Jansen: 'Daarna hebben we stukje voor stukje in de attractie geknipt om het mechanisme waar de jongen in vastzat los te maken. Toen van bovenaf naar beneden werd geroepen dat de arm los was, was de opluchting groot. Dan slaak je wel even een zucht.' Enkele manschappen helpen de ambulancemedewerkers om de jongen naar beneden te tillen. 'Daarna hebben we op de kermis direct een eerste nagesprek gehouden. Het was voor ons allemaal een flinke klus, dan wil je je verhaal wel even kwijt.' De eerste twee TS'en en de HV gaan buiten dienst en alle eenheden keren daarna terug naar de kazerne voor de evaluatie. Daar doet iedereen nogmaals zijn verhaal. Na de evaluatie besluit de ploeg weer in dienst te gaan. 'We hebben die dag nog een paar meldingen gehad. Dat was goed. Het is fijn om wat afleiding te hebben.' Na zijn dienst keert Jansen naar huis. 'Thuis heb ik mijn verhaal nog een keer gedaan. Ik praat normaal nooit over inzetten, maar deze inzet was bijzonder en vergeet ik nooit meer. Het is fijn als er dan thuis een luisterend oor is.'

English translation:

P1 injury Intermezzo ice cream parlor, Heuvelring in Tilburg. Commanding officer Klaas Wim Jansen was the first to receive this report on 22 July. "We responded and drove to the fairground, without me having any idea of what had happened." When he hears en route that a child is trapped under a fairground attraction, it goes quiet in the fire engine. Jansen: "'Oh shit' was my first thought. 'I sympathize with you guys, lots of luck', added the operator. Operators never say that. Afterwards, it turned out he'd seen images from a surveillance camera in the vicinity."

The fire fighters are met at the fairground by a traffic controller. "I could see from the look on the traffic controller's face, and the way he waved us through, that the incident was a serious one. After that, everything went so quickly. Fifty meters further, and we'd arrived. There were lots of people, screaming in panic." At first, Jansen doesn't see the child. He is pointed out by bystanders that the victim is about two meters high, trapped under a trolley of the fairground attraction. "It didn't look good." The ride, for small children, consists of a row of cars, linked together, that ride slowly upward and downward. "The little boy had fallen out of a ride car and landed underneath it. Bystanders were already lifting the ride car, so the full weight of it wasn't on his stomach any more. He was conscious and very calm. He hasn't screamed once during the whole operation, but the look in his eyes spoke volumes. His eyes were screaming for help. Now and then he closed his eyes. When we started talking to him, he briefly opened his eyes again and looked at us with that expression."

At that point, the fire fighters don't know how exactly the little boy is trapped. The crew of the first TS truck stabilizes the ride car, then disconnects and removes it. Only then do they see that the little boy's arm is jammed in the slit of the mechanism that moved the cars forward. "It wasn't until we removed the

covers from underneath the ride, that we could see properly how his little arm was stuck. This is never going to work, I thought. It looked really complicated", according to Jansen. He and the commanding officers of the second fire truck and the other emergency vehicle decide to start by cutting through the railing with the lighting to get a better view. At that point, a bystander turns against the fire fighters. "A woman walked toward one of my crew just as he was about to cut through the railing. He looked at me. Neither of us knew who she was. She might have been a concerned parent, but when she started pulling at my crew member, it was clear. Luckily, the police intervened quickly", Jansen tells. "Then there's a commotion a bit further away, when a second person turns on the police. You're just taken aback for a moment, because you're trying to save a child and people are turning on you like that. Soon after that, all attention went back to the operation and the child, while the police dealt with the fight further along."

Once the railing is cut, the fire fighters bend the steel plate, so they have a better view of the construction of the ride. Jansen: "After that, we have cut through the ride, a bit at a time, to free the mechanism the boy was stuck in. When it was shouted downward from above that the arm was loose, the relief was great. You really heave a sigh then." A few crew members help the ambulance staff lift the boy down. "Afterward, we immediately had an initial meeting about the operation, there at the fairground. It was a big operation for all of us, so you want to talk about it." The first two TS trucks and the HV go off duty, and all units return to the fire station for the evaluation. Everyone tells their story there again. Following the evaluation, the crews decide to go back on duty. "We have been called out a few more times that day. That was fine. It's good to have some distraction." After his shift, Jansen returns home. "At home, I have told my story again. I don't usually talk about operations, but this one was special, and I'll never forget it. It's nice when there's someone at home to talk to then."

frontiers in Psychology

ORIGINAL RESEARCH
published: 31 March 2021
doi: 10.3389/fpsyg.2021.613357

Extended Perspective Shift and Discourse Economy in Language Processing

*Jesse A. Harris**

Department of Linguistics, University of California, Los Angeles, Los Angeles, CA, United States

Research spanning linguistics, psychology, and philosophy suggests that speakers and hearers are finely attuned to perspectives and viewpoints that are not their own, even though perspectival information is not encoded directly in the morphosyntax of languages like English. While some terms seem to require a perspective or a judge for interpretation (e.g., epithets, evaluative adjectives, locational PPs, etc.), perspective may also be determined on the basis of subtle information spanning multiple sentences, especially in vivid styles of narrative reporting. In this paper, I develop an account of the cues that are involved in evaluating and maintaining non-speaker perspectives, and present an economy-based discourse processing model of perspective that embodies two core principles. First, perspectives are subject to a "speaker-default," but may shift to a non-speaker perspective if sufficient contextual cues are provided. Second, the processor follows the path of least resistance to maintaining perspective, opting to maintain the current perspective across sentences as long as the shifted perspective continues to be coherent. The predictions of the model are tested in a series of offline and online studies, manipulating the form of an attitude report and the tense of the sentence that follows. Implications for processing perspective and viewpoint in speech and narrative forms are explored.

Keywords: perspective taking, free indirect discourse, tense, discourse processing, language processing

OPEN ACCESS

Edited by:
Emar Maier,
University of Groningen, Netherlands

Reviewed by:
Thomas Weskott,
University of Göttingen, Germany
Marta Abrusan,
École Normale Supérieure, France

***Correspondence:**
Jesse A. Harris
jharris@humnet.ucla.edu

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

Received: *02 October 2020*
Accepted: *11 January 2021*
Published: *31 March 2021*

Citation:
Harris JA (2021) Extended Perspective Shift and Discourse Economy in Language Processing. Front. Psychol. 12:613357. doi: 10.3389/fpsyg.2021.613357

1. INTRODUCTION

Language users are clearly sensitive to perspectives other than their own. And yet, there appears to be no uniform or determinate way to convey point of view, even though many lexical items, e.g., *beautiful* or *nearby*, seem to require that a viewpoint be assessed for interpretation (e.g., Mitchell, 1986; Partee, 1989; Lasersohn, 2005; Stephenson, 2007, among many others in linguistics and philosophy). Perspective has been described in various ways: as an "origo" (Bühler, 1934), an "empathic identification" (Kuno and Kaburaki, 1977), a "judge" (Lasersohn, 2005), an "evaluator" (Patel-Grosz, 2012), or, as we do here, a "perspectival center" (Harris, 2012; Patel-Grosz, 2012; Hinterwimmer, 2017). Still, little is known about how a perspectival center is calculated by the human language processing system, and we are, at present, left with a wealth of open foundational questions: Given the myriad of potentially relevant information sources that signal perspective, how does the processor determine which perspective is at play? What types of cues signal a perspective shift? How do these cues interact? Is perspective shift costly to maintain over longer stretches of discourse or text?

Most experimental research takes as its starting point the assumption that, in non-narrative genres, a presumptive pragmatic default favors the speaker's perspective, which then interacts with presentational, surface cues to shift to a non-speaker center (Smith, 2003, 2009; Harris, 2012), even if the mechanisms motivating these shifts differ dramatically. Many lexical and contextual cues have been explored experimentally: epithets (Harris and Potts, 2009, 2011; Kaiser, 2015) or other expressives, like *damn* (Frazier et al., 2015), verbs of saying (Harris and Potts, 2009), evaluative adjectives (Kaiser and Lee, 2018; Kaiser, to appear), as well as subtle prosodic and non-verbal modulations (Harris and Potts, 2011).

This project turns to how perspective shift is achieved in vivid narration style, in which the perspectival center shifts to a third-person character. Utilizing the perspective shifting tendencies of narrative parenthetical reports, as in *It was a time of utter chaos, lamented Mary*, this paper explores how the tense of a following continuation influences whether the perspective stays shifted with the third-person or reverts back to the speaker. To this end, I present a series of offline and online experiments that manipulate the report type of an utterance and the tense of the sentence that follows. The results confirm that, compared to standard indirect styles of reporting, narrative parenthetical reports provide a strong cue for an alternate perspectival center (Reinhart, 1975, 1983). Further, a shifted perspective is more likely to extend to the following sentence when it is presented in Present tense, placing the reported event within a single, continuous "contextual now." I argue that the results support an economy-based model of perspective processing, in which the discourse processor discourages perspective shifting generally, but maintains a shifted perspective when possible for continuity.

2. PERSPECTIVE AND PERSPECTIVE SHIFT

Intuitively, a PERSPECTIVE relates to a particular attitude, opinion, or belief regarding an individual or situation. It is roughly synonymous with *viewpoint* and *point of view*. In this sense, a perspective then is just a body of perceptions, commitments, and information, broadly construed, that hold for an attitude holder in a particular situation. Following Harris and Potts (2009), a constituent or clause C is said to be SPEAKER-ORIENTED if the speaker expresses a commitment to the content of C in using it; otherwise, C is said to be NON-SPEAKER-ORIENTED, and expresses, perhaps by proxy of the speaker, the commitments and perspectives of another individual in the discourse.[1] I will use the term PERSPECTIVAL CENTER as a shorthand for how the views and commitments of a relevant attitude holder are represented within the discourse. The term is intended to be compatible with multiple formalizations of linguistic subjectivity.

Some lexical items seem to require a perspectival center for interpretation. For example, certain locational predicates like *local* or *nearby* is evaluated with respect to a particular location (Mitchell, 1986; Partee, 1989; Recanati, 2011). Expressives and other evaluative terms, like predicates of personal taste, likewise convey a particular viewpoint (e.g., Lasersohn, 2005; Potts's, 2005, among many others). By using an epithet like *that jerk Bill*, the speaker conveys that she (or another attitude holder) stands in a particular, usually negative, emotional relationship with the referent, Bill. Similarly, the use of *tasty* in *that dish is tasty* indicates that an attitude holder takes a positive gustatory stance with respect to the dish under discussion.

Since expressives typically convey the experience or opinion of an individual,[2] they are sometimes said to directly update the context with the relevant attitude (Potts, 2007; Barker et al., 2011). Expressive content is thus notoriously challenging to deny. In Potts's (2005) examples (1-2), an epithet like *that bastard* communicates a certain degree of speaker commitment to the attitude conveyed by the epithet. To then immediately renounce or deny that commitment is strikingly incoherent (1). Expressive judgments also seem impervious to standard forms of denial. While it's easy enough for Speaker A to assert another opinion (2a), a direct denial of Speaker A's attitudinal stance (2b) is inappropriate (unless perhaps to convey the implication that A's attitude is insincere or fleeting). Nonetheless, the veridicality of A's attitude itself can be called into question, even if it has not been denied directly (2c).

(1) That bastard Conner was promoted. #But probably he's not a bastard.

(2) *Speaker A:* That bastard Connor was promoted.

Speaker B:

a. I don't think Connor is a bastard.
b. # You don't think Connor is a bastard.
c. You don't really think Connor is a bastard, do you?

As predicates of personal taste convey personal or subjective experience, they are subject to what has sometimes been called "faultless disagreement," in which opposing opinions regarding personal taste might each be true (e.g., Kölbel, 2004). Intuitively, Abe's claim in (3) is false if it generically extends to others, including Betty, but true enough if it is merely limited to his own preferences.

(3) a. Abe: Cilantro is tasty.
b. Betty: Ugh, no it isn't. It tastes like soap.

The above properties have fueled the claim that expressives and related terms are inherently speaker-oriented (Cruse, 1986; Kaplan, 1999; Potts's, 2005). An intuitive explanation is that one's feelings and emotions are one's own, and that other discourse

[1]For present purposes, I take a commitment to represent "a kind of *persona* of beliefs" (Hamblin, 1970). An individual *X*'s commitment is inferred from the content of a particular expression *E*, and licenses the inference that *X* is committed to the presuppositions associated with *E* (Peters, 2016). Commitments are distinct from *genuine beliefs*, as they may be provisional, disingenuous, or adopted the purposes of the discourse (Hamblin, 1970, 1971). Commitments may also be shared or dependent on another source (e.g., Gunlogson, 2008; Harris, 2014; Malamud and Stephenson, 2014).

[2]The use of the term "individual" may not be entirely accurate, in that predicates of personal taste have been argued to display elements of normativity (Buetow, 2014) or genericity (Moltmann, 2010; Pearson, 2013).

participants simply lack the appropriate first-hand experience to warrant a denial (e.g., Willer and Kennedy, to appear).

And yet, many cases of non-speaker perspective shift have been documented in the literature. Following the intuitions of Amaral et al. (2007), Wang et al. (2005), and others, Harris and Potts (2009) showed experimentally that such terms could be understood as reflecting a non-speaker perspective, provided there were sufficient clues in the context to warrant a shift in perspective (see also Harris, 2009, 2012; Kaiser, 2015). Their results found that changing the valence of a single word (e.g., *high* to *low*) in context sentences like (4) modulated whether the rate at which subjects attributed an epithet like *the bastard* in the target sentence (5) to the attitude holder (Sheila) or the speaker. In (4A), giving someone a high grade is unlikely to result in considering the professor a jerk, and so the non-speaker interpretation remains unlikely. However, receiving a low grade would plausibly result in a negative assessment, increased non-speaker interpretations of the target (4B).

(4) *Contexts:*
   A. My classmate Sheila said that her history professor gave her a high grade.
   B. My classmate Sheila said that her history professor gave her a low grade.

(5) *Target:* The jerk always favors long papers.

Harris and Potts (2009) proposed that such terms normally default to a speaker-oriented interpretation because shifts in perspective constitute a risky communicative strategy (see also Lasersohn, 2007; Potts, 2007). Such terms do not merely describe a situation, but also evaluate it against a particular perspective or point of view. Without a morphosyntactic or highly conventionalized means to convey non-speaker commitment, speakers run the risk of the content "leaking" from a reportive context. Thus, speakers tend to either avoid non-speaker oriented uses of expressives or limit such cases to rich contexts where their intentions are likely to be recoverable.

Predicates of personal taste can also be understood from a non-speaker perspective under similar circumstances. Stephenson (2007) provided examples in which *tasty* is understood as non-speaker oriented if another perspective is highly salient (6) or embedded under an epistemic modal (7).

(6) a. Mary: How's that new brand of cat food you bought? (Attributed to Kai von Fintel)
   b. Sam: I think it's tasty, because the cat has eaten a lot of it.

(7) The cat food might be tasty. (Attributed to Danny Fox)

In questions, predicates of personal taste are readily associated with the addressee, participating in "interrogative flip" (e.g., Speas and Tenny, 2003; Tenny, 2006). The examples below are striking in that the speaker's dissenting judgment is clearly known (8) or can be easily inferred (9).

(8) *Betty hates cilantro, but knows Abe loves it. She points to his order of Cilantro Lime Shrimp and asks:* Is your dish tasty?

(9) *To a hungry cat:* How about some tasty cat food?

In these cases, it's clear that the speakers are not interested in whether or not they would enjoy the item; the query is entirely oriented toward the perspective and personal tastes of the addressee. Other uses are possible, of course. For example, uttered in a context where Betty does likes cilantro but is unsure about shrimp, (8) could be easily understood as a query into the dish's quality, information that she could use to determine whether she might like it herself.

In an experimental study, Kaiser (to appear) found that the perspective associated with a predicate of personal taste (*disgusting*) was influenced by whether a character was presented as having a direct sensory experience. Experiencer predicates with olfactory (10b), gustatory (10c), and visual sensation (10c) were compared against a non-experiencer baseline (10a).

(10) a. When I came into the room, Eliza put the muffin on the platter. It was disgusting.
   b. When I came into the room, Eliza smelled the muffin on the platter. It smelled disgusting.
   c. When I came into the room, Eliza tasted the muffin on the platter. It tasted disgusting.
   d. When I came into the room, Eliza saw the muffin on the platter. It looked disgusting.

Subjects selected forced-choice answers to questions like *Whose opinion is it that the muffin looked/smelled/tasted/was disgusting?* to indicate whose experience (*the narrator's* or *Eliza's*) was represented by the subjective adjective (*disgusting*). While all experiencer predicates increased the rate of perspective shift, predicates describing taste and smell experiences tended to shift perspective more often (∼80–85%) than visual experience (∼50%), compared to the baseline (<25%). Presumably, visual experience can, in a sense, be shared between the speaker and another agent, making it less clearly subjective. The results point to the role that experience and evidential warrant play in shifting the perspectival center away from the speaker default. The result aligns with Hinterwimmer (2019) account, in which experiencer predicates generate particularly salient events of the represented experience, thereby providing highly salient anchors for perspective shift.

Even if the evaluation of a term tends to default to the speaker's point of view, shifts in perspective are ubiquitous in text and speech. Single sentence examples can obscure these intuitions, but extended perspective shifts are readily found in many forms of narrative, where events are often presented within a character's perspective. For example, narratives often convey a text through the lens, as it were, of a character telling the story from her experience (see for example Banfield, 1982, Fludernik, 1993, or Emmott, 1999 for review).

The following passage (11) from Henning Mankell's (2009) *The Troubled Man* illustrates the variety of perspectives that may come into play in narrative discourse. The sentences are segmented using the numbered | symbol, marking plausibly non-speaker-oriented passages with an asterisk ∗ superscript.

(11) $|_1$ The meeting with Atkins had unsettled him. $|_2$ He searched for a Berlin wall inside himself, but failed to find one. $|^*_3$ Was his life really so restricted that major events taking place in the outside world never had much effect on him? $|^*_4$ What aspects of life had upset him? $|^*_5$ Pictures of children who had been badly treated, of course – but he had never been sufficiently moved to do anything about it. $|^*_6$ His excuse was always that he was too busy with work. $|^*_7$ I sometimes manage to help people by making sure that criminals are removed from the streets, he thought. $|^*_8$ But aside from that? $|_9$ He gazed out over the fields where nothing was yet growing, but he failed to find what he was looking for. (Mankell's, 2009, 85)

In sentences 3–8, the text contains a variety forms, pronouns, tenses, and non-standard reporting styles, giving rise to the inference that the perspective has shifted from the narrator to the main character's internal thoughts and experience. Although many authors have identified the same kinds of perspective shifting devices in non-reportive narrative styles (e.g., Banfield, 1982; Fludernik, 1993; Smith, 2003), the cues remain somewhat heterogeneous.

The remainder of the paper is organized as follows. A novel taxonomy of perspective shifting cues is proposed in section 2.1, where two natural classes are introduced: cues involved in *evaluating* a perspective, and cues responsible for *maintaining* a shifted perspective. The key features of an economy-based discourse processing model are presented in section 2.2. Then, sections 3–5 present the results of three experimental studies probing the prediction of this model. Finally, section 6 concludes.

2.1. Perspective Taking

This paper centers on two theoretical issues central to perspective taking in language. The first addresses what features of a context *initiate* perspective shift. While this is largely an empirical question, any term which is perspectively rich or conventionally associated with a non-speaker perspective could in principle be a candidate. The second issue addresses what factors *maintain* a perspective shift, once shifted. On the assumption that perspective shift is possible, albeit difficult, to maintain across larger stretches of discourse, it is reasonable to expect that a non-speaker perspective is facilitated by cues that reference the speech, thought, or attitudes of the perspective to which the context has shifted.

Prior literature has identified a host of seemingly heterogenous forms contributing to linguistic subjectivity, exemplified by the partial list in (12) from Smith (2003, p. 176):

(12) Communication verbs
Deictic adverbials: place, time
Direction and location PPs
Epithets
Evaluative verbs and adverbs
Evidential adverbials, adjectives, and verbs
Experiencer predicates
Possessive pronouns
Progressive viewpoint
Propositional complements
Situation type: State and activity sentences

As discussed below, other indicators of perspective, such as certain tense-aspect combinations, interjections, fragments, and questions, may be added to the list.

In the following sections, I propose that perspectival cues may be roughly subdivided into at least two natural classes of cues to perspective by their primary role of establishing (TYPE A) or maintaining (TYPE B) a non-speaker perspective. A basic and incomplete sketch of the proposed typology is provided below, summarized in **Figure 1**. I argue that such cues not only work together, conspiring to strengthen the chance of a perspective shift, but also have distinct roles in interpretation (see also Wiebe, 1994, for the idea that cues to perspective differ in strength). The two types are briefly sketched below. Special attention is given to narrative parenthetical reports and tense, factors that were manipulated in the experiments presented below.

2.1.1. Type A: Evaluating Perspective

The Type A class consists of elements which are semantically or conventional associated with a perspectival center for interpretation. This type may be further divided into two subcategories: PRESENTATIONAL and partially or QUASI-CONVENTIONAL cues. Type A presentational cues contain an evaluative or representational component for which a judge or perspective is required for interpretation, as in epithets and predicates of personal taste. Additional examples plausibly include communication verbs, deictic adverbs, and evaluative verbs and adverbs like *amazingly*. Examples of Type A quasi-conventional cues consist of cases of root clause phenomena (e.g., Emonds, 1970), including narrative parenthetical reports (13a), exclamative particles (13b), inversion (13c), and questions (13d), among others.[3]

(13) a. All she wanted was a little happiness, thought Mary.
b. God, what a mess today was turning out to be.
c. Never had she been so worried in all her life.
d. Why did so many bad things have to happen to her?

These cues have been studied extensively in the literature on narrative discourse, particularly in works dealing with the phenomena of represented speech and consciousness (e.g., Banfield, 1982; Fludernik, 1993). It is easy to read (13) as a passage conveying the perspective of the parenthetical subject, Mary, rather than the viewpoint of a narrator or speaker. Indeed, Parenthetical reports (PRs) like (13a) are often used to illustrate narrative shifts in perspective. We briefly digress to describe how PRs differ from Standard reports (SRs), and why the former have been implicated in perspective shift.

Although early accounts (Emonds, 1970; Ross, 1970) attempted to derive PRs from SRs, there is strong evidence that they are distinguished both syntactically and interpretively. As Banfield (1982) observed, PRs and SRs have syntactically

[3]Moore (2011) details how similar devices marked direct speech in early English before the use of quotation marks was standardized. The convergence of these forms is unlikely to be accidental (Fleischman, 1990).

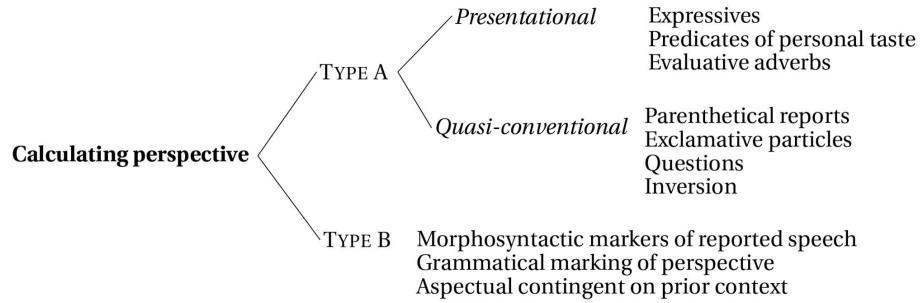


FIGURE 1 | Typology of cues used to calculate perspective. Cues are intended to be representative and not exhaustive.

distinct distributions. For example, PRs generally do not allow complementizers (14), but do permit speaker-oriented adverbs (15), imperatives (16), and direct questions (17) in the reported content.

(14) a. A week ago, John thought that he would marry Ann tomorrow.
b. * That he would marry Ann tomorrow, John thought a week ago.

(15) a. * John said (that) oh he was so tired.
b. Oh, he was so tired, John said.

(16) a. Go home! (She shouted.)
b. # She shouted go home!

(17) a. # She$_i$ wondered should she$_i$ go home?
b. Should she$_i$ go home before her$_i$ husband returned? (Mary$_i$ wondered.)

Regarding interpretation, Reinhart (1983) noted that PRs lack an ambiguity that SRs exhibit. The SR in (18) may be interpreted as either a transparent belief, in which the term "his mother" is interpreted *de re*, and is not attributed to Oedipus, or an opaque belief, in which the contents of the belief are interpreted *de dicto*.

(18) *Standard Report:* Oedipus believed (that) his mother wasn't his mother.
a. *Transparent:* Oedipus believed that Jocasta wasn't his mother.
b. *Opaque*: Oedipus believed the contradictory thought that his mother wasn't his mother.

However, PRs lack the transparent reading, and can be interpreted only in the opaque, and in this case a contradictory, *de dicto* sense.

(19) *Parenthetical Report:* His mother wasn't his mother, Oedipus believed.
a. *Transparent unavailable:* ~~Oedipus believed that Jocasta wasn't his mother.~~
b. *Opaque*: Oedipus believed the contradictory thought that his mother wasn't his mother.

A second, and related, piece of evidence is that a PR represents the parenthetical subject's words or inner speech more directly. Thus, interjections of an opposing perspective from another attitude holder are perceived as infelicitous. Oedipus' thought in (20) may be conveyed as a PR (20a), but only if the report is consistent with Oedipus' actual thought. Interjecting that thought with a speaker-oriented appositive like "his mother in reality" in (20b) is not part of the content of the report, and renders the sentence semantically incoherent.

(20) Oedipus thinks: "Jocasta is married to Laius"
a. Jocasta was married to Laius, Oedipus thought.
b. # Jocasta, his mother in reality, was married to Laius,
Speaker perspective
Oedipus thought.

In her seminal analysis of free indirect discourse (FID), Banfield (1982) provided the category of "represented speech and thought," distinct from directly quoted (DQ) or reported speech. These reports are syntactically realized by an Expressive node consisting of SELF and NOW values, which represent the attitude holder and moment of consciousness, respectively. These values favor the speaker and present, but may be shifted to take on values corresponding to another attitude holder. Many subsequent analyses proposed that expressions are semantically evaluated against two potentially different contexts: one anchored to the utterance situation, the other to the thought or perception situation (e.g., Schlenker, 2004; Sharvit, 2008; Stokke, 2013; Eckardt, 2014; Hinterwimmer, 2017; Abrusán, 2020, among others). Although these accounts differ a good deal in implementation, the two contexts diverge in cases of free indirect discourse and other kinds of perspective shift. This view nicely captures the fact some context-dependent expressions are sensitive to different contexts in FID. While tense and first person pronouns are understood from the speaker's perspective, as in direct discourse, third person pronouns and adverbials optionally reflect the perspective of the represented attitude holder. Crucially, the past tense in FID is understood as past to the speaker, while nonetheless representing the character's contextual here and now (Banfield, 1982; Fludernik, 1993). Example (21) from Sharvit (2008) illustrates the properties that FID shares with DQ

(underlined) and the properties it shares with standard indirect reports (**in bold**).

(21) Examples from Sharvit (2008)

a. As he looked at my picture, John thought: "Yes, I want to marry her today." (DQ)

b. As he looked at my picture, John thought that **he** want**ed** to marry **me** that day. (SR)

c. John looked at my picture. Yes(, he thought,) **he** wanted to marry **me** today. (FID)

Interestingly, the speech or thought that is represented in FID reports shares some features with DQ, in that various adverbs, like the speaker-oriented adverb *yes* or the temporal adverb *today* in (21a,c), portray the point of view of the reported subject, rather than the speaker, or narrator in the text. As such, it may be tempting to treat FID as a special case of DQ. However, there are important differences between FID and DQ, as well: in FID, a third person pronoun may refer to the subject of the reporting clause, which is *inconsistent* with DQ, but consistent with Standard reports (21b). As noted by Sharvit (2008), there is some disagreement about how first and second person pronouns pattern in FID, but at least in the case above, the first person pronoun appears to refer to the speaker, rather than the attitude holder. Consequently, the extent to which such reports can be treated as an instance of mixed quotation (e.g., Maier, 2015, 2017) remains a matter of some debate (Reboul et al., 2016, for review).

It is important to note that Reinhart (1983) distinguished between two subtypes of parenthetical report: the Narrative parenthetical and the Discourse parenthetical. Only Narrative parentheticals unambiguously represent the point of view of the parenthetical subject. The two types can also be distinguished syntactically. For example, Narrative PRs can postpose the subject to a position after the verb (*said John*), but Discourse PRs cannot. Narrative PRs allow backwards anaphora (22), but do not permit negation (23), in contrast with Discourse PRs.

(22) *Backwards anaphora*

a. He$_i$ has never seen a thing like that, said John$_i$. (Narrative PR)

b. * John$_i$ has never seen a thing like that, said he$_i$. (Discourse PR)

(23) *Negation*

a. * She was absolutely wrong, it did not seem to her. (Narrative PR)

b. You couldn't loan me five dollars, I don't suppose. (Discourse PR)

To ensure that the PRs are truly associated with perspective shift in the experiments below, only cases with postposed subjects were used, forcing a Narrative PR reading.

The discussion so far has concentrated on the cues that signal a shift in perspective. However, shifts in perspective may *extend* across several sentences in vivid styles of reporting like FID. The next sections turns to the linguistic elements that are involved in maintaining a non-speaker perspective.

2.1.2. Type B: Maintaining Perspective

In some languages, certain aspects of perspective appear to be encoded morphosyntactically in lexical or grammatical forms. In Japanese, for example, there is a well-known lexical contrast between verbs of giving (24). Although both verbs lexically describe a giving relation, they do so from a different perspective. The verb *yatta* is said to be appropriate in situations described from the point of view of the subject (Taroo), whereas *kureta* is used to describe the event from the point of view of the dative object (Hanako); see Kuno and Kaburaki (1977) among others for discussion.

(24) a. Taroo wa Hanako ni okane o *yatta*
Taroo Hanako to money gave
'Taroo gave money to Hanako'

b. Taroo wa Hanako ni okane o *kureta*
Taroo Hanako to money gave
'Taroo gave money to Hanako'

Another example of grammatically encoded perspective comes from German, which employs a dedicated marker of reportive mood, sometimes called *Konjunktiv I* (e.g., Zifonun et al., 1997; Sode, 2014) or the *reportive subjunctive* (Fabricius-Hansen and Sæbø, 2004). It conveys that the speaker is not committed to the content of the clause, and is licensed when subordinate to a verb of saying like *behauptete* (claimed), or for extended narrative discourse, in which a non-speaker viewpoint is reported.

(25) Er behauptete, dass jemand das Auto angefahren
He claimed that somebody the car on-driven
habe, …
have$_{PresSub}$
'He claimed that somebody had driven into the car, …'

Although English lacks morphological indicators of a particular perspective, certain tense-aspect combinations are sometimes counted among those cues that signal non-speaker perspective (e.g., Smith, 2003). Such combinations will be designated TYPE B cues. Whereas Caenepeel and Sandström (1992) argue that the past progressive introduces a perspective, I'll assume that tense-aspect configurations do not so much initiate a perspective shift as *maintain* a shifted perspective already active in the previous context. As such, nothing special needs to be said about such cases, as they simply follow from what we already know about how tense and aspect interact with the linguistic context. In this way, Type B cues are elements that are responsible for maintaining a perspective, once shifted. These cases will be referred to as EXTENDED SHIFTS of perspective.

A simplified version of a relatively standard three parameter account of tense and aspect is adopted for illustration. Only enough to follow the logic of the experimental design is presented, and many important developments and complications are omitted. Following Reichenbach (1947), Klein (1994), and many others, *tense* encodes the relationship between two temporal intervals: the TU (the time of utterance) and the TTOP (a contextually determined topic time). Ignoring the future tense for simplicity, the present tense is used when TU and TTOP are

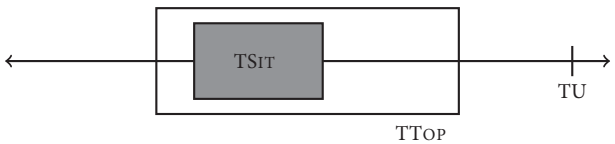
simultaneous or overlap (TU = TTOP), whereas the past tense is used when TTOP precedes TU (TTOP < TU). *Aspect* encodes the relationship between TTOP and TSIT (the time that the event described by the clause occurred). In perfective aspect forms, for example, TTOP is said to contain TSIT (TSIT ⊂ TTOP).[4] The basic ingredients of the system are illustrated with the Past and Present perfect sentences in (26).

(26) *Past and present perfect*

a. Mary had gone to bed.

i. *Past tense*: The topic time of the sentence precedes the time of utterance (TTOP < TU).

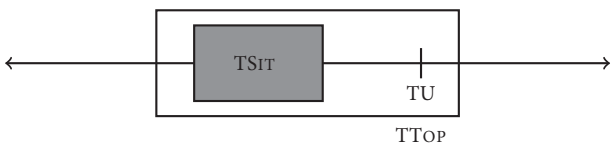
ii. *Perfect aspect*: The time of going to bed is contained by the topic time (TSIT ⊂ TTOP).



b. Mary has gone to bed.

i. *Present tense*: The topic time of the sentence coincides with the time of utterance (TTOP = TU).

ii. *Perfect aspect*: The time of going to bed is contained by the topic time (TSIT ⊂ TTOP).



In both cases, the event of going to bed had already been initiated prior to the end of TTOP. The only semantic difference is whether TU and TTOP overlap. With a few key exceptions relevant to narration, TU is anchored to the time of the present experience. However, the temporal location of TTOP depends on contextual information, and can be anchored to events in the narrative sequence. Klein (1994) adopted the view that sentences make reference to a topic situation (Kratzer, 2008), and analogized TTOP to the time span associated with it. The TTOP can be provided by explicit linguistic context, such as a preceding sentence or a frame or temporal adverbial, but is usually left implicit (Partee, 1973). In the Present perfect, TTOP may anchor to TU, though other reference points are possible. However, in the Past, sentences like (26a) seem incomplete without a contextually relevant topic time, e.g., *Sally came home at midnight. John had gone to bed.*

Tense interacts with perspective in intriguing ways. Intuitively, the interpretation of a sentence following a perspective shift depends on the tense and aspect of the continuation. To take the second sentences in (27) as an example, while the Past perfect version (27a) can be interpreted either as speaker or non-speaker oriented, the Present perfect counterpart (27b) is more readily understood as a representing a non-speaker experience.

(27) a. Mary wasn't feeling well, thought John. She *had* fallen asleep on the couch again.

b. Mary wasn't feeling well, thought John. She *has* fallen asleep on the couch again.

An extended shift in these cases may be achieved by somewhat different means. Shifting in the Past perfect case plausibly represents an instance of classic FID, in which the perspective of the main character is cast as anterior to TU. In the Present perfect case, however, the TU overlaps with TTOP, which semantically contains TSIT. Provided that the second sentence rhetorically expands on the reported event, the perspective of the character seemingly extends beyond the initial report.

The difference may also be described in terms of how the "contextual now," a contextually given reference point for the perspectival center, is established. An extended shift in the Past perfect would require the comprehender to infer that the contextual now remains in the present experience of the parenthetical subject. Such a reading is made more accessible when additional Type A indicators of non-speaker perspective are provided (28).

(28) a. Mary wasn't feeling well, thought John. Oh no – looks like she had fallen asleep on the couch again.

b. Mary wasn't feeling well, thought John. Had she fallen asleep on the couch again?

In the Present perfect case, however, a shifted contextual now can be established directly as the TU via the semantics of the Present.

There is a long tradition of positing a shift in TU to capture atypical uses of tense in vivid narrative forms (Reichenbach, 1947; Partee, 1973, 1984; Kamp and Rohrer, 1983; Comrie, 1985; Fleischman, 1990; Hornstein, 1990; Caenepeel, 1995; Michaelis, 1998, among many others). One well-known example among many is an account of the HISTORICAL PRESENT in which TU is back-shifted to the period described in the narrative:

(29) It was 1812, just before the Battle of Borodino. The anticipation of the coming struggle is palpable. Napoleon has just woken. He is getting ready to inspect the troops and see that they are ready for the battle that will determine the fate of Europe. (Hornstein, 1990)

A similar case can be made for FUTURE-IN-PAST, in which the contextual now can be understood as situated within an event described in the past (Kaufmann, 2005; Eckardt, 2017).

(30) Peter left the camp full of hope. He would propose to Sue, he would marry her and live happily for the rest of his life. (Eckardt, 2017)

Although I will remain agnostic about the precise mechanisms behind perspective shift in these cases, I hope that the intuitions

[4]The perfect is extremely complex and has several distinct uses; see Comrie (1976), Klein (1994), and Portner (2011) for discussion. Consideration of these uses is beyond the scope of the present project.

are clear. What remains now is to provide an account of how the two kinds of cues might conspire to signal and maintain extended perspective shift within the human language processing system.

2.2. Processing Perspective

The fact that disparate and complex linguistic information signals perspective raises an important series of questions for language processing research, of which just three are mentioned below. First, how does the processing system represent perspectival information about attitude holders? Second, is this information difficult to access and deploy during sentence comprehension? That is, is calculating and shifting a perspective center costly? Finally, once adopted, is perspective shift costly to maintain over longer stretches of text and discourse?

While these questions will not be addressed in any significant depth here, a basic picture is emerging. Evidence from numerous experimental studies suggests that comprehenders track the knowledge state and (presumed) viewpoint of attitude holders both in live conversation and in text. An early strand of research concentrated on mismatches between *common ground* information, i.e., information which other discourse agents are expected to know, and *privileged ground information*, i.e., information which is private to some individual or subgroup of the community, but not to all members of a discourse. Studies from Boaz Keysar and colleagues argue against an information processing system in which common ground information predominates over privileged ground. According to the monitoring and adjustment (Horton and Keysar, 1996) and the perspective adjustment (Keysar et al., 2000) models, common ground information does not have an impact until later stages of planning and comprehension. That is, language users plan and process utterances *egocentrically* first, only using information from the common ground in later stages to filter out unwanted ambiguities. However, Hanna et al. (2003) found evidence that participants are able to access common ground information early and circumscribe referential domains in a listening task, suggesting a *simultaneous access* view of common ground processing (see also Heller et al., 2008, 2010, and other since).

Further, readers and listeners appear to anticipate characters' likely emotional states while reading. Accordingly, they are surprised when a story does not conform to those expectations (Gernsbacher et al., 1992, 1998). In addition, language users consult a variety of potentially non-linguistic cues from the environment, and appear to generate expectations during interpretation (see Van Berkum (2010) for review). These cues include gender and social class (Van Berkum et al., 2008), political leanings (Van Berkum et al., 2009), and so on. Such information accumulates moment by moment within a representation of context (e.g., Clark, 1996). Thus, it seems clear that, at a minimum, readers and listeners monitor and use alternative perspectives during language comprehension.

How does the language processing system select the appropriate perspective? The discourse-economy approach taken here adopts a presumptive view of perspective processing in which calculating perspective follows a path of least resistance. Instead of continually generating hypotheses about whose perspective the speaker intended to discuss, it simply defaults to the presumption that the speaker is committed to the attitude expressed.

(31) **Speaker as Perspectival Center (SPC):** Take the speaker as the perspectival center, all else being equal.

The formulation of the principle highlights the presumptive nature of the proposal. Assuming from the maxim of Quality (Grice, 1975) that cooperative speakers are expected to believe what they assert, it follows that they should expect to be publicly committed to the content of their assertions (see also Levinson, 1983; Atlas, 2005), including presuppositions that project (e.g., Tonhauser et al., 2013). Following a similar logic as Harris and Potts (2009), speakers conveying non-speaker sentiment run the risk of being understood as committing to the perspective, unless they takes pains to signal otherwise. The strategies for signaling non-speaker orientation may vary depending on the genre, with specialized conventions for narrative that differ from other, perhaps more marked, forms of communication.

Although non-speaker uses of an epithet or a socially charged expressive to the incorrect individual carries with it a certain risk, the consequences of misattribution of perspectival terms are sure to vary according to the expression and the context of utterance. For example, the social ramifications of attributing a slur or epithet to the wrong attitude holder are generally greater than misinterpreting a predicate of personal taste. However, cases in which something important hinges on the contextual understanding of *beautiful* or a directional term like *left* can be constructed: for example, emergency instructions to a pilot from an air traffic controller. While these differences raise potentially interesting possibilities regarding whether speakers adjust their strategies according to the consequences of misattribution, the principle behind such cases remains the same: using a perspectival term to express a non-speaker orientation runs the risk of misattribution, leading to a violable preference for speaker-oriented uses.

A strictly speaker-centric view predicts resistance to perspective shifting generally, but has little say about the maintenance of perspetive, however. On the one hand, the processor might be eager to return to the speaker's perspective whenever possible. This view would be broadly compatible with accounts of perspective shifting that require a specialized mechanism for every instance of non-speaker perspective, e.g., a propositional operator that shifts parameters of the context. On the other hand, the processor might follow a path of least resistance when it comes to point of view. In this case, while shifting to a non-speaker perspective might be effortful and require a good deal of evidence, a shifted perspective might not require additional resources to maintain, provided the resulting text is coherent. This view would be broadly compatible with accounts that posit text level operations (e.g., Smith, 2003) or an extra-linguistic representation of perspective (e.g., Harris, 2012). In favor of the second position, I posit another discourse-economy principle favoring processing inertia:

(32) **No Shift Principle (NSP):** Don't shift perspectives unless required, e.g., evidence of incompatible viewpoint.

The NSP is predicted to interact with tense through the calculation of the contextual now. Forms like Present tense situate the TU within the contextual now. As long as the content within a sequence of clauses is consistent, the processor should avoid changing the contextual now as a reference point, thus ensuring that the perspectival center remains continuous.

In general, SPC and NSP discourage perspective shift, favoring speaker orientation by default. Once shifted, however, NSP favors to maintain the current perspective, as long the situation described is compatible with the ascribed viewpoint, resulting in an extended shift interpretation. Both principles can be understood as economical in the sense that they advocate against making potentially unnecessary updates to the discourse representation without evidence, but neither one prevents such updates from occurring, even rapidly.

Three experiments were designed to test how the availability of extended perspective shift might vary as a function of report type and tense. Attitudes were introduced with either Narrative parenthetical reports or Standard reports, and the continuation sentence following the report was presented in either Past or Present perfect.

(33) A. *Standard report:* John said that Mary wasn't feeling well.

B. *Narrative parenthetical report:* Mary wasn't feeling well, thought John.

Continuation:

i. *Past perfect:* She had fallen asleep on the couch again.

ii. *Present perfect:* She has fallen asleep on the couch again.

The central prediction was that extended shift interpretations of a Present perfect continuation would be more accessible when following a Narrative parenthetical (33B.ii). Further, if changing perspective exacts a processing cost, even with sufficient evidence for a shift, these continuations should be faster to process and interpret, as they allow the processor to maintain a single, shifted perspective.

3. EXPERIMENT 1: TWO-PERSON PRODUCTION-LISTENING STUDY

This experiment was designed to test the effect of tense on the interpretation of a sentence following different types of report. If the SPC is correct, then Narrative parenthetical reports should elicit more extended shift interpretations in perspective than Standard counterparts. According to the NSP, the availability of extended shifts should be increased by Present tense in the continuation, by allowing the contextual now to be uninterrupted, which results, by hypothesis, in a non-speaker perspective.

3.1. Participants

Fifteen pairs of undergraduate students at the University of Massachusetts Amherst participated in this study, for a total of 30 subjects. Participants were recruited from the Psychology department's subject pool and were offered class credit for their time. Upon entering the testing room, a Producer and a Listener was selected from the pair at random. Producers and Listeners were seated at separate computers, facing away from each other, so that non-verbal communication, such as facial cues and gestures thought to facilitate perspective shift (Harris and Potts, 2011), would not be a factor in the task.

3.2. Materials and Methods

The experiment consisted of 21 triplets in three conditions: *Standard report - Past tense* (34a), *Narrative parenthetical report - Past tense* (34b), and *Narrative parenthetical report - Present tense* (34c).[5] Labels are provided below for convenience and were not shown to participants. Each item consisted of a Report and a Continuation, and were followed by an interpretation question probing the viewpoint of the second sentence (35). Materials are provided in Appendix A.

(34) a. *Standard report - Present tense*
Mary said that there was a storm today.
Clouds have been brewing all morning.

b. *Narrative parenthetical report - Past tense*
There was a storm today, said Mary.
Clouds had been brewing all morning.

c. *Narrative parenthetical report - Present tense*
There was a storm today, said Mary.
Clouds have been brewing all morning.

(35) *Producer & Listener answer interpretation question in turn*
Was it Mary who said the second sentence, or someone else?

i. It was Mary. (*Extended shift response*)

ii. It was someone else. (*Speaker response*)

Items were interspersed with 33 items from two unrelated experiments. All items in the study probed perspective or commitment in some form or another. Experimental materials were presented in randomized counterbalanced order, so that pairs saw or heard only one item from each triplet.

Participants were instructed verbally before the experiment began, as well as in written instructions printed on the computer screen. The procedure for Producers and Listeners was somewhat different. Producers were first presented with the target sentence pairs. After reading the sentences, they were asked an interpretation question. After answering the question, they pressed a key to reveal the target sentence again, at which point the Producer read the target out loud to the Listener. Producers were instructed to "perform" the sentences, rather

[5]For simplicity, I have ignored complications that arise from the possibility of a "simultaneous" vs. "back-shifted" reading of the past embedded under a past tense verb (e.g., Ogihara and Sharvit, 2012). The sentences were designed to be biased toward a simultaneous reading, in which the event of the subordinate clause (*there was a story today*) overlap with the time of the report. The availability of multiple temporal interpretations of the embedded clause could be explored in future studies.

than to simply read them. All productions were recorded by a headset microphone connected to the computer.[6] After the Producer finished speaking, the Listener was presented with an interpretation question, much like the one that the Producer answered. Participants were allowed to break whenever desired, and typically finished the experiment in less than an hour.

3.3. Results

The response data was analyzed as a logistic linear mixed effect regression models with Condition as the fixed effect predictor. After models with maximal random effect models failed to converge (Barr et al., 2013), a model with by-subject and by-item random intercepts was computed. Predictors were treatment coded, so that each level was compared against Standard report - Present tense as the reference level.[7] The difference between Standard report—Present tense ($M = 51\%, SE = 5$) and Narrative parenthetical report—Past tense ($M = 54\%, SE = 5$) conditions was not significant, $z < 1$. However, the Narrative parenthetical report—Present tense condition ($M = 73\%, SE = 5$) prompted more extended shift responses than the Standard report—Present tense condition, $z = 5.21, p < 0.001$.

An additional, exploratory model was constructed in which the role of the participant was added as a predictor. The Narrative parenthetical report—Past tense condition was again the most likely to be associated with an extended shift interpretation, $z = 4.71, p < 0.001$. There was a numerical trend toward an interaction with this condition and Participant role, in which Listeners were marginally more likely to give an extended shift response than Producers were, $z = 1.68, p = 0.09$.

Participants agreed on their interpretations in 64% of trials in this study and in 62% of trials in the experiment overall. While the extent to which participants agreed on the interpretation of the experimental item did not differ significantly over conditions, there was a numerical trend in which the Narrative parenthetical report -Present tense condition ($M = 70\%, SE = 5$) elicited greater agreement, $z = 1.64, p = 0.10$, than Narrative parenthetical-Past ($M = 59\%, SE = 5$). Neither of the Narrative parenthetical conditions differed from the Standard-Present ($M = 65\%, SE = 5$) condition in terms of agreement rates. Limiting the data to trials in which Participants agreed—i.e., the 64% percent of trials in which Producer and Listener judged the token the same, we observe a similar effect as before: extended shift responses were more likely in the Narrative parenthetical-Present condition than in the others, $z = 3.87, p < 0.001$, and no other effects were observed.

3.4. Discussion

The experiment above investigated the effects of Present and Past perfect progressive sentences following different kinds of report contexts. This primary purpose was to test the basic intuition that the Present tense would prompt more extended perspective shifts when following a Narrative parenthetical than when following a Standard report. The reliable increase of extended shift responses in the Narrative parenthetical report - Present tense condition suggests that neither a Narrative report nor Present tense is sufficient to reliably signal an extended shift on its own. In addition, there was a marginal increase in agreement between participants when multiple cues converge, supporting the idea that multiple cues can increase the efficacy of signaling a perspective shift (Harris and Potts, 2009, 2011; Smith, 2009). However, the experiment lacked a Standard report - Past tense condition, preventing a full exploration of how the conditions may have interacted. The next study reports the results of a fully crossed designed in a interpretation judgment study to assess whether the interaction between condition is superadditive—i.e., the effects of each condition are greater than the sum of the effect of each condition individually.

4. EXPERIMENT 2: INTERPRETATION STUDIES

This experiment follows up on the previous experiment by adding a fourth condition to make a fully crossed design. The method was simplified to a standard single-participant forced-choice interpretation task. The central prediction was again that extended perspective shifts would be more likely in Present tense sentences following Narrative parenthetical reports, compared to all other conditions.

4.1. Participants

Thirty-six self-reported native speakers of English were recruited online from Amazon Mechanical Turk and were compensated with $6 for their participation. Only subjects who had performed at least 50 previous assignments and received a 98% approval rating or above were permitted to participate in the experiment.

4.2. Materials and Methods

The design consisted of 16 quartets crossing Report type (Standard report, Parenthetical report), which was always presented in past tense, and Tense (Past, Present) of the Target sentence. Aspect for these sentences varied across conditions, but were always the same within a quartet. See Appendix B for materials.

[6] Audio recordings were not analyzed statistically, but were reviewed impressionistically for cues of quotation, e.g., those discussed in Jansen et al. (2001) and Oliveira and Cunha (2004). Although no such cues were observed, a more systemic analysis is required to properly address the issue.

[7] As observed by a reviewer, random intercept linear mixed effects regression models are anti-conservative, and increase the chance of Type I error. Accordingly, Barr et al. (2013) recommend fitting "maximal" random effects structures, with by-subject and by-item random slopes and intercepts whenever possible. However, Matuschek et al. (2017) note that maximal models run the risk of inflating Type II error. To address these issues, Bayesian maximal mixed effect models were computed with `brms` (Bürkner, 2017), specified with 10,000 iterations and 1,000 warm up cycles. All effects in each of the three experiments exhibited the same pattern as those reported here with a standard 89% credible interval (McElreath, 2020) with the exception of the reaction times on questions in Experiment 3, for which divergent transitions could not be resolved. A simpler random intercepts Bayesian model showed the same pattern as the linear mixed effects regression model. Since the effects were replicated multiple times in different paradigms with different subjects, the effects are likely to be highly robust.

TABLE 1 | Experiment 2: mean percent and standard error of extended shift responses for each condition.

	Tense	
Report type	**Past**	**Present**
Standard	56% (4)	51% (4)
Parenthetical	67% (4)	78% (3)

(36) *Report sentence*

a. *Standard report*
Mary said that there was a storm today.

b. *Parenthetical report*
There was a storm today, said Mary.

(37) *Target sentence*
Clouds (had | have) completely covered the sky.

(38) *Interpretation question*
Was it Mary who said the second sentence, or someone else?

i. It was Mary (*Extended shift response*)

ii. It was someone else, like a narrator (*Speaker response*)

Materials were presented in the second half of a two-part study, and followed a sentence judgment task testing unrelated manipulations. The second part consisted solely of a forced-choice judgment task. Some of the items probed perspectival information. Experimental materials were interspersed with 20 items from unrelated experiments, and presented in counterbalanced and individually randomized order.

4.3. Results

The mean percent of extended shift responses and standard errors are provided in **Table 1** and depicted in **Figure 2**. The data were analyzed as linear mixed effect regression models with Report type, Tense, and their interaction as fixed effects and by-subject and by-items random intercepts, after a model with maximal random effects structures did not converge. Predictors were sum-coded so that Standard report and Past tense conditions served as the statistical baseline.

Parenthetical reports ($M = 73\%, SE = 3$) elicited more extended shift interpretations than Standard reports ($M = 53\%, SE = 3$), $z = 5.62, p < 0.001$. Although there was no general effect of Tense, Report type and Tense conditions interacted, $z = 2.40, p < 0.05$. Present tense targets following Parenthetical reports increased extended shift responses by 11 percentage points, but decreased these responses by 5 percentage points following Standard reports.[8]

[8]The continuation sentence was mistakenly presented in simple Past and Present, rather than in the Perfect, for three items (6, 7, and 10 in Appendix B). Removing these items did not change the patterns reported above. There was still a main effect of Report type and an interaction between Report type and Tense. The means were nearly identical to those obtained from the entire dataset.

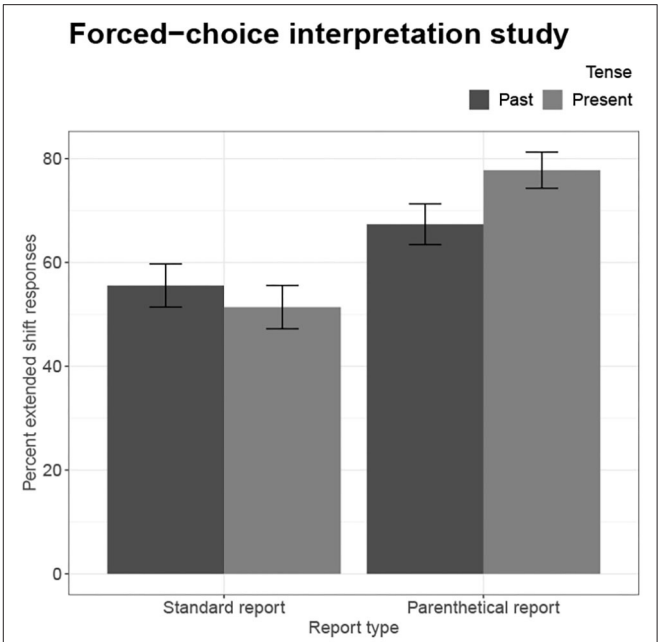


FIGURE 2 | Experiment 2: Forced-choice interpretation study.

4.4. Discussion

The results can be summarized by two central findings. First, Narrative parenthetical reports elicited more extended shift responses than Standard reports did. Second, Present tense in a continuation had a greater effect on interpretation when following a Narrative parenthetical report, supporting the idea that the effect of Narrative parentheticals and Tense on perspective is greater than the effect of either condition individually.

In this experiment and the previous one, there were relatively high proportions of extended shifted interpretations across the board. This is perhaps unexpected given the predictions of the SPC, which biases toward a speaker viewpoint. Several explanations are possible. The first is that the speaker default is empirically incorrect. A second possibility is that the proportion of shifted interpretations is partially artificial and due to experimental factors. For example, the questions may have been unclear, introducing additional noise into the responses. Given that a shifted interpretation was selected in approximately half of the trials, participants may have been more prone to guess on these trials. Further, as other items in both of these experiments probed perspective, subjects may have become desensitized to the speaker default, opting for shifted perspective responses more often. This issue was addressed in the third, and final, experiment.

5. EXPERIMENT 3: SELF-PACED READING STUDY

This study was designed with several objectives in mind. First, the previous experiments were conducted using offline measures,

which do not provide temporal information about how long participants read the passages or spent on the interpretation questions. According to an economy-based account, maintaining a shifted perspective is preferred and shifting perspective, even back to the speaker, exacts a cognitive cost. If correct, this account predicts a processing advantage for Present tense continuations following a Narrative parenthetical report. The central purpose of this experiment was to test this prediction.

Second, in the previous experiments, the reports were presented in Past tense regardless of the report type. Therefore, continuations in Present tense did not match with the tense of the report. It is possible that the mismatching tense was perceived as a marked narrative progression, leading participants to seek a non-standard extended shift interpretation (see Fludernik, 1993, for comments on the relationship between markedness and perspective). Although this explanation of Experiment 2 is unlikely, given that there was no independent effect of tense, it does raise the possibility of an alternate mechanism for extended perspective shift. In the third experiment, the tenses in the report and continuation always matched. The central prediction is again that the interpretation of Present tense continuations would be differentially affected by Report type, resulting in increased extended shift responses for Present tense continuations after a Narrative parenthetical.

Finally, to address potential habituation to non-speaker perspectives, no other items involving perspective shift were included, and interpretation questions were presented after only half the items. If the response rates observed in Experiments 1 and 2 were due, at least in part, to saturation with shifted perspectives, then the overall rate of extended shifts should be reduced.

5.1. Participants

Forty-eight subjects were recruited from the Claremont Colleges and compensated with course credit or $10 cash. All subjects self-reported as native speakers of English.

5.2. Materials and Methods

Materials consisted of 16 two-sentence discourses (39) modeled after those in Experiment 2. Conditions crossed Report type (Standard report, Parenthetical report) and Tense (Past, Present). The tense of the report and the continuation were matched, e.g., *was* paired with *had* and *is* with *has* or *have*, to avoid any effects of tense mismatch. Interpretation questions like (40) below were presented after half of the items.

Discourses were presented in a self-paced non-cumulative moving window fashion. The first sentence was always presented in its entirety on the first line. The second sentence was presented on region by region on the following line, as demarcated in (39). Regions of interest were identified as those that encode tense and/or aspect information: *Auxiliary* (Region 2), *Perfect participle* (Region 3), and *Verbal participle* (Region 4). Interpretation questions (40) followed half of the items to reduce participant saturation.

(39) *Report type*

a. | Mary said that there (was | is) a storm.

b. | There (was | is) a storm, said Mary.

Target sentence

$|_1$ Clouds $|_2$ (had | have) $|_3$ been $|_4$ forming $|_5$ in the sky $|_6$ all afternoon.

(40) *Interpretation question (half of items)*

Who said the second sentence?

i. It was Mary. (*Extended shift response*)

ii. It was someone else, like a narrator. (*Speaker response*)

Sentences were counterbalanced, so that each subject saw only one condition from the quartet. Items were presented in individually randomized order and interspersed with 66 items from unrelated experiments, 36 of which were also presented on two lines, and 28 non-experimental filler items, 12 of which were presented on two lines. None of these items were associated with perspective shift. The experiment lasted approximately 40 min on average per participant.

5.3. Results

Prior to analysis, outliers were censored according to a procedure known as winsorization (Dixon, 1960; Tukey, 1962). Observations above the bottom and top 5th percentile were transformed to the 5th and 95th percentile, respectively. The results did not differ in kind from those obtained with untransformed data. Adjusted means and standard errors for all regions are provided in **Table 2**. The second row of the table lists the mean time subjects spent reading the interpretation question and making a decision, as well as the percent of responses indicating an extended perspective shift.

Linear mixed effect regression models with Report type, Tense, and their interaction as fixed effects were created for regions of interest. Conditions were sum-coded as in Experiment 2. Decision time and responses to the interpretation question were modeled separately. Random effects were specified as by-subject and by-item random intercepts, after models with random slopes failed to converge. When interactions between conditions were significant in reaction time measures, the estimated marginal means of the model for Past and Present tenses were compared within each Report type using the *emmeans* package (Lenth, 2019), for which model estimates and 95% confidence intervals are reported. All significant effects are reported.

5.3.1. Reading Times on Regions of Interest

Statistical models are provided in **Table 3**. On the auxiliary (*has / have* or *had*) region, there was a general advantage for Parenthetical ($M = 404, SE = 7$) over Standard reports ($M = 425, SE = 10$), $t = -2.31, p < 0.05$. No effects were observed on the perfect participle (*been*) region. On the verbal particle (*V-ing*) region, there was a crossed interaction, $t = -3.13, p < 0.001$; see **Figure 3A**. While Present tense elicited a 38ms penalty in Standard reports (*Marginal means:* $\beta = 34.80, SE = 15.90$, 95% CI $[3.62, 66.06], p < 0.05$), it elicited a 33ms advantage in Narrative parenthetical reports (*Marginal means:* $\beta = -35.50, SE = 15.90$, 95% CI $[-66.73, -4.28]$, $p < 0.05$).

TABLE 2 | Experiment 3: self-paced reading means and standard errors in parentheses.

		Region						
Report type	**Report Tense**	**1: Report**	**2: Subject**	**3: Auxiliary**	**4: Perfect Participle**	**5: Verbal Participle**	**6: Spillover**	**7: Final**
Standard	Past	1939 (38)	570 (20)	419 (12)	388 (11)	410 (11)	486 (15)	677 (34)
	Present	1974 (39)	597 (24)	438 (18)	392 (11)	445 (18)	486 (14)	700 (31)
Parenthetical	Past	1774 (35)	584 (19)	411 (9)	400 (14)	443 (15)	491 (15)	714 (34)
	Present	1637 (31)	565 (15)	401 (11)	380 (8)	408 (11)	493 (16)	681 (30)
		Decision time on question	**Extended shift responses**					
Standard	Past	2,242 (38)	34% (5)					
	Present	2,244 (43)	47% (5)					
Parenthetical	Past	2,024 (43)	44% (5)					
	Present	1,891 (32)	81% (4)					

TABLE 3 | Experiment 3: linear mixed effect regression models for regions of interest.

Region	**Parameter**	**Estimate**	**Std. Error**	***t*-value**	***p*-estimate**
Auxiliary	(Intercept)	414.60	16.14	25.69	< 0.001
	Parenthetical	−10.67	4.62	−2.31	< 0.05
	Present	0.25	4.62	0.06	0.96
	Parenthetical × Tense	−6.58	4.62	−1.42	0.15
Perfect participle	(Intercept)	388.47	15.51	25.05	< 0.001
	Parenthetical	−0.70	3.86	−0.18	0.86
	Present	−3.41	3.86	−0.88	0.38
	Parenthetical × Tense	−5.37	3.86	−1.39	0.17
Verbal participle	(Intercept)	426.51	18.02	23.67	< 0.001
	Parenthetical	−0.84	5.62	−0.15	0.88
	Present	0.17	5.62	0.03	0.98
	Parenthetical × Tense	−17.59	5.62	−3.13	< 0.001

TABLE 4 | Experiment 3: linear mixed effect regression models for responses to and time spent on interpretation questions.

Measure	**Parameter**	**Estimate**	**Std. Error**	***t*/*z*−value**	***p*-estimate**
Responses	(Intercept)	0.15	0.30	0.49	0.63
	Parenthetical	0.70	0.14	5.00	< 0.001
	Present	0.78	0.14	5.54	< 0.001
	Parenthetical × Present	0.41	0.13	3.03	< 0.001
Decision time	(Intercept)	2100.15	33.22	63.22	< 0.001
	Parenthetical	−142.89	17.81	−8.02	< 0.001
	Present	−32.66	17.81	−1.83	0.07
	Parenthetical × Present	−33.70	17.81	−1.89	0.06

5.3.2. Interpretation Questions

Statistical models are provided in **Table 4**. Analysis of the interpretation questions revealed that Parenthetical reports were associated with increased extended shift responses, $z = 5.00, p < 0.001$, as was Present tense, $z = 5.54, p < 0.001$. As illustrated in **Figure 3B**, Present tense resulted in a greater increase in extended shift responses in the Narrative parenthetical report condition (diff = 37%) than in the Standard report condition (diff = 13%), $z = 3.03, p < 0.001$.

Decision times on interpretation questions showed that subjects were faster to read and respond to Narrative parenthetical report conditions, $t = -8.02, p < 0.001$, and were marginally faster in Present tense conditions, $t = -1.83, p = 0.07$. These conditions marginally interacted, in that there was a greater effect of Tense on Parenthetical reports than on Standard reports, $t = -1.89, p = .06$. Planned comparisons revealed a 133ms advantage for Present Narrative parenthetical reports (*Marginal means:* $\beta = 132.72, SE = 50.4$, 95% CI $[33.6, 232], p < 0.01$), but only a 2ms difference in Standard reports; see **Figure 3C**.[9]

5.4. Discussion

The responses to interpretation questions patterned very closely with the central results from the previous two experiments. Present tense continuations following Narrative parenthetical reports were associated with increased extended shift responses at a greater rate than Past tense continuations were. As the tense in the report and the continuation matched, the interaction cannot be attributed to tense mismatch leading to increased extended shift interpretations as a result of a more marked sequence. In addition, the percentage of extended shifts in conditions other than the Narrative parenthetical - Present tense condition was

[9]Note that the analysis suffered a considerable loss of power, as there were only 8 interpretation questions per subject. The lack of full significance in a case where the difference between conditions is so large may be the result of an underpowered analysis.

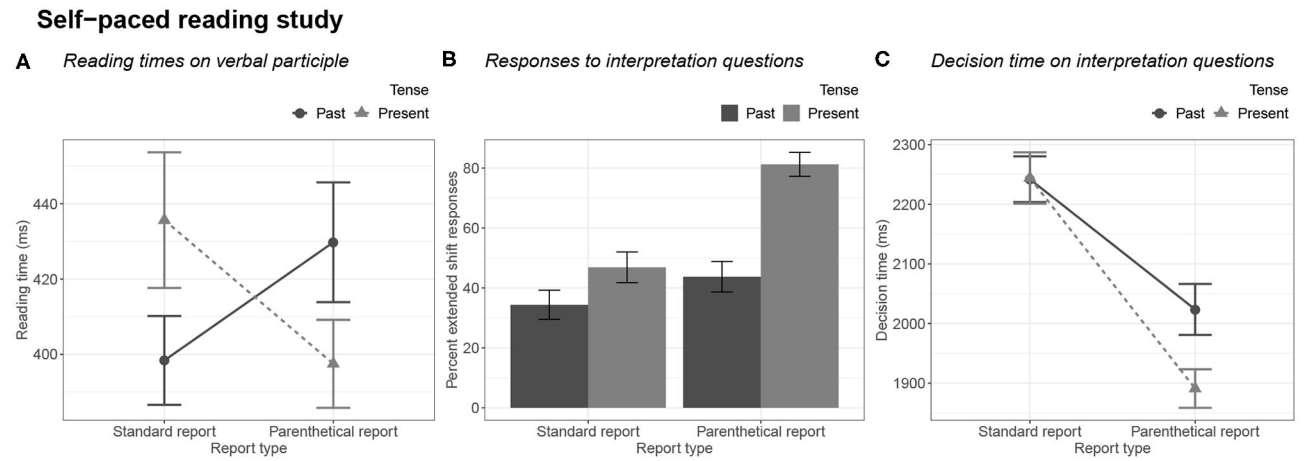


FIGURE 3 | Experiment 3: Self-paced reading study. **(A)** Reading times on verbal participle. **(B)** Reading times on interpretation questions. **(C)** Decision time on interpretation questions.

reduced, suggesting that the high rates observed in Experiments 1 and 2 may be partially attributed to experimental artifacts.

Decision times on interpretation questions also revealed an interaction between conditions, showing a greater advantage for Present tense continuations following a Narrative parenthetical report compared to a Standard report. Both of these results are highly compatible with the NSP, in that a non-speaker perspective after a perspective shift was predicted to be the preferred, and most economical, interpretation.

This approach made a similar prediction for the online, incremental processing of perspective. The expected interaction was observed, as participants read Present tense continuations faster after a Narrative parenthetical. However, the effect appeared on the verbal particle, two words after the morpheme marked for tense. While a delayed effect is very common in self-paced reading studies (Mitchell, 2004), it should be noted that the two previous regions trended in the same direction. The interaction is again compatible with the predictions of the discourse-economy account, in which a maintaining a non-speaker perspective is preferred after a perspective shift.

In addition, the interaction revealed an unexpected cost for Present tense continuations following Standard reports. To speculate briefly, it is possible that this effect reflects a pragmatic clash regarding the speaker's presumed epistemic warrant and discourse coherence. Let's assume that the Present tense, when anchored to a speaker perspective, indicates that the speaker has enough reliable evidence about the ongoing situation to license a statement about it. For example, the speaker would be taken as the evidential source for the statement *Clouds have been forming all morning*. However, one use of a Standard report is to mark the attitude holder as the source for report (e.g., Simons, 2007), resulting in a mild pragmatic clash. A similar clash would not arise in Past tense cases as there is no inference that the speaker has acquaintance with the events described, at least beyond what has been reported to her. Although this explanation remains speculative, it presents a possible avenue for exploring the role of discourse coherence relations between sentences in future studies.

6. GENERAL DISCUSSION

In this paper, we have concentrated on the phenomenon of *perspective shift*, in which a sentence (or sub-sentential constituent) may be presented as reflecting the perspective of an attitude holder other than the speaker. The topic has received increased attention in the theoretical linguistics literature, where several different subtypes of perspective shifting, and analyses thereof, have been proposed. Interest has also increased steadily among researchers in psychology and psycholinguistics, as a growing body of research suggests that perspectives are tracked and utilized during language comprehension.

I've sketched the outlines of an economy-based discourse model encapsulated by two basic principles. First, I proposed a general speaker default (SPEAKER AS PERSPECTIVAL CENTER; SPC), in which shifting away from a speaker's perspective requires ample evidence (following Smith, 2003; Potts's, 2005; Harris and Potts, 2009, among others). This principle serves to constrain perspective shift to cases where the speaker has provided the audience with sufficient cues to retrieve the intended message. Second, the NO SHIFT PRINCIPLE (NSP) proposed that perspectival continuity is preferred whenever possible. Together, the two principles predict that although shifting to a non-speaker perspective tends to be avoided, it is preferred after shifting has occurred. The principles converge on the idea that the processor doesn't continually assess perspective, but instead shifts only as required. Theoretically, this reluctance to shift results in making processing perspective more efficient.

The central predictions of the model were supported in three experiments. Following Reinhart (1975, 1983) and others, Narrative parenthetical reports were assumed to signal a non-speaker source for the reported content. Sentences following the report were interpreted from a non-speaker perspective at a

greater rate when presented in Present than in Past tense. The result was interpreted as a preference to continue the non-speaker perspective from the report in an extended shift interpretation. An online advantage for these sequences was also observed in a self-paced reading study.

Multiple cues indicating perspective and perspective shift have been discussed in the literature (e.g., Banfield, 1982; Wiebe, 1994; Smith, 2003). I proposed two classes of cues characterized by whether they (i) are associated with a perspective as an evaluator through quasi-conventional means (Type A), or (ii) grammatically determine how events are represented (Type B). I suggested that these two types are further distinguished by their primary role of establishing or maintaining a perspective shift, respectively. Following Smith (2003) and Harris and Potts (2009, 2011), the experimental results indicate that cues to perspective might work together to magnify the effect of perspective tracking. This observation leads to a natural follow up to determine whether additional perspectival devices facilitate perspective shift even further, and whether Type A cues may also serve to maintain an extended perspective shift, as discussed in connection with (28) in section 2.1.2.

The experiments presented here addressed whether the way an attitude is presented interacts with the tense of the following sentence in maintaining a shifted perspective. As the results were replicated across several different experimental paradigms with different subjects and filler sentences, the findings are likely to be fairly robust. However, the scope of the manipulations remained relatively constrained, and it remains to be seen whether the processing principles extend to other tenses, constructions, or perspectival cues, and, in particular, whether such cues also interact in extended perspective shift. As speakers may be wary of perspectival information "leaking" into their own public commitments (Lasersohn, 2007; Potts, 2007), they may tend to signal a non-speaker perspective with multiple cues to increase recoverability, particularly when the content is, in some way, important or controversial. Although how perspectival cues might interact remains an open question, existing evidence suggests that comprehenders are sensitive to multiple, interacting cues.

This issue was addressed in part by Harris and Potts (2009), who investigated whether appositive relative clauses, underlined in (41a–b), could shift to a non-speaker perspective when syntactically embedded under a verb of saying (Schlenker, 2007, 2009). Although embedding did increase the likelihood of perspective shift, they found that items containing an evaluative predicate like *paranoid* in the context were not only associated with more perspective shifted interpretations, but also elicited higher rates of perspective shift when embedded, compared to items without evaluative terms.

(41) [*Context:*] I am increasingly worried about my roommate. She seems to be growing paranoid.
   a. *Embedded:* The other day, she told me that we need to watch out for the mailman, a possible government spy.
   b. *Matrix:* The other day, she refused to talk with the mailman, a possible government spy.

Harris and Potts (2009) concluded that multiple factors contribute to non-speaker interpretations of subjective content, and that the combination of these factors seems to strengthen perspective shift. They proposed that non-speaker interpretations were marked as non-conventional and required additional evidence to shift perspective. Harris (2012, p. 13) further argues that the speaker default, and its violations, can be understood in terms of markedness, along the lines of Horn (1984):

> The distribution of speaker-orientation thus appears to follow a distribution of markedness. Speakers use expressions that are conventionally more marked when attempting to convey unfamiliar or otherwise less accessible meanings, at the cost of brevity dictated by the maxim of Manner. It is fairly natural to treat this distribution in terms of a default which enjoins the hearer to interpret a clause as speaker-oriented, unless there is specific evidence to the contrary [(31)].

Crucial to a successful perspective shift is the identification of a prominent discourse agent who can be associated with the shifted perspective (e.g., Hinterwimmer, 2019). Salem et al. (2018) conducted two online experiments investigating whether discourse particles and rhetorical questions increased the salience of a protagonist attitude holder in German. Increased salience was hypothesized to reduce time readers spent on an anaphor referring to the protagonist in passages of FID text. However, these markers did not influence the speed at which readers processed the anaphor online, and failed to reliably affect speed or accuracy on comprehension questions following the experimental passage. Such results would seem to contrast with those reported in Experiment 3.

It is possible that the accessibility of perspective shift is highly dependent on experimental design factors. The question of cue interaction was the subject of an additional exploratory pilot study ($N = 36$). Several different kinds of subjective constructions were presented to subjects to judge whose perspective was being portrayed with instructions similar to Experiment 2. One set of sentences (13 items) were modeled after the past tense conditions of the experiments above, and crossed report type with the presence of an interjection, such as *my goodness*.

(42) a. There was a storm today, said Mary. (My goodness.) Clouds had completely covered the sky.
   b. Mary said that there was a storm today. (My goodness.) Clouds had completely covered the sky.

Without an interjection, subjects provided extended shift responses after Narrative parenthetical and Standard report types at equal rates (52%). Interjections increased extended shift responses more for Narrative parenthetical (81%) than Standard (66%) reports, $\beta = 0.22, SE = 0.11, z = 2.03, p < 0.05$.

Also included in the pilot were passages with experiencer predicates and subjective terms (16 items). Statements were compared with questions in past and present tense:

(43) Deb hugged her son.
   a. He smelled like smoke.

b. He smells like smoke.
c. Did he smell like smoke?
d. Does he smell like smoke?

Present tense statements (43a) increased the rate of extended shifts (80%) compared to past tense statements (43a; 58%), compatible with the experiments above. However, tense had no effect on passages containing questions, which appeared to perform at ceiling (90% for both tenses).

While a more complete picture is still emerging, these preliminary results support the claim that calculating perspective relies on multiple, interacting factors. Some cues may exert varying degrees of influence; questions, for example, appear to be strong indicators of a non-speaker orientation, obscuring the effects of tense. This finding is highly compatible with a taxonomical distinction between cues that indicate a shift in perspective and those that maintain a perspective once shifted.

As noted above, perspective shifting has received many treatments. Multiple approaches treat perspective shift in terms of a silent operator which may adjust context-sensitive values in linguistic expressions or by altering parameters within the context. Approaches further differ according to the level of representation that the operator takes scope over. On sentence-based accounts, a context-shifting operator is posited at the sentence level, and presumably must be posited for each shifted sentence. On discourse-based accounts, a single perspective-shifting operator may scope over multiple, connected sentences in a fashion reminiscent to accounts of modal subordination (Roberts, 1989), as in Smith (2003) and Fabricius-Hansen and Sæbø (2004).

At first blush, the discourse-based account would appear to be more compatible with the claim that perspective shifting is avoided if the contextual values are consistent with the shifted viewpoint. In contrast, sentence-based accounts would not necessarily predict that extended perspective shifts would be as accessible as they are, as the processor would need to posit another context-shifting operator at each sentence. However, both approaches were designed to formalize how context-sensitive expressions are resolved within a particular representational framework, and not to make predictions about language processing behavior. To bridge the gap, they would each need to provide a principled account of the conditions under which context-shifting operators are licensed in order to make testable predictions.

Licensing conditions might be captured in various ways: for example, as an algorithm that determines whether a non-speaker perspective is available. Wiebe (1991, 1994) proposed an explicit procedure to determine the perspective of an utterance from features indicating subjectivity in a sentence (e.g., the forms listed in (12)) and properties of the current context, such as the most recent perspectival center. Abrusán (to appear) has recently offered a rich account that integrates this function with Segmented Discourse Representation Theory (Asher et al., 2003), a theory that models the mental states as a discourse context anchored to an attitude holder and the object of belief. This framework allows us to incorporate coherence relations between sentences, as well as the information structural status of referents in text, into an explicit procedure for identifying the perspectival center.

Another approach was detailed by Harris (2012), who proposed that a conversational scoreboard maintains information about discourse agents (Lewis, 1979), including the current perspectival center. The processor was assumed to passively construct a model of situations mentioned and implied by a text and the attitudes of central characters in text (e.g., Garrod and Sanford, 1985, 1995; Sanford and Garrod, 1998). The commitments, attitudes, and likely beliefs of discourse agents are collected into a representation called an "agent profile." He proposed that values in the scoreboard are updated by a general abductive inferencing procedure that provides the most reasonable interpretation given what is known about discourse agents via their profiles and presentational cues provided in the text. Assuming that abductive inferencing is computationally costly and prone to error, discourse economy provides a shortcut through this expensive inferencing process.

In summary, the three experiments reported here support the idea that Type A cues, like Narrative parenthetical reports, collude with Type B cues, like Tense, to produce an extended perspective shift. And while shifting a perspective away from the speaker may be costly to achieve, maintaining a non-speaker perspective itself may, at times, be the most economical option. In general, the results suggest that the perspectival center is not calculated at each moment in interpretation. Instead, the processor defaults to a speaker perspective unless it encounters evidence promoting a likely shift in interpretation, but may settle on a single perspective for longer stretches of discourse. Although the question of precisely how to model perspective remains open, the studies above support a model of perspective that is subject to general economy considerations, avoiding re-evaluating the perspectival center whenever possible.

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 the Internal Review Boards at the University of Massachusetts Amherst and Pomona College. The patients/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.

ACKNOWLEDGMENTS

Experiments 1 and 2 originally appeared in my dissertation (Harris, 2012). I am grateful to my committee, Lyn Frazier, Chris Potts, Barbara Partee, Adrian Staub, and Chuck Clifton, for their support and guidance. Experiment 3 was presented as the poster Shifting viewpoints and discourse economy at the 27th Annual CUNY Human Sentence Processing Conference, held at The Ohio State University in 2014. I thank members of the conference, especially Elsi Kaiser, for insightful comments and discussion. Thanks also to Mara Breen for her help programming Experiment 1 and to the undergraduate RAs for administering this experiment at UMass Amherst. Thanks to Joel Fishbein for his assistance in collecting the data in Experiment 3 at Pomona College. Finally, I thank the reviewers for their many insightful comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.613357/full#supplementary-material

REFERENCES

Abrusán, M. (2020). The spectrum of perspective shift: protagonist projection versus free indirect discourse. *Linguist. Philos*. doi: 10.1007/s10988-020-09300-z

Abrusán, M. (to appear). "Computing perspective shift in narratives," in *The Language of Fiction*, eds E. Maier and A. Stokke (Oxford University Press).

Amaral, P., Roberts, C., and Smith, E. A. (2007). Review of *The Logic of Conventional Implicatures* by Chris Potts. *Linguist. Philos*. 30, 707–749. doi: 10.1007/s10988-008-9025-2

Asher, N., Asher, N. M., and Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Atlas, J. D. (2005). *Logic, Meaning, and Conversation: Semantical Underdeterminacy, Implicature, and Their Interface*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195133004.003.0001

Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Boston, MA; Routledge and Keegan Paul, Ltd.

Barker, C., Bernardi, R., and Shan, C.-C. (2011). "Principles of interdimensional meaning interaction," in *Semantics and Linguistic Theory (SALT) 20*, eds N. Li and D. Lutz, 109–127. doi: 10.3765/salt.v20i0.2569

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, 255–278. doi: 10.1016/j.jml.2012.11.001

Buetow, H. (2014). *Taste disagreements and predicates of personal taste* (Ph.D. thesis). University of Massachusetts Amherst, Amherst, MA, United States.

Bühler, K. (1934). *Sprachtheorie*. Jena: Fischer.

Bürkner, P.-C. (2017). BRMs: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw*. 80, 1–28. doi: 10.18637/jss.v080.i01

Caenepeel, M. (1995). Aspect and text structure. *Linguistics* 33, 213–253. doi: 10.1515/ling.1995.33.2.213

Caenepeel, M., and Sandström, G. (1992). "A discourse-level approach to the past perfect in narrative," in *Semantics of Time, Space and Movement and Temporal Reasoning*, eds M. Aumague, A. Borillo, M. Borillo, and M. Bras (Toulouse: Groupe Langue), 167–188.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620539

Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.

Comrie, B. (1985). *Tense*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139165815

Cruse, D. A. (1986). *Lexical Semantics*. Cambridge University Press.

Dixon, W. J. (1960). Simplified estimation from censored normal samples. *Ann. Math. Stat*. 31, 385–391. doi: 10.1214/aoms/1177705900

Eckardt, R. (2014). *The Semantics of Free Indirect Discourse: How Texts Allow Us to Mind-Read and Eavesdrop*. Leiden: Brill. doi: 10.1163/9789004266735

Eckardt, R. (2017). Perspective and the future-in-the-past. *Glossa* 2. doi: 10.5334/gjgl.199

Emmott, C. (1999). *Narrative Comprehension: A Discourse Perspective*. Oxford: Oxford University Press.

Emonds, J. (1970). *Root and structure-preserving transformations* (Ph.D. thesis). MIT, Cambridge, MA, United States.

Fabricius-Hansen, C., and Sæbø, K. J. (2004). In a mediative mood: the semantics of the German reportive subjunctive. *Nat. Lang. Semant*. 12, 213–257. doi: 10.1023/B:NALS.0000034514.27887.d9

Fleischman, S. (1990). *Tense and Narrativity: From Medieval Performance to Modern Fiction*. Austin, TX: University of Texas Press.

Fludernik, M. (1993). *The Fictions of Language and the Languages of Fiction: The Linguistic Representation of Speech and Consciousness*. London: Psychology Press.

Frazier, L., Dillon, B., and Clifton, C. (2015). A note on interpreting damn expressives: transferring the blame. *Lang. Cogn.* 7, 291–304. doi: 10.1017/langcog.2014.31

Garrod, S., and Sanford, A. (1995). Incrementality in discourse understanding. *Increment. Interpret*. 11, 99–122.

Garrod, S., and Sanford, A. J. (1985). On the real-time character of interpretation during reading. *Lang. Cogn. Process*. 1, 43–59. doi: 10.1080/01690968508402070

Gernsbacher, M. A., Goldsmith, H. H., and Roberston, R. R. W. (1992). Do readers menally represent characters' emotional states? *Cogn. Emot*. 6, 89–111. doi: 10.1080/02699939208411061

Gernsbacher, M. A., Hallada, B. M., and Robertson, R. R. W. (1998). How automatically do readers infer fictional characters' emotional states? *Sci. Stud. Read*. 2, 271–300. doi: 10.1207/s1532799xssr0203_5

Grice, H. P. (1975). "Logic and conversation," in *Syntax and Semantics, Volume 3: Speech Acts*, eds P. Cole and J. Morgan (New York, NY: Academic Press), 41–58.

Gunlogson, C. (2008). A question of commitment. *Belgian J. Linguist*. 22, 101–136. doi: 10.1075/bjl.22.06gun

Hamblin, C. L. (1970). *Fallacies*. London: Methuen and Co.

Hamblin, C. L. (1971). Mathematical models of dialogue. *Theoria* 37, 130–155. doi: 10.1111/j.1755-2567.1971.tb00065.x

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, J. A. (2009). "Epithets and perspective shift: experimental evidence," in *University of Massachusetts Occasional Papers: Papers in Pragmatics*, eds M. Biezma and J. A. Harris (Amherst, MA: GLSA Publishing), 49–76.

Harris, J. A. (2012). *Processing perspectives* (Ph.D. thesis). University of Massachusetts Amherst, Amherst, MA, United States.

Harris, J. A. (2014). "Signaling non-speaker commitment in transparent free relatives: a paired speaker-hearer judgment study," in *The Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue*, eds V. Rieser and P. Muller (Edinburgh: DialWatt), 71–79.

Harris, J. A., and Potts, C. (2009). Perspective-shifting with appositives and expressives. *Linguist. Philos*. 36, 523–552. doi: 10.1007/s10988-010-9070-5

Harris, J. A., and Potts, C. (2011). "Predicting perspectival orientation for appositives," in *Proceedings from the 45th Annual Meeting of the Chicago Linguistic Society* (Chicago, IL: Chicago Linguistic Society), Vol. 45.

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., Grodner, D., and Tanenhaus, M. K. (2010). "The real-time use of information about common ground in restricting domains of reference," in *Semantic and Pragmatics: From Experiment to Theory*, eds U. Sauerland and K. Yatsushiro (London: Palgrave Macmillan), 228–248.

Hinterwimmer, S. (2017). "Two kinds of perspective taking in narrative texts," in *Semantics and Linguistic Theory*, Vol. 27, 282–301. doi: 10.3765/salt.v27i0.4153

Hinterwimmer, S. (2019). Prominent protagonists. *J. Pragmat.* 154, 79–91. doi: 10.1016/j.pragma.2017.12.003

Horn, L. R. (1984). "Towards a new taxonomy for pragmatic inference: Q and R based implicature," in *Meaning, Form, and Use in Context: Linguistic Applications, Georgetown University Round Table on Languages and Linguistics*, ed D. Schiffrin (Washington, DC: Georgetown University Press), 11–42.

Hornstein, N. (1990). *As Time Goes by: Tense and Universal Grammar*. Cambridge, MA: MIT Press.

Horton, W. S., and Keysar, B. (1996). When do speakers take into account common ground? *Cognition* 59, 91–117. doi: 10.1016/0010-0277(96)81418-1

Jansen, W., Gregory, M. L., and Brenier, J. M. (2001). "Prosodic correlates of directly reported speech: evidence from conversational speech," in *Proceedings of the ISCA Workshop on Prosody in Speech Recognition and Understanding* (Red Bank, NJ: International Speech Communication Association), 77–80.

Kaiser, E. (2015). "Perspective-shifting and free indirect discourse: experimental investigations," in *Semantics and Linguistic Theory*, Vol. 25, 346–372. doi: 10.3765/salt.v25i0.3436

Kaiser, E. (to appear). Effects of sensory modality on the interpretation of subjective adjectives: Comparing sight, smell and taste. Berkeley Linguistics Society.

Kaiser, E., and Lee, J. H. (2018). "Predicates of personal taste and multidimensional adjectives: an experimental investigation," in *Proceedings of WCCFL*, eds W. G. Bennett, L. Hracs, and D. R. Storoshenko, 224–231.

Kamp, H., and Rohrer, C. (1983). "Tense in texts," in *Meaning, Use, and Interpretation of Language*, eds R. Bäuerle, C. Schwarze, and A. von Stechow (Berlin: Mouton de Gruyter), 250–269. doi: 10.1515/9783110852820.250

Kaplan, D. (1999). What is meaning? Explorations in the theory of *Meaning as use*. Brief version-draft.

Kaufmann, S. (2005). Conditional truth and future reference. *J. Semant.* 22, 231–280. doi: 10.1093/jos/ffh025

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

Klein, W. (1994). *Time in Language*. New York, NY: Routledge.

Kölbel, M. (2004). Faultless disagreement. *Proc. Aristotelian Soc.* 104, 53–73. doi: 10.1111/j.0066-7373.2004.00081.x

Kratzer, A. (2008). "Situations in natural language semantics," in *The Stanford Encyclopedia of Philosophy. Stanford*, ed E. N. Zalta (Palo Alto, CA: Metaphysics Research Lab, Stanford University).

Kuno, S., and Kaburaki, E. (1977). Empathy and syntax. *Linguist. Inq.* 8, 627–672.

Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguist. Philos.* 28, 643–686. doi: 10.1007/s10988-005-0596-x

Lasersohn, P. (2007). Expressives, perspective and presupposition. *Theoret. Linguist.* 33, 223–230. doi: 10.1515/TL.2007.015

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.4.3.01.

Levinson, S. C. (1983). *Pragmatics*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511813313

Lewis, D. (1979). Scorekeeping in a language game. *J. Philos. Logic* 8, 339–359. doi: 10.1007/BF00258436

Maier, E. (2015). Quotation and unquotation in free indirect discourse. *Mind Lang.* 30, 345–373. doi: 10.1111/mila.12083

Maier, E. (2017). "The pragmatics of attraction: explaining unquotation in direct and free indirect discourse," in *The Semantics and Pragmatics of Quotation*, Vol. 3, eds P. Saka and M. Johnson (Berlin: Springer), 259–280. doi: 10.1007/978-3-319-68747-6_9

Malamud, S. A., and Stephenson, T. (2014). Three ways to avoid commitments: declarative force modifiers in the conversational scoreboard. *J. Semant.* 32, 275–311. doi: 10.1093/jos/ffu002

Mankell, H. (2009). *The Troubled Man*. New York, NY: Random House, Inc.

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

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. New York, NY: CRC Press. doi: 10.1201/9780429029608

Michaelis, L. A. (1998). *Aspectual Grammar and Past-Time Reference*. East Sussex, UK: Psychology Press.

Mitchell, D. C. (2004). "On-line methods in language processing:16 ntroduction and historical review," in *The On-Line Study of Sentence Comprehension: Eyetracking, ERP and Beyond*, eds M. Carreiras and C. Clifton, Jr. (Psychology Press), 15–32.

Mitchell, J. E. (1986). *The formal semantics of point of view* (Ph.D. thesis). University of Massachusetts, Amherst, Amherst, MA, United States.

Moltmann, F. (2010). Relative truth and the first person. *Philos. Stud.* 150, 187–220. doi: 10.1007/s11098-009-9383-9

Moore, C. (2011). *Quoting Speech in Early English*. Cambridge, UK: Cambridge University Press.

Ogihara, T., and Sharvit, Y. (2012). "Embedded tenses," in *Handbook of Tense and Aspect*, ed R. I. Binnick (Oxford: Oxford University Press), 638–668. doi: 10.1093/oxfordhb/9780195381979.013.0022

Oliveira, M. Jr., and Cunha, D. A. (2004). "Prosody as marker of direct reported speech boundary," in *Speech Prosody 2004, International Conference*, eds. B. Bel and I. Marlien (Nara, Japan), 263–266.

Partee, B. (1984). "Compositionality," in *Varieties of Formal Semantics: Proceedings of the Third Amsterdam Colloquium*, eds F. Landman and F. Veltman (Amsterdam: Foris Publications), 281–311.

Partee, B. H. (1973). Some structural analogies between tenses and pronouns in English. *J. Philos.* 70, 601–609. doi: 10.2307/2025024

Partee, B. H. (1989). "Binding implicit variables in quantified contexts," in *Papers from the Twenty Fifth Meeting of the Chicago Linguistic Society*, eds C. Wiltshire, R. Graczyk, and B. Music (Chicago, IL), 42–365.

Patel-Grosz, P. (2012). (*Anti-*)*locality at the interfaces* (Ph.D. thesis). Massachusetts Institute of Technology, Cambridge, MA, United States.

Pearson, H. (2013). A judge-free semantics for predicates of personal taste. *J. Semant.* 30, 103–154. doi: 10.1093/jos/ffs001

Peters, S. (2016). "Speaker commitments: presupposition," in*Proceeding of Semantics and Linguistic Theory 26*, eds M. Moroney, C. R. Little, J. Collard, and D. Burgdorf, 1083–1098. doi: 10.3765/salt.v26i0.3951

Portner, P. (2011). "Perfect and progressive," in *Semantics: An International Handbook of Natural Language Meaning*, Vol. 3, eds C. Maienborn, K. von Heusinger, and P. Portner (Berlin: Walter de Gruyter), 1217–1261. doi: 10.1515/9783110255072.1217

Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford, UK: Oxford University Press. doi: 10.1093/acprof:oso/9780199273829.003.0003

Potts, C. (2007). The expressive dimension. *Theoret. Linguist.* 33, 165–197. doi: 10.1515/TL.2007.011

Reboul, A., Delfitto, D., and Fiorin, G. (2016). The semantic properties of free indirect discourse. *Annu. Rev. Linguist.* 2, 255–271. doi: 10.1146/annurev-linguistics-011415-040722

Recanati, F. (2011). *Truth-Conditional Pragmatics*. Oxford, UK: Oxford University Press. doi: 10.1093/acprof:oso/9780199226993.001.0001

Reichenbach, H. (1947). *Elements of Symbolic Logic*. New York, Macmillian.

Reinhart, T. (1975). "Whose main clause? Point of view in sentences with parentheticals," in *Harvard Studies in Syntax and Semantics*, eds S. Kuno (Cambridge, MA: Harvard), 127–171.

Reinhart, T. (1983). "Point of view in language: the use of parentheticals," in *Essays on Deixis*, ed G. Rauh (Tübingen: Narr), 169–194.

Roberts, C. (1989). Modal subordination and pronominal anaphora in discourse. *Linguist. Philos.* 12, 683–721. doi: 10.1007/BF00632602

Ross, J. R. (1970). "On declarative sentences," in *Readings in English Transformational Grammar*, eds R. Jacobs and P. Rosenbaum (Waltham, MA: Ginn), 222–272.

Salem, S., Weskott, T., and Holler, A. (2018). On the processing of free indirect discourse. *Linguist. Found. Narrat. Spoken Sign Lang.* 247, 143–171. doi: 10.1075/la.247.07sal

Sanford, A. J., and Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Process.* 26, 159–190. doi: 10.1080/01638539809545043

Schlenker, P. (2004). Context of thought and context of utterance (a note on free indirect discourse and the historical present). *Mind Lang.* 19, 279–304. doi: 10.1111/j.1468-0017.2004.00259.x

Schlenker, P. (2007). Expressive presuppositions. *Theoret. Linguist.* 33, 237–245. doi: 10.1515/TL.2007.017

Schlenker, P. (2009). "Supplements within a unidimensional semantics II: epistemic status and projection," in *Proceedings of the 40th Annual Meeting of the North East Linguistic Society*, eds S. Kan, C. Moore-Cantwell, and R. Staubs (Amherst, MA. GLSA Publishing), 167–182.

Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguist. Philos*. 31, 353–395. doi: 10.1007/s10988-008-9039-9

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117, 1034–1056. doi: 10.1016/j.lingua.2006.05.006

Smith, C. S. (2003). *Modes of Discourse: The Local Structure of Texts*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511615108

Smith, C. S. (2009). "Accounting for subjectivity (point of view)," in *Text, Time, and Context: Selected Papers of Carlota S. Smith, Volume 89 of Studies in Linguistics and Philosophy*, eds R. P. Meier, H. Aristar-Dry, and E. Destruel (New York: Springer), 371–393. doi: 10.1007/978-90-481-2617-0_15

Sode, F. (2014). *Zur Semantik und Pragmatik des Konjunktivs der Indirektheit im Deutschen* (Ph.D. thesis). Humboldt-Universität zu Berlin, Berlin, Germany.

Speas, M., and Tenny, C. (2003). "Configurational properties of point of view roles," in *Aymmetry in Grammar*, ed A. M. di Sciullo (John Benjamins Publishing Company). doi: 10.1075/la.57.15spe

Stephenson, T. (2007). Judge dependence, epistemic modals, and predicates of personal taste. *Linguist. Philos*. 30, 487–525. doi: 10.1007/s10988-008-9023-4

Stokke, A. (2013). Protagonist projection. *Mind Lang*. 28, 204–232. doi: 10.1111/mila.12016

Tenny, C. L. (2006). Evidentiality, experiencers, and the syntax of sentience in Japanese. *J. East Asian Linguist*. 15, 245–288. doi: 10.1007/s10831-006-0002-x

Tonhauser, J., Beaver, D., Roberts, C., and Simons, M. (2013). Toward a taxonomy of projective content. *Language* 89, 66–109. doi: 10.1353/lan.2013.0001

Tukey, J. W. (1962). The future of data analysis. *Ann. Math. Stat*. 33, 1–67. doi: 10.1214/aoms/1177704711

Van Berkum, J. J. A. (2010). "The neuropragmatics of 'simple' utterance comprehension: an ERP review," in *Semantic and Pragmatics: From Experiment to Theory*, eds U. Sauerland and K. Yatsushiro (New York, NY: Palgrave Macmillan), 276–316.

Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., and Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychol. Sci*. 20, 1092–1099. doi: 10.1111/j.1467-9280.2009.02411.x

Van Berkum, J. J. A., 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

Wang, L., Reese, B., and McCready, E. (2005). The projection problem ofnominal appositives. *Snippets* 10, 13–14.

Wiebe, J. M. (1991). *Recognizing subjective sentences: a computational investigation of narrative text* (Ph.D. thesis). SUNY Buffalo, Buffalo, NY, United States.

Wiebe, J. M. (1994). Tracking point of view in narrative. *Comput. Linguist*. 20, 233–287.

Willer, M., and Kennedy, C. (to appear). Assertion, expression, experience. Inquiry.

Zifonun, G., Hoffmann, L., and Strecker, B. (1997). *Grammatik der deutschen Sprache*. Berlin: de Gruyter.

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.

Copyright © 2021 Harris. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Communication

ORIGINAL RESEARCH
published: 10 May 2021
doi: 10.3389/fcomm.2021.629757

Mental State Verb Production as a Measure of Perspective Taking in Narrations of Individuals With Down Syndrome

*Isabel Neitzel and Martina Penke**

Department of Rehabilitation and Special Education, University of Cologne, Cologne, Germany

Introduction: Perspective taking is an important ability to enrich narrations by empathizing with a real or fictional character. Mental state verbs (MSV) are a good indicator for this ability as they serve to reflect the mindset that the narrator attributes to a protagonist. Especially syntactic abilities have been argued to be relevant for MSV use. Investigating persons with Down Syndrome (DS) is likely to provide important insights into the relationship of MSV use and syntactic abilities: MSV are mostly used in complex sentence structures, which are a frequent difficulty for individuals with this syndrome. Indeed, previous investigations have found first evidence for impaired MSV production in individuals with DS, indicating a link to syntactic abilities and expressive vocabulary. Our aim was to provide evidence on MSV production and on the syntactic context of MSV production in individuals with DS and to target a possible connection to both cognitive and language abilities using specific language assessments. Typically-developing (pre-)school children were included as a comparison group to identify impaired respectively developmentally-adequate performance.

Method: 28 individuals with Down syndrome (aged 10; 0–20; 1 years) participated in a battery of cognitive, narrative and language measures. MSV-performance and syntactical context of MSV use were compared to data from 33 typically-developing children aged 3–9 years. We also analyzed the relationship between MSV production and language performance (vocabulary, syntax measures, mean length of utterance).

Results: The total number and types of MSV used were comparable for individuals with DS and TD. Moreover, a syntactic analysis indicated that individuals with DS and TD use MSV in the same syntactic contexts. Nevertheless, the syntactic difficulties of participants with DS are reflected in their frequent use of MSV in sentence-fragments. Correlations over the DS group yielded that syntactic abilities were not decisive for the richness and diversity of MSV in narrations.

Conclusion: Our findings suggest a comparable performance in MSV use in individuals with DS and school-aged TD children. The data indicate that MSV production is possible even with an impaired syntax suggesting unimpaired perspective taking abilities in individuals with DS.

Keywords: perspective taking, down syndrome, narration, mental state language, syntax, development of mental state verbs, syntactic impairment

OPEN ACCESS

Edited by:
Sonja Zeman, Ludwig Maximilians Universität München, Germany

Reviewed by:
José Sanders, Radboud University Nijmegen, Netherlands
Sue Buckley, Down Syndrome Education International, United Kingdom

****Correspondence:***
Martina Penke
martina.penke@uni-koeln.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *15 November 2020*
Accepted: *22 April 2021*
Published: *10 May 2021*

Citation:
Neitzel I and Penke M (2021) Mental State Verb Production as a Measure of Perspective Taking in Narrations of Individuals With Down Syndrome. Front. Commun. 6:629757. doi: 10.3389/fcomm.2021.629757

INTRODUCTION

Telling a story is an important part of social interaction in everyday life. A narration shows two main characteristics (Tompkins et al., 2013): the inclusion of (macrostructural) narrative elements such as actor, actions or scene and the perspective of the protagonist(s). Children need to understand the protagonist's intentions to comprehend a story as a whole (Kim, 2015). Furthermore, taking a protagonist's view is an important step of productive narrative development. Taking a protagonist's view requires the ability of perspective taking which allows to enrich narrations by empathizing with a (fictional) character. The ability of perspective taking requires both cognitive and linguistic abilities. Concerning cognitive functions, this especially involves Theory of Mind (ToM) development (see Symons, 2004 for an overview of perspective taking development and the connection between perspective taking and mentalizing). Theory of Mind (ToM) refers to an individual's ability of decoding other people's intentions and emotions. ToM enables a person to understand mental states within others to comprehend and predict their assumptions and behaviors. A critical achievement in ToM acquisition is false belief understanding (Tompkins et al., 2019, p. 111), which is used in most studies investigating ToM (Devine and Hughes, 2014; Beaudoin et al., 2020) and which reveals the connection between ToM and perspective taking. In the most common false belief task, the Sally-Anne-task by Baron-Cohen et al., (1985), the child witnesses that the puppet Anne hides the toy of the puppet Sally in a new place while Sally is away. To answer the target question correctly ("Where will Sally look for the marble?"), the child needs to take the perspective of Sally, who does not know the new place of the toy, and to suppress the own knowledge about the real hiding place in answering the question.

Linguistically, so-called mental state verbs (MSV) are important means for expressing the view of a protagonist or character in a narration (overview in Van Krieken et al., 2017). MSV are content verbs pointing towards mental states, beliefs or wants of a protagonist (Perner et al., 2003) or another character in a story. A common classification of MSV is semantic and differentiates three main categories (e.g. Jenkins et al., 2003): volition/desire (want, hope…), cognition (think, say…) and emotion/effect (like, hate…). Consider the following narration: "Yesterday, Susan walked along the neighbor's house when the little dog of the owner ran out and barked terribly behind the fence. After Susan returned home, she told her mother that she hated this frightening dog." In this example, the MSV "told" does not only show the mental state of the protagonist (hateful, frightened) but it also clearly points to a change in perspective of the narrator as it introduces the view of the fictional story character whereas the narrator might like dogs or does at least not know the (fictional) neighbor's dog. Van Krieken et al. (2017) point out that perspective taking in narratives can be expressed in different dimensions, e.g. an emotional and a cognitive dimension. The authors give a review of different linguistic elements to formulate a protagonist's view in these dimensions. One way to do this is to use MSV expressing the emotional or mental states of oneself or other persons.

In typical development, first MSV occur by the age of two and their occurrence strongly increases from the age of three onwards (see Channell, 2020, for an overview). During development, a connection between usage of MSV and the production of complex syntax seems plausible, as the description of other people's mental states often requires complex sentence structures, especially the mastery of syntactic complementation (De Villiers and Pyers, 2002) as in our narrative example before: "Susan told her mother that she hated this frightening dog." De Villiers & Pyers suggest that the mastery of this complex sentence construction enables children to represent other person's beliefs, a prerequisite for taking another individual's perspective. Thus, the acquisition of complex syntax involving sentence complements might constitute a prerequisite for using MSV expressing another person's perspective. While in English, only the verb *think* strongly encompasses the usage of a complement clause, Perner et al. (2003) point out that in German, substantially more verbs take a sentential complement in the form of a subordinate clause (e.g. want *that*, say *that*, think *that*…). Notably, all these verbs meet the definition of MSV, thus, bolstering the assumption that MSV use might be related to the mastery of complex syntactic structures, specifically involving sentence complementation and subordinate clauses.

A population which might be especially affected by the assumed connection between perspective taking (as expressed by MSV) and syntactic abilities are individuals with Down syndrome (DS). Individuals with DS display marked deficits in the production and comprehension of complex syntactic structures involving subordination (Abbeduto et al., 2007; Wimmer et al., 2020). Furthermore, individuals with this syndrome frequently show difficulties in narration (Neitzel & Penke, under review; overview in Segal and Pesco, 2015) and Theory of Mind (Neitzel and Penke, 2021). Thus, the investigation of MSV production in individuals with DS might provide important insights into the interrelation between MSV usage as a linguistic measure of perspective taking and the mastery of complex syntax production. Although investigations on this topic are sparse, the few existing studies addressing MSV production in individuals with DS have indeed pointed to an association of syntactic impairments and deficits in MSV usage.

Beeghly and Cicchetti (1997) compared the productive vocabulary of 2–5 year-old individuals with DS (n = 39) and typically-developing children (TD) (n = 38) with respect to internal state language. In contrast to MSV which specifically focus on people's feelings or intention, internal state language includes MSV but also verbs of perception or physiology. MSV were included but not coded as a separate category. The children were separated into two groups concerning their cognitive abilities: group 1 included individuals with a mental age of less than 30 months (individuals with DS: n = 18), whereas group 2 contained individuals with a mental age of above 30 months (DS: n = 21). Data were collected in a mother-child picture-book situation and during a period of semi-structured play. The authors describe that children with DS produced less internal state words than TD children in both mental-age groups. Although mental age was significantly associated to internal

state word production in individuals with DS, their level of internal state language stayed behind expectations for their mental age. Furthermore, the authors determined the mean length of utterances (MLU) as a measure for syntactic development and found children with DS to be significantly delayed in comparison to TD children. This study points to an impairment of mental state language and an association between mental state language and syntax in individuals with DS. However, the missing of a specific category for MSV and the use of an unspecific measure of syntactic development leave questions regarding the link between MSV production and syntactic abilities open.

In contrast to Beeghly and Cicchetti (1997) and Grela (2002) found the number of MSV to be comparable between individuals with DS and TD children. The author computed lexical verbs in language samples from a free play situation with the child's mother at home, containing 294 utterances per child, and found no significant group difference for MSV between the DS and TD group. However, this investigation only included seven individuals with DS (chronological age: 6;2–12;2) and seven TD participants, matched for MLU, limiting claims regarding MSV and their association to syntactic development in individuals with DS.

Larger groups of individuals with DS were investigated in three recent studies. The study of Ashby et al. (2017) included 23 children with DS (chronological age: 10–16 years) and a group of younger TD children matched for performance in an intelligence measure. All children were asked to tell a story based on a nonverbal picture book. The authors coded each utterance of the produced narrations for 15 inferential aspects. MSV were contained in one category measuring "internal states" but were not evaluated separately. According to the authors' analysis, children with DS showed significantly less inferential language than TD children. The authors employed MLU as a measure of syntactic complexity and suggested that the difficulties in inferential language might be due to a morpho-syntactic problem, rather than to a specific narrative or inferential deficit. While Ashby et al. (2017) suggest "a clear link between complexity of sentence structure and use of inferential language during narration in individuals with DS" (p. 105) they state that the direction a causal relationship might take remains unclear.

In 2020, Channell described difficulties in MSV production for a large group of individuals with DS (n = 40, chronological age 6–11 years) and found evidence for impaired MSV production in narrations obtained by a nonverbal picture book, both in terms of the total number of utterances including MSV and in terms of the diversity of MSV used. Furthermore, the author computed correlations between MSV use and measures for nonverbal cognition, an emotion knowledge task, an expressive vocabulary test and once again MLU as a syntactic measure. Significant correlations only held between MSV use and MLU as well as between MSV use and expressive vocabulary. Based on these findings, the author suggested a connection between MSV use and syntactic abilities and named expressive vocabulary as another key factor for MSV production. However, the investigation included no control participants with TD.

In a very recent study, Martzoukou et al. (2020) investigated narrative abilities of a group of 20 Greek adults with DS (chronological age 19–46 years). The participants were asked to retell two 6-picture-stories from the narrative instrument MAIN after auditory presentation. Macrostructural analyses of the re-narrations involved the enumeration of internal state words. Although MSV were not specifically coded, cognitive and emotional MSV were included in this category. In addition, to explore the participant's ability to recognize a character's internal state, all individuals were asked comprehension questions addressing the internal states of the characters after their re-narrations. MLU as well as a sentence-repetition test were adopted as measures of syntactic ability. Martzoukou et al. (2020) report that in comparison to two groups of typically-developing preschool children (each n = 20) - one matched by nonverbal cognition, one matched by expressive vocabulary abilities - individuals with DS produced less internal state terms in their re-narrations, although the comprehension questions revealed an adequate understanding of the character's beliefs, thoughts and emotions. In addition, Martzoukou et al. report poor morpho-syntactic abilities in their participants with DS reflected in significantly poorer performance in sentence-repetition compared to both groups of control children. Moreover, MLU was significantly lower compared to the control group matched for expressive vocabulary. Based on these findings, Martzoukou et al. (2020) conclude that individuals with DS might not be impaired in understanding the internal state of a (fictional) character but that their poor morpho-syntactic abilities might limit their expression of internal states.

Summarizing, the few existing studies provide some evidence for difficulties in MSV production in children, adolescents, and adults with DS, connecting these difficulties to impaired syntactic abilities and expressive vocabulary. However, none of the studies specifically focused on the relation between the production of MSV and the development of particular syntactic constructions that might be especially relevant to express perspective taking via MSV (i.e. sentence complementation, see De Villiers and Pyers, 2002).

A possible association between perspective taking, MSV and syntactic abilities is of special interest for clinicians working with atypically-developing individuals that show syntactic impairments. If these were associated to each other, one could assume that individuals who show syntactic disabilities might also show impairments in expressing another character's perspective. These difficulties might result in impaired narrative abilities and difficulties in story understanding (Kim, 2015). As narration is an important part of everyday interaction with others, narrative deficits might cause problems with respect to social relationships when children stay behind the narrative skills of their school peers. Furthermore, narrative abilities are a frequent aim in oral and written school exercises and research has proven a close connection between narrative abilities and later literacy (Griffin et al., 2004) or even mathematical skills (O'Neill et al., 2004). Thus, a limited expression of one's own and other person's perspectives might lead to limitations in every day social interaction or school life, suggesting MSV as a relevant topic for therapeutic intervention.

TABLE 1 | Chronological age (CA) and nonverbal mental age (MA) over participants with DS and TD groups. Nonverbal mental age was not available for TD participants.

	Group DS (n = 28)	Group TD: age groups		
		TD3 (n = 12)	TD5 (n = 11)	TD9 (n = 10)
CA (y; mm)				
Mean	14; 05	3; 08	5; 04	9; 06
SD	2; 06	0; 02	0; 04	0; 05
Range	10; 00–20; 01	3; 03–3; 11	5; 00–5; 11	9; 00–9; 11
MA (y; mm)		*n.a.*	*n.a.*	*n.a.*
Mean	5; 03			
SD	1; 02			
Range	3; 05-<7; 11			

Note: MA of one participant in the DS group exceeded the norming sample of the SON-R (<7; 11).

The aim of our study was to provide specific evidence on MSV production in narrations of individuals with DS and its implications for perspective taking abilities in individuals affected by this syndrome. We focused on the types of MSV produced and on the syntactic structures in which MSV appeared. By uncovering parallels and differences to typical development, we aimed to obtain deeper insight into the particular problems individuals with DS might display with respect to expressing perspective taking. If MSV usage was dependent on expressive vocabulary or on syntactic abilities, especially related to the mastery of complex syntax structures, individuals with DS that show impaired lexical/syntactic abilities should display difficulties in MSV usage, visible in the total number and the diversity of MSV. If, on the other hand, our analyses were to show that MSV production in individuals with DS was adequate despite impaired lexical or syntactic abilities, this would suggest that the expression of mental states in persons with DS via the use of MSV is independent from language and/or poor cognitive development.

We compared MSV production in narrations of 28 children and adolescents with DS to MSV production of 33 preschool and school children aged 3, 5 and 9 years. Furthermore, our participants with DS completed a battery of measures targeting nonverbal cognitive development as well as specific language measures. These included a measure for expressive vocabulary that was found to be related to MSV production by Channell (2020) and a measure targeting the development of complex syntactic structures.

MATERIALS AND METHODS

Participants

28 children and adolescents with Down syndrome participated in a comprehensive assessment battery including different cognitive and language measures. The data were obtained in two to three testing sessions per participant which lasted about 45–60 min and included sufficient breaks. All participants (n = 15 females, n = 13 males) were recruited by contacting special-need schools and parent organizations. Most participants displayed a free trisomy (n = 26); one participant was diagnosed with mosaic trisomy and the form of trisomy was unknown for one of the participants. All children and adolescents attended inclusive or special-need schools. Informed consent for participation in the study was obtained from participants as well as from their parents. The research project was approved by the Ethics Committee of the Medical Department of the University of Cologne (number of approval 18–121). Age characteristics of all groups are given in **Table 1**. Mean chronological age over the group was 14; 05 (given in y;mm, *SD* 2;06, range 10;00-20;01). Mean nonverbal mental age as measured with the SON-R 2 ½-7 (Tellegen et al., 2007; see 2.2) was 5;03 (*SD* 1;02, range 3;05–8;00). All participants were monolingual German speakers and had no other reported medical conditions or psychiatric diagnoses. Visual loss was corrected by glasses in all cases. For nine participants, parents reported a slight hearing loss (10–25 dB), however, hearing loss did not exceed 25 dB in any of these participants.

In addition, narrations of typically-developing (TD) children were taken from the Bamberg-corpus available on *Childes* (Berman and Slobin, 1994). Written transcripts of the so-called "Frog story" (see 2.2.1) were obtained from 3 year-olds (n = 12, range 3; 3–3; 11), 5 year-olds (n = 11, range 5;00-5;11) and 9 -year-olds (n = 10, range 9;00–9;11) (see **Table 1**). These age groups seemed especially suitable for a comparison concerning MSV production and syntactic context of MSV use as MSV vocabulary as well as complex syntax both develop at preschool age, more precisely between the age of three and five (see Channell, 2020 for an overview). Recall that the mean nonverbal mental age of the individuals with DS was 5; 03 years. Therefore, a comparison of the performance of the group with DS to these three groups of TD children allowed for a comparison to a younger group of TD children, to an older group of TD children, and a group of TD children matched in chronological age to the mental age of the individuals with DS. As neither sex nor mental age or performance in language measures were given for the participants from this corpus, these factors could not be included into statistical analyses comparing performance for participants with DS and TD children.

Instruments

Narrative Task

All participants' narratives were collected using a non-verbal picture-book ("Frog, where are you?"; Mayer, 2003). The so-called "Frog story" is a widely used instrument to obtain narrations in children. With its richness in protagonists and the black-white pictures, it is also suitable for older children and adolescents with DS. The story consists of 24 pictures illustrating the story of a young boy and his dog, who search for their pet frog which has escaped from his glass. On their way, they meet several animals and experience some adventures. For the participants with DS, we followed the procedure typically chosen when eliciting narratives of the Frog story. The book was first presented nonverbally by the experimenter so that the participant was able to learn about the protagonists and the storyline. Afterwards, the participant was asked to retell the story picture by picture. The images were presented while telling the story and the experimenter took care that every picture was

TABLE 2 | Categories for syntactic context of MSV and examples from the DS group.

Syntactic context of MSV	Example (participant)
[i] in a main clause	*Der Hirsch will trinken./*The deer wants to drink. (P02)
[ii] in a main clause followed by a subordinate clause	*Max möchte Timmy helfen, weil er hängt noch aufn Hirschkopf auf fest./*Max wants to help Timmy because he is still stuck on the deer's head on. (P17)
[iii] in a main clause followed by direct speech	*Dann sagte der Junge: Wo ist der Frosch denn hin?/*Then the boy said: Where did the frog go? (P09)
[iv] in a complex sentence (e.g. subordinate clause or question)	*Und der Hund ist enttäuscht, dass die Bienen sich geärgert haben./*And the dog is disappointed that the bees were annoyed. (P30)
[v] in sentence fragment	*Geflippt aus./*Freaked out. (P03)

included into the story. During the narration, only unspecific questions such as "What (else) happens here?" were asked by the experimenter. The narration was documented using video and audio recording. Afterwards, all stories were transcribed using ELAN 5.3 (Max Planck Institute, 2018). The written transcript was used for further analyses. For TD children, we used the written transcripts available on *Childes* (see 2.1) for further analyses.

Perspective Taking Measure

We analyzed usage of MSV in the frog-story narrations as a measure for perspective taking. MSV expressing volition (e.g. *want*), cognition (e.g. *think*) and emotion (e.g. *rejoice*) were included in this measure. MSV in the formulaic utterance '*(I) don't know*' were excluded from the analyses. For each Frog-story narration, we counted the total number of verbs produced and the total number of MSV (tokens) to determine the density of the verb inventory for each participant. We also determined the number of different verbs and different MSV (types) as an indication of the diversity of the verb lexicon. All categorizations to the three MSV types were coded by the same rater for both DS and TD participants and co-checked by a trained second rater. The categorizations used for the analyses were consistent in both ratings. A full table of all included MSV and their categorization is displayed in **Supplementary Table SA**. Subsequently, proportions of MSV in relation to verbs total were calculated for verb types and verb tokens for each narration and participant.

Furthermore, we determined the syntactic context in which each MSV appeared. Each utterance containing a MSV was coded for one of five syntactic contexts the MSV appeared in: [i] in a main clause, [ii] in a main clause followed by a subordinate clause, [iii] in a main clause followed by direct speech[1], [iv] in a complex sentence (e.g. a subordinate clause or question), [v] in a sentence fragment (see examples in **Table 2**). We also calculated the MLU (in words) for all participants since MLU had been used as a measure of syntactic development in all previous investigations.

[1]Direct speech is defined as the combination 'of two (or more) main clauses, representing the quoted utterance(s) on the one hand, and a framing clause on the other.' (Nordqvist, 2001, p. 58).

Non-verbal Cognition

Non-verbal cognition of children and adolescents with DS was assessed using the *Sniders-Omen Nonverbal Intelligence Test* (SON-R 2 ½-7; Tellegen et al., 2007). This measure includes a reasoning scale, which consists of three non-verbal subtests (categories, analogies, situations) and allows to compute the children's nonverbal mental age (see 2.1).

Specific Syntax Measure

To explore the relationship between MSV use and the mastery of complex syntactic structures, we used the *ESGRAF 4-8* (Motsch and Rietz, 2016), a standardized German instrument targeting the production of different complex syntactic structures in a circus setting. Specifically, the *ESGRAF 4-8* elicits subordinate clauses, *wh*-questions and main clauses with topicalized sentence constituents. *ESGRAF 4-8* offers norming data for 4–8 year-old children. The mental age of the children with DS was used to choose the norming scale for each participant.

The first subtest of *ESGRAF 4-8* assesses the production of *wh*-questions and topicalized sentences. In the first part of the subtest, the child is asked to produce 12 *wh*-questions to identify a toy animal hidden in a box (e.g. *Wo wohnst du?* "Where do you live?"). Subsequently, the child is asked to take the perspective of the animal and to formulate what she would like to eat and what she has eaten, resulting in a maximum of 24 utterances with a topicalized object (e.g. *Brot mag ich nicht.* "Bread, I don't like.") or a fronted adverbial (Adv VS) (e.g. *Heute esse ich Brot. Morgen esse ich den Apfel.* "Today, I eat bread. Tomorrow, I eat the apple."). In total, the subtest contains 36 items. A point is scored for each produced sentence that correctly contains the fronted element (*wh*-pronoun, object or adverb) before the finite verb in second structural position followed by the subject (max. 36 raw points). An imitation aid is possible for each item if the child is not able to produce one of the items by herself. The child scores one point for the production of the correct sentence after the imitation aid as well.

The production of subordinate clauses was assessed by the second subtest of the *ESGRAF 4-8*. In German, MSV are often accompanied by subordinate clauses that express the sentential complement of a MSV (e.g. *Susan told her mother that she hated this frightening dog.*). In contrast to English, German subordinate clauses display a different order of sentence constituents compared to main clauses. Whereas main clauses typically display SVO order, in subordinate clauses the finite verb is placed clause-finally, resulting in SOV (e.g. *Der Junge ist*

traurig, weil sein Frosch entwischt ist. "The boy is sad because his frog has escaped."). The subtest targets the acquisition of subordinate clauses by evaluating verb placement. It contains 20 items and is separated into three short tasks that require the child to produce subordinate clauses with different subordinating conjunctions while ordering items for the circus' magician, answering questions about magic tricks, or repeating magic rules. According to the manual, a point is scored if a subordinate clause with correct verb placement is produced independently or imitated after exemplification of the experimenter (=max. 20 raw points).

Vocabulary Measure

Expressive vocabulary was assessed using the *AWST-R* (Aktiver Wortschatztest für 3- bis 5-jährige Kinder – Revision; Kiese-Himmel, 2005), which is a standardized German vocabulary test. The instrument contains 75 items (51 nouns, 24 verbs). The verbs included in the measure are concrete verbs (e.g. *schneiden* "cut") and do not contain emotional or mental state verbs. All items are presented as colored photographs and the child is asked to name the pictures. According to the manual, the child scores one raw point (max. 75) for each item named correctly. Norming data is available for children aged 3; 00 to 5; 05 years.

Data Analysis

TD participants were separated into three age groups according to their chronological age: (i) three-year-olds (TD3, $n = 12$), (ii) five-year-olds (TD5, $n = 11$) and (iii) nine-year-olds (TD9, $n = 10$). Based on the written transcripts of the Frog-story, we counted occurrences of verbs (types and tokens) as well as MSV for each participant (DS and TD) and narration. Each utterance was coded for the type of MSV: volition, cognition or emotion verb. In a syntactic analysis, we coded the syntactic context for each MSV (see 2.2.2) per utterance and calculated the mean length of utterances (MLU) for words in each participant. Group means were computed for the three groups of TD children and the group with DS. Differences between the group with DS and the groups of TD children were computed using non-parametric statistical analyses to account for the fact that proportions were not normally distributed. Pairwise comparisons for each group combination (e.g. TD3 vs. DS) were computed by post hoc analyses using *SPSS 25* (IBM, 2017).

For participants with DS, raw scores were calculated for the vocabulary measure and the two syntax measures according to the manual. Based on these raw-scores, the norming data given in the respective tests and the nonverbal mental age of each participant as determined by the SON-R, we subsequently derived T-scores for each individual and each test. T-scores allow to address whether individual scores obtained by participants with DS are at the level expected for their nonverbal cognitive development. A T-score above 40 indicates performance that is appropriate for the participant's nonverbal mental age, whereas a T-score below 40 indicates substandard performance indicative of impairment in the respective test. Note that for 10 participants with DS norming data could not be obtained for the *AWST-R* as this measure only includes norming data for 3; 0 to 5; 05 year-old children. Pearson's correlations were computed using *SPSS* to analyze the relation between MSV use and language performance (vocabulary, MLU, complex syntax) as well as nonverbal mental age in participants with DS.

RESULTS

Density and Diversity of MSV in Narrations

In total, participants produced 420 utterances containing MSV in their narratives. Of these, 224 utterances were produced by TD participants ($n = 33$) und 196 utterances by the 28 participants with DS. **Table 3** gives an overview on the production of verbs and MSV. A total (cumulated number for all participants over the group), mean, *SD* and range are given for each measure and column. Key measures following Channell (2020) are the density of MSV (column 2), which means the number of utterances including MSV in total, and the diversity of MSV (column 5), which indicates the number of different MSV.

All participants with DS produced MSV. For TD children, only three of the 3 year-old participants did not produce any MSV. A high heterogeneity with respect to the number of MSV produced occurs in all groups, indicating that the observed variability in MSV production is not a specific problem of the group with DS but a typical pattern in MSV production. For TD children, the data indicate a vocabulary growth in verbs overall as well as in MSV that is particularly pronounced between the TD group aged 5 years and the group of 9 year-olds. The fact that three TD children at age three did not produce MSV underlines the start of MSV acquisition in this age range. Due to the increase of the overall number of verbs as well as of the number of MSV, the proportion of MSV in verbs shows only a slight increase over the three age groups of TD children. With respect to verb diversity, the data display a similar pattern with a pronounced increase in verb and MSV diversity between ages 5 to 9. The increase in verb diversity is supported by a Kruskal-Wallis analysis of variance that yields a significant difference between the TD3 and TD9 groups ($p = 0.020$). Due to comparable increases in verb diversity and MSV diversity, the proportion of different MSV in different verbs remains relatively stable over the three age groups.

Against this performance of the TD children, the group of participants with DS scores somewhat better in verb and MSV density and diversity measures than the group of 3 and 5 year old TD children, but scores below the group of nine-year-old TD children. Given a mean mental age of 5; 03 years in the tested participants with DS, their performance with respect to the density and diversity of MSV use thus seems adequate for their mental age. However, performance for the group of participants with DS is clearly below the expectations for their chronological age as the group performs below the group of TD9 children with respect to MSV density and diversity, despite a higher chronological age of the individuals with DS.

Types of MSV

MSV were assigned to three categories: volition, cognition and emotion. **Table 4** shows the numbers and proportions of MSV

TABLE 3 | Enumeration of verbs and MSV in narrations over group with DS and TD groups as well as group comparison (pairwise) for all groups using a Kruskal-Wallis variance analysis.

	1: *n* of verbs	2: *n* of MSV (Density)	3: Proportion of MSV in verbs	4: *n* of different verbs	5: *n* of different MSV (=Diversity)	6: Proportion of diff. MSV in diff. verbs
Group DS (*n* = 28)						
Total	1934	196		791	102	
Mean	69.07	7.00	0.11	28.25	3.64	0.13
SD	30.57	4.17	0.04	10.65	2.25	0.05
Range	5–122	1–20	0.06–0.25	4–48	1–10	0.06–0.25
Group TD3 (*n* = 12)						
Total	762	73		255	33	
Mean	63.50	6.08	0.08	21.25*	2.75	0.12
SD	34.05	6.80	0.08	7.72	2.49	0.10
Range	25–146	0–20	0.00–0.26	15–44	0–7	0.00–0.30
Group TD5 (*n* = 11)						
Total	681	69		252	27	
Mean	61.91	6.27	0.09	22.92	2.45	0.10
SD	23.97	6.47	0.06	5.32	2.34	0.08
Range	29–102	1–22	0.01–0.22	17–31	1–7	0.03–0.23
Group TD9 (*n* = 10)						
Total	769	82		324	45	
Mean	76.90	8.20	0.10	32.40*	4.50	0.13
SD	32.69	5.45	0.04	9.30	2.32	0.04
Range	51–156	3–19	0.04–0.14	23–56	2–10	0.09–0.19
Group comparison (Kruskal-Wallis)	n.s.	n.s.	n.s.	*p = 0.020 for TD3 vs. TD9	n.s.	n.s.

TABLE 4 | Enumeration and proportion of MSV types (volition, cognition and emotion) over age groups TD3, TD5 and TD9 and group DS.

Category (number / proportion)	Group			
	TD3 (*n* = 12)	**TD5 (*n* = 11)**	**TD9 (*n* = 10)**	**DS (*n* = 28)**
Volition	26 / 0.36	26 / 0.38	10 / 0.12	41 / 0.21
Cognition	43 / 0.59	38 / 0.55	53 / 0.65	130 / 0.66
Emotion	4 / 0.05	5 / 0.07	19 / 0.23	25 / 0.13
Total	73 / 1.0	69 / 1.0	82 / 1.0	196 / 1.0

for each category of MSV and participant group. All three types of MSV were produced in each group. Cognition verbs were the most frequently produced type of MSV in all groups (mean proportion between 0.55 and 0.66). Emotion verbs increase over age in the TD groups (0.05–0.23). Volition verbs are at a high level in the TD3 and TD5 groups (0.36/0.38), whereas the TD9 group and participants with DS produce less volition verbs (0.12/0.21). A group comparison using a Kruskal Wallis variance analysis was significant for emotion verbs (p = 0.032) over all four groups. However, none of the pairwise comparisons was significant, although a comparison between the TD5 and the TD9 group approached significance (p = 0.052). Over all groups, *want* was the most frequent volition verb (e.g. *Der Junge möchte auf den Fels klettern.* "The boy wants to climb on the hill."), *say/call* was the most frequent cognition verb (e.g. *Der Junge ruft: "Frosch, wo bist du?"* "The boy calls: Frog, where are you?") and *laugh* was the most frequent emotion verb (e.g. *Der Hirsch lacht über den Jungen.* "The deer laughs at the boy.").

Syntactic Context of MSV

The syntactic context of MSV use was analyzed for all TD groups and the DS group. For each participant group, **Figure 1** shows proportions of each syntactic context a MSV appeared in. All participant groups produced MSV verbs in each of the five categorized syntactic contexts, however proportions differed. In the TD children, we see that with increasing age fewer MSV are produced in main clauses while the proportion of MSV that are accompanied by a sentential complement in a subordinate clause increases substantially from 0.04 in the group of TD3 children to 0.21 in the TD9 group. A smaller increase can also be seen for MSV in main clauses + direct speech. MSV in sentence fragments have disappeared in the group of TD9 children. Likewise, MSV in *wh*-clauses or subordinate clauses decrease substantially from the TD3 group to TD9. While the former findings are well in line with a progredient development of syntactic abilities from age three to age nine, this latter observation (decrease of MSV in complex clauses) is unaccounted for by this explanation.

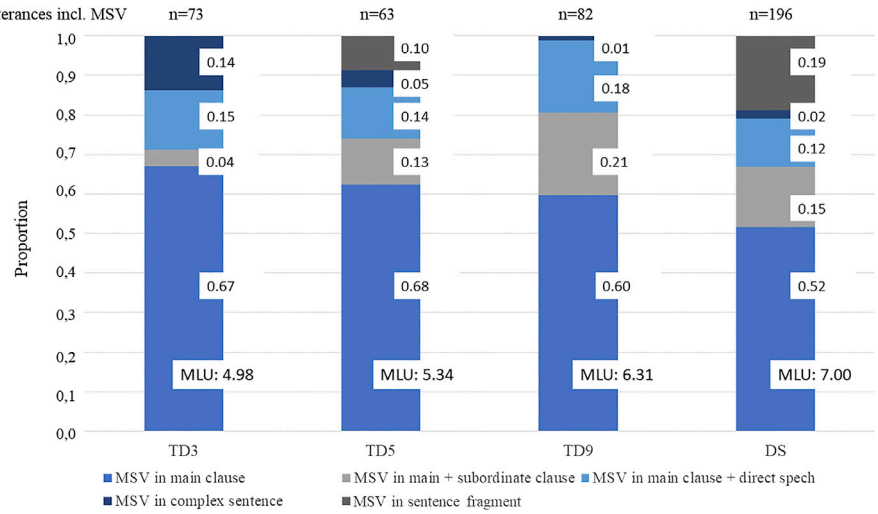


FIGURE 1 | Syntactic context of MSV over all groups: TD3 = 3 year-olds (*n* = 12), TD5 = 5 year-olds (*n* = 11), TD9 = 9 year-olds (*n* = 10), DS = participants with Down syndrome (*n* = 28).

As in the TD groups, participants with DS produce MSV in all five types of syntactic constructions. Note particularly the relatively high proportion of MSV that are accompanied by a sentential complement in a subordinate clause (0.15 for the group of participants with DS compared to 0.04 in TD3 and 0.13 in TD5). Comparable proportions of MSV use between TD children and the group of participants with DS can also be observed for MSV appearing in main clauses and for MSV accompanied by direct speech. A strong difference to the TD groups, however, holds with respect to MSV in sentence fragments. Here the proportion is particularly high in the group of participants with DS while MSV in sentence fragments do not occur in the data of TD3 and TD9 children. The high proportion of MSV in sentence fragments in participants with DS most likely reflects impairments in syntactic development which are frequently observed in individuals with DS.

The MLU in words shows a clear growth over the TD age groups, from a mean MLU of 4.98 in TD3 (*SD* 0.51) and 5.34 (*SD* 0.61) in TD5 to a mean MLU of 6.31 in the group of TD9 children (*SD* 0.51), reflecting progress in syntactic development. Although the MLU is even higher for the group of participants with DS (mean MLU 7.01, *SD* 2.94), this value has to be interpreted with caution as it is influenced by the frequent production of incomplete sentence fragments that are conjoined within one utterance (e.g. *Wir wissen noch nicht was sind die beiden was meint*. "We do not know yet what the two are what means."), leading to long but syntactically incoherent utterances.

Language Performance of Participants with DS

Table 5 shows the performance of the participants with DS in the standardized expressive vocabulary test (see 2.2.5) and the two syntax measures of the ESGRAF (see 2.2.4). Mean score in raw points was 49.71 (*SD* 13.44) for expressive vocabulary. Mean T-score was 52.06 (*SD* 15.09). Participants achieved a higher raw score in the syntax measure targeting the production of *wh*-clauses and topicalized main clauses than in the subtest targeting the production of subordinate clauses. This, however, does not reflect a better performance but is due to the fact that more points could be obtained in the first subtest of the ESGRAF compared to the second. Consequently, performance in T-scores was comparable over the two measures and suggests impaired syntax production (mean T-score below 40) in both measures.

Associations of Cognitive Development, Language Measures and MSV

To explore whether our measures of perspective taking are connected to age, cognitive development or language performance in individuals with DS, we computed Pearson's correlations (see **Table 6**). Diversity, which means the number of different MSV, correlated with density (number of utterances including MSV) at a very high level ($r = 0.877$, $p < 0.01$). The correlation between diversity and MLU showed a significance with a medium effect size ($r = 0.301$, $p < 0.05$). Correlations were not significant for diversity, density and the language measures targeting expressive vocabulary and the production of complex syntactic structures.

DISCUSSION

The main aim of our study was to describe perspective taking as reflected by MSV production in individuals with DS and to compare their performance to MSV use of TD children at (pre-)school age. Furthermore, we wanted to evaluate the influence of language abilities on MSV use, specifically of abilities related to complex syntax and expressive vocabulary which have been discussed as decisive factors in the literature. We

TABLE 5 | Raw scores and T-scores for standardized vocabulary (*AWST-R*) and syntax measures (*ESGRAF 4–8*) in participants with DS.

	Expressive vocabulary		Syntax *wh*-questions and topicalized main clauses		Syntax subordinate clauses	
	Raw points (max. 75)	T-scores	Raw points (max. 36)	T-scores	Raw points (max. 20)	T-scores
Mean	49.71	52.06	30.26	38.11	12.56	38.08
SD	13.44	15.09	12.37	7.06	5.48	10.60
Range	12–64	22–75	4–44	27–49	0–19	15–54

Note: Norming data (T-scores) are given for all participants in ESGRAF 4-8 *and for* n = *17 participants in* AWST-R.

TABLE 6 | Correlations for Density and Diversity with chronological age (CA) and nonverbal mental age (MA) (in months), MLU, and raw scores of syntax and vocabulary measures (*r*-values) for the DS group (*n* = 28).

	CA	MA	Density	Diversity	Raw score Main Clauses	Raw score Subord. clauses	Raw score expressive vocabulary	MLU
Density (*n* utterances incl. MSV)	0.296	0.346		0.877**	0.278	0.127	0.347	0.171
Diversity (*n* diff. MSV)	0.351	0.359	0.877**		0.324	0.246	0.370	0.301*

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

analyzed number, types and syntactic context of MSV in frog-story narrations of 28 children and adolescents with DS as well as 33 children aged 3, 5 and 9 years. Our investigation is the first study to assess syntactic context of MSV production systematically in individuals with DS and to include specific measures of syntactic abilities that have been discussed as relevant for MSV production. Our investigation yielded four main results, which are discussed in the following: [i] Density and diversity of MSV were comparable for individuals with DS and TD. [ii] Types of MSV produced are similar for the TD groups and the DS group. [iii] A syntactic analysis shows similarities with respect to the syntactic contexts in which MSV are used between individuals with DS and children with TD. [iv] MSV production in individuals with DS seems independent from language measures related to vocabulary development and the acquisition of complex syntactic constructions.

Number and Proportion of MSV

Statistical comparisons yielded no significant differences between the groups of TD children and the group of participants with DS concerning number and proportion of verbs and MSV. Density and diversity of MSV were numerically comparable to the data of the TD3 and TD5 groups but lower than the values of the TD9 group. In particular, the similarity in the production of MSV in individuals with DS and children aged five (TD5) suggests that the performance in MSV production in individuals with DS is appropriate for their cognitive developmental level. The lower numeric values in density and diversity in the DS group in comparison to the TD9 group, however, indicate a performance that is not appropriate for their chronological age, given that the DS group shows a considerably higher chronological age (DS mean: 14; 05) and thus individuals had more time to expand their vocabulary, including MSV. The finding of mental-age adequate performance that is, however, not commensurate to chronological age is well in line with former investigations on expressive vocabulary in individuals with DS (see meta-analysis of Næss et al., 2011, and Witecy and Penke, 2019, for German individuals with DS) and gives no indication that MSV would pose a particular problem to individuals with DS.

Our findings are consistent with the findings of Grela (2002) who also described similarities of MSV production in a small group of seven children with DS and seven TD children matched by MLU. Our finding of mental-age appropriate production of MSV in individuals with DS is, however, in contrast to the studies of Channell (2020), who described impaired performance of MSV production in her participants with DS in a similar narration task, and Martzoukou et al. (2020), who found poorer use of internal state terms in individuals with DS than in cognition-matched TD children. While Channell did not include a control group of TD children against which performance of the DS group could be measured and evaluated as mental-age appropriate, the discrepancy to the findings of Martzoukou et al. (2020) awaits explanation. Note that Martzoukou et al. did not focus on MSV in their analysis but evaluated all sorts of internal state words including adjectives (such as *happy*, *sad* etc.) – a class of words that might be particularly vulnerable in individuals with DS (Beeghly and Cicchetti, 1997).

Types of MSV

We analyzed three different types of MSV: volition, cognition and emotion verbs. All types of verbs were produced by both populations (DS & TD) and in all TD age groups. Statistical comparisons yielded no significant differences in pairwise comparisons between the groups but suggested an age effect in emotion verbs. This finding underlines a comparable quality of MSV use between individuals with DS and TD children. The increase of emotion verbs in TD children suggests that emotion verbs might be acquired later than volition and cognition verbs.

This seems plausible as the description of another person's emotions requires a higher stage of sensitivity for others. This is a flaw of younger children who have been said to be relatively egocentric during development until the end of Piaget's preoperational phase at the age of seven (see Heo et al., 2011 for an overview). Hence, school children might be more sensible for a protagonist's emotions in storytelling, leading to a higher production of emotion verbs.

Syntactic Context of MSV

We furthermore explored five different syntactic contexts of MSV use: [i] in a main clause, [ii] in a main clause followed by a subordinate clause, [iii] in a main clause followed by direct speech, [iv] in a complex sentence (e.g. subordinate clause or question), [v] in a sentence fragment (see **Table 2**). Interestingly, MSV were produced in all five syntactic contexts by both populations (TD and DS) and over all TD age groups. This indicates that individuals with DS do not differ substantially in MSV use from TD children and are able to produce MSV in similar syntactic structures, including those where the MSV is accompanied by a sentential complement. Nevertheless, our participants with DS show a significantly higher number of MSV in sentence fragments than TD children, reflecting more pronounced difficulties of individuals with DS to use a MSV in a syntactically complex utterance. This observation conforms to the performance of the participants with DS in the two subtests of the *ESGRAF 4-8* which also indicate a performance with respect to the production of complex syntactic structures that is below mental-age expectations and, thus, indicative of a syntactic impairment. Against this background of an impaired syntactic development in individuals with DS, a striking finding of our investigation is that MSV production is possible despite developmental deficits in syntax and that it is not dependent on specific complex syntactic structures but that MSV can successfully be produced even in sentence fragments.

Our findings indicate that MSV are produced in various syntactic contexts already by young TD children and in individuals with DS who are impaired with respect to their syntactic development. This attenuates the assumption that use of MSV might be dependent on the acquisition of complex syntactic structures related to sentence complementation. De Villiers and Pyers (2002) have suggested that the understanding of another person's mind hinges on the acquisition of sentential complementation. The ability to produce complex sentence structures also allows complex cognitive inferences to the mental state of others. However, even the youngest age group of TD3 children and the syntactically impaired participants with DS produced MSV in a variety of syntactic structures, arguing against a direct association between MSV use and the acquisition of sentential complementation. Our findings are a first indication that syntactic abilities in individuals with DS are not decisive for their MSV production and perspective taking skills.

Relations Between MSV Production and Language Variables

To evaluate the relationship between language factors (complex syntax as well as expressive vocabulary) and MSV production in individuals with DS, we computed Pearson's correlations for density and diversity of MSV, age (chronological as well as mental), MLU in the narrations produced and performance in three standardized measures evaluating the production of complex syntactic structures as well as expressive vocabulary.

Previous investigations had identified syntactic development measured by MLU as well as expressive vocabulary as factors related to the production of MSV in individuals with DS (Ashby et al., 2017; Channell, 2020). With respect to expressive vocabulary, we could not confirm the finding of Channell (2020) that expressive vocabulary accounts for MSV production in individuals with DS. Note that while Channell evaluated expressive vocabulary against the chronological age of her participants we assessed it against the mental age of the participants with DS. While our participants display relatively good expressive vocabulary skills according to their mental age (mean T-score 52.06), Channell's participants displayed an impaired vocabulary performance. Given the substantial difference between chronological (mean 14; 05 years) and mental age (mean 5; 03 years) in our participants with DS, it seems likely that their expressive vocabulary – although adequate for their mental age – was limited for their chronological age. Due to the lack of a standardized test spanning the mental and chronological age range of our participants, this issue could not be determined. Nevertheless, the lack of a correlation between expressive vocabulary and MSV use is in contrast to Channell's findings and requires further research.

Similar to previous investigations by Channell (2020) and Ashby et al. (2017), we found a significant correlation between diversity of MSV and MLU ($r = 0.301$, $p < 0.05$). The exact nature of this relationship is, however, difficult to evaluate. Our data suggest that a high MLU does not indicate unimpaired syntactic abilities in individuals with DS but might also come about by concatenations of syntactic fragments within one utterance. Moreover, we included a specific measure of syntactic abilities but found no significant correlation between density and diversity of MSV and syntactic abilities in the group of participants with DS. This suggests that MSV use might be relatively independent at least from more advanced syntactic abilities, a finding that is supported by the frequent use of MSV in sentence fragments. Note also that MLU values cannot be directly compared between our investigation and previous studies because MLU is typically calculated on words in German but on morphemes in English (see Beeghly and Cicchetti, 1997; Ashby et al., 2017; Channell, 2020). Future studies should directly test specific syntactic abilities rather than relying on MLU as a measure for syntactic development in evaluating the developmental link between syntax and MSV use. The lack of a correlation between a measure of complex syntax and MSV use in our study suggests that MSV use seems to be more independent from syntactic development than was previously assumed. Specifically, it does not hinge on the acquisition of subordinate or complement clauses. Our finding that the ability to express perspective taking via MSV use is not dependent on syntactic abilities is also supported by Martzoukou et al. (2020) who found morpho-syntactic abilities as measured by a sentence-repetition not to be related to the use of internal state terms.

The findings of Martzoukou et al. and the results of our investigation suggest that the ability to understand a character's internal state and to take this character's perspective might be largely independent of language abilities and seems to be unaffected in individuals with DS. Participants in the Martzoukou et al. study were able to infer the internal state of story characters, participants in our study were able to produce MSV expressing a story character's view, beliefs and desires. The syntactic impairments that are typically observed in individuals with DS might, however, limit the ability to express another character's perspective verbally. In the study by Martzoukou et al., this was reflected by a lower number of complement clauses produced by the participants with DS. The use of MSV, that we adopted as a measure of perspective taking abilities, proved to be less affected by syntactic deficits. The amount and type of produced MSV were comparable to TD children, and MSV would also surface in ungrammatical sentence fragments. However, the relatively large proportion of MSV in sentence fragments produced by our participants with DS points to the limitations of these individuals when trying to express another character's perspective.

Limitations and Outlooks

It is important to consider some limitations of our study. While the overall number of participants in the TD groups ($n = 33$) and the DS group ($n = 28$) is sufficiently high and comparable in our study, the number in each TD age group is quite low, limiting type and power of the statistical analyses that could be conducted. Another limitation of our study is that for TD children no standardized measures on cognitive and syntactic development as well as size of expressive vocabulary were available, precluding direct comparisons between the TD children and the participants with DS with respect to these measures.

Our study targeted MSV production as a measure for perspective taking, disregarding other aspects of narrations that might also be indicative of the ability to take perspective in individuals with DS, such as character introduction or the consideration of the listener's knowledge. Also, perspective taking abilities might be associated to Theory of Mind as both abilities require sensitivity for other people's intentions and emotions. As individuals with DS often also show impairments in Theory of Mind abilities (Neitzel and Penke, 2021), the connection between perspective taking in form of MSV production and Theory of Mind may deserve closer examination.

A deeper investigation of this topic should also involve a qualitative assessment of MSV use in individuals with DS. Our study focused on the use and syntactic context of MSV whereas the semantically-adequate expression of mental states was not evaluated. It might, thus, be the case that while individuals with DS use a mental age-appropriate number of MSV, these provide a less adequate or incorrect identification of a character's mental states. Further research should target the difference between understanding a character's mental state and perspective and the verbal expression of another character's internal state and perspective to support the view that the former is unaffected in individuals with DS whereas the latter might be limited by impaired syntactic abilities.

Previous studies have mostly analyzed internal state language, including but not limited to MSV, as a measure of inferential language in individuals with DS. While Beeghly and Cicchetti (1997), Ashby et al. (2017) as well as Martzoukou et al. (2020) described impaired internal state language in individuals with DS, we focused on MSV and found the production of MSV to be mental-age adequate. However, this does not allow general conclusions concerning strengths and weaknesses in internal state language as a generic category of mental state language in individuals with DS.

As previous investigations have found impaired narrative ability in individuals with DS (overview by Segal and Pesco, 2015), future analyses should include not only MSV production as microstructural ability but also evaluate a connection to macrostructural and overall narrative abilities in this population. These aspects might be important as perspective taking and the understanding of other person's mental states are also reflected in a story's macrostructure, e.g. by attention to internal reactions of the characters or the communication of a moral in the end of a story.

Finally, longitudinal data of individuals with DS and TD would evidently be better suited to investigate and compare MSV development and other perspective taking abilities in typically- and atypically-developing children. Longitudinal data is however difficult to obtain as longitudinal studies including individuals with cognitive disabilities face multiple challenges (Witecy and Penke, 2017): First, participant recruitment, which is a difficulty in research on individuals with DS, might fail for long-term studies due to unpredictable health development and other factors. Second, drop-out of participants is more likely in less abled participants (Carr, 2005) which might distort the findings. This is especially problematic for individuals with DS since the slow progress in cognitive development demands observation over a long period. Finally, multiple testing is in danger to affect the outcome.

CONCLUSION

In our investigation, we focused on perspective taking in the form of MSV production in individuals with DS. Previous investigations have provided indications that MSV production might be impaired in this population and suggested that these difficulties might be related to expressive vocabulary or impaired syntactic abilities. We have compared MSV production in narrations of individuals with DS and three groups of TD children aged three, five and nine years. Our analyses indicate that MSV production is adequate for MA in our participants with DS regarding density and diversity of MSV as well as the syntactic context MSV appear in. Despite previous suggestions that MSV use might be dependent on syntactic development, specifically on the ability to construct sentential complements, we have not found interdependencies between MSV production and complex syntactic abilities. Rather, the production of MSV seemed to be quite independent of specific syntactic structures and MSV even occurred in sentence fragments in our participants with DS. Our investigation indicates that individuals with DS have mental age-

appropriate abilities of perspective taking in form of MSV production and that MSV can be produced even when syntactic abilities are impaired. These findings lead to a better understanding of the connection between MSV production and language performance in general. Our findings suggest that MSV production constitutes a measure of perspective taking abilities that is relatively unaffected by syntactic abilities and therefore even applicable for individuals with syntactic impairments such as individuals with DS. Further research on other measures of perspective taking is necessary to corroborate the assumption that perspective taking per se is unaffected in individuals with DS, while the ability to verbally express another character's perspective might be limited by syntactic impairments.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors on demand.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Medical Department of the University of Cologne. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

IN collected the data and developed the concept for this paper. This concept work was supported by MP. IN computed all analyses which were controlled by and discussed with MP. This paper was written by IN and revised by MP.

ACKNOWLEDGMENTS

We would like to thank all participants and their families as well as all institutions and networks that were involved in the recruitment of participants. Furthermore, we would like to thank Dr. Steffen Neitzel-Grieshammer for support in data processing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.629757/full#supplementary-material

REFERENCES

Abbeduto, L., Warren, S. F., and Conners, F. A. (2007). Language Development in Down Syndrome: From the Prelinguistic Period to the Acquisition of Literacy. *Ment. Retard. Dev. Disabil. Res. Rev.* 13 (3), 247–261. doi:10.1002/mrdd.20158

Ashby, S. A., Channell, M. M., and Abbeduto, L. (2017). Inferential Language Use by Youth with Down Syndrome During Narration. *Res. Developmental Disabilities.*, 71, 98–108. doi:10.1016/j.ridd.2017.10.002

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the Autistic Child Have a "theory of Mind" ? *Cognition* 21 (1), 37–46. doi:10.1016/0010-0277(85)90022-8

Beaudoin, C., Leblanc, E., Gagner, C., and Beauchamp, M. (2020). Systematic Review and Inventory of Theory of Mind Measures for Young Children. *Front. Psychol.* 10, 2905. doi:10.3389/fpsyg.2019.02905

Beeghly, M., and Cicchetti, D. (1997). Talking about Self and Other: Emergence of an Internal State Lexicon in Young Children with Down Syndrome. *Dev. Psychopathol* 9, 729–748. doi:10.1017/s0954579497001417

Berman, R., and Slobin, D. I. (1994). In collaboration with Ayhan Aksu, Michael Bamberg, Virginia Marchman, Tanya Renner, Eugenia Sebastian, and Christiane von Stutterheim) in *Different Ways of Relating Events in Narrative: A Crosslinguistic Study*. Hillsdale, NJ: Erlbaum Associates.

Carr, J. (2005). Stability and Change in Cognitive Ability over the Life Span: A Comparison of Populations with and without Down's Syndrome. *J. Intellect. Disabil. Res.* 49 (12), 915–928. doi:10.1111/j.1365-2788.2005.00735.x

Channell, M. M. (2020). Cross-sectional Trajectories of Mental State Language Development in Children with Down Syndrome. *Am. J. Speech Lang. Pathol.* 29 (2), 760–775. doi:10.1044/2020_ajslp-19-00035

De Villiers, J. G., and Pyers, J. E. (2002). Complements to Cognition: A Longitudinal Study of the Relationship between Complex Syntax and False-Belief-Understanding. *Cogn. Development.* 17, 1037–1060. doi:10.1016/s0885-2014(02)00073-4

Devine, R., and Hughes, C. (2014). Relations between False Belief Understanding and Executive Function in Early Childhood: a Meta-Analysis. *Child Dev.* 85 (5), 1777–1794. doi:10.1111/cdev.12237

Grela, B. G. (2002). Lexical Verb Diversity in Children with Down Syndrome. *Clin. Linguistics Phonetics.* 16 (4), 251–263. doi:10.1080/02699200210131987

Griffin, T. M., Hemphill, L., Camp, L., and Wolf, D. P. (2004). Oral Discourse in the Preschool Years and Later Literacy Skills. *First Lang.* 24 (2), 123–147. doi:10.1177/0142723704042369

Heo, J., Han, S., Koch, C., and Aydin, H. (2011). Piaget's Egocentrism and Language Learning: Language Egocentrism (LE) and Language Differentiation (LD). *J. Lang. Teach. Res.* 2 (4), 733–739. doi:10.4304/jltr.2.4.733-739

IBM (2017). *IBM SPSS Statistics for Windows*, Version 25.0. Armonk, NY: IBM Corp.

Jenkins, J. M., Turrell, S. L., Kogushi, Y., Lollis, S., and Ross, H. S. (2003). A Longitudinal Investigation of the Dynamics of Mental State Talk in Families. *Child Development* 74, 905–920. doi:10.1111/1467-8624.00575

Kiese-Himmel, C. (2005). AWST-R. Aktiver Wortschatztest für 3- bis 5-jährige Kinder. Göttingen: Hogrefe.

Kim, Y.-S. (2015). Language and Cognitive Predictors of Text Comprehension: Evidence from Multivariate Analysis. *Child Dev.* 86, 128–144. doi:10.1111/cdev.12293

Martzoukou, M., Nousia, A., and Marinis, T. (2020). Narrative Abilities of Adults' with Down Syndrome as a Window to Their Morphosyntactic, Socio-Cognitive, and Prosodic Abilities. *Front. Psychol.* 11, 2060. doi:10.3389/fpsyg.2020.02060

Mayer, M. (2003). *Frog, where Are You? (A Boy, a Dog, and a Frog)*. New York: Dial books.

Max Planck Institute (2018). *Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive*. Available at: https://archive.mpi.nl/tla/elan.

Motsch, H., and Rietz, C. (2016). *ESGRAF 4-8. Grammatiktest für 4- bis 8-jährige Kinder*. München: Reinhardt.

Næss, K. A., Laster, S. A., Hulme, C., and Melby-Lervåg, M. (2011). Language and Verbal Short-Term Memory Skills in Children with Down Syndrome: A Meta-Analytic Review. *Res. Developmental Disabilities.*, 32 (6), 2225–2234. doi:10.1016/j.ridd.2011.05.014

Neitzel, I., and Penke, M. (2021). Theory of Mind in Children and Adolescents with Down Syndrome. *Res. Developmental Disabilities.*, 113, 103945. doi:10.1016/j.ridd.2021.103945

Nordqvist, A. (2001). The Use of Direct and Indirect Speech by 1 ½-to 4-Year-Olds. *Psychol. Lang. Commun.* 5 (1), 57–66.

O'Neill, D., Pearce, M., and Pick, L. (2004). Preschool Children's Narratives and Performance on the Peabody Individualized Achievement Test – Revised: Evidence of a Relation between Early Narrative and Later Mathematical Ability. *First Lang.* 24 (2), 149–183. doi:10.1177/0142723704043529

Perner, J., Sprung, M., Zauner, P., and Haider, H. (2003). Want That is Understood Well before Say That, Think That, and False Belief: A Test of de Villiers's Linguistic Determinism on German-Speaking Children. *Child Development* 74 (1), 179–188. doi:10.1111/1467-8624.t01-1-00529

Segal, A., and Pesco, D. (2015). Narrative Skills of Youth with Down Syndrome: a Comprehensive Literature Review. *J. Dev. Phys. Disabil.* 27, 721–743. doi:10.1007/s10882-015-9441-5

Symons, D. K. (2004). Mental State Discourse, Theory of Mind, and the Internalization of Self-Other Understanding. *Developmental Rev.* 24 (2), 159–188. doi:10.1016/j.dr.2004.03.001

Tellegen, P. J., Laros, J. A., and Petermann, F. (2007). *SON-R 2½-7. Non-verbaler Intelligenztest*. Göttingen: Hogrefe.

Tompkins, V., Farrar, M. J., and Montgomery, D. E. (2019). Speaking Your Mind: Language and Narrative in Young Children's Theory of Mind Development. *Adv. Child Development Behav.* 56, 109–140. doi:10.1016/bs.acdb.2018.11.003

Tompkins, V., Guo, Y., and Justine, L. M. (2013). Inference Generation, Story Comprehension, and Language Skills in the Preschool Years. *Read Writ.* 26, 403–429. doi:10.1007/s11145-012-9374-7

Van Krieken, K., Hoeken, H., and Sanders, J. (2017). Evoking and Measuring Identification with Narrative Characters–A Linguistic Cues Framework. *Front. Psychol.* 8, 1190. doi:10.3389/fpsyg.2017.01190

Wimmer, E., Witecy, B., and Penke, M. (2020). "Syntactic Problems in German Individuals with Down Syndrome: Evidence from the Question of Wh-Questions," in *New Trends in Language Acquisition within the Generative Perspective*. Editors P. Guijarro Fuentes and C. Suarez Gomez (Basel: Springer), 151–163.

Witecy, B., and Penke, M. (2017). Language Comprehension in Children, Adolescents, and Adults with Down Syndrome. *Res. Developmental Disabilities.* 62, 184–196. doi:10.1016/j.ridd.2017.01.014

Witecy, B., and Penke, M. (2019). Lexical-semantic Abilities of Adults with Down Syndrome. *Logos* 27 (2), 84–95. doi:10.7345/prolog-1902084

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

frontiers in Communication

ORIGINAL RESEARCH
published: 14 May 2021
doi: 10.3389/fcomm.2021.627144

Egophoricity and Perspective: A View From Spoken Swedish

*Henrik Bergqvist**

Department of Linguistics, Stockholm University, Stockholm, Sweden

From the point of view of everyday talk and especially, casual conversation, it is obvious that language use is highly perspectivized with a clear focus on the speech-act participants. This fact is supported by observations regarding the pervasiveness of egophoric pronouns and the frequent use of the modal particles *ju* and *väl* in spoken Swedish. The paper demonstrates how egophoric pronouns, modal particles, and mental verbs are used to signal the epistemic perspective of the speech-act participants, i.e., when the knowledge and attention of the speech-act participants are at stake. These formally distinct resources show patterns of co-distribution that permit an analysis of forms in terms of how they signal shared/private access to events from the perspective of the speaker and the addressee.

Keywords: egophoricity, epistemic marking, dialogicity, Swedish, modal particles

OPEN ACCESS

Edited by:
Petra Hendriks,
University of Groningen, Netherlands

Reviewed by:
Lotte Hogeweg,
Radboud University Nijmegen, Netherlands
Elisabeth Engberg-Pedersen,
University of Copenhagen, Denmark

***Correspondence:**
Henrik Bergqvist
henrik.bergqvist@ling.su.se

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *08 November 2020*
Accepted: *07 April 2021*
Published: *14 May 2021*

Citation:
Bergqvist H (2021) Egophoricity and Perspective: A View From Spoken Swedish. Front. Commun. 6:627144. doi: 10.3389/fcomm.2021.627144

INTRODUCTION

Egophoric contexts target the perspective of the speaking/to-be speaking subjects, and in such contexts, grammatical constructs like sentence-type, pronouns, tense-aspect marking, epistemic modals, and adverbs are used to position the speech-act participants with respect to talked-about events. This paper demonstrates how a subset of such grammatical resources are used for signaling differences and overlaps in the epistemic perspective of the discourse participants. These formally distinct devices signal different aspects of epistemic perspective in Swedish grammar and may be compared to recent accounts of epistemic marking systems in the languages of the world (e.g., Bergqvist and Kittilä, 2020; Grzech et al., 2020). The term perspective is used with reference to the notion of "multiple perspective" defined by Evans (2006), which permits a separation of epistemic and non-epistemic parameters in the configuration of different perspectives involving the speech-act participants and third parties (see Evans et al., 2018, for a discussion; cf. Bergqvist, 2015). As such, the term perspective is used in a technical sense that separates it from partly overlapping terms like viewpoint and stance (see section Status, Stance, and Territories of Information, below)[1].

The main argument of the present paper is that traditional categories such as pronouns, modal particles, mental verbs, and sentence-type distinctions are prominent resources for signaling the epistemic perspective of the speech-act participants and that this is done, in unnoticed ways. When an analysis of such forms is based on how they are used in spoken discourse, it becomes evident that epistemic function determines their distribution in non-random ways. Consider Example (1):

[1]Multiple perspective consists of (grammatical) constructions that "encode potentially distinct values, on a single semantic dimension, that reflect two or more distinct perspectives or points of reference" (Evans, 2006, p. 99).

(1) Swedish

A: *ja vi bara går vi tar inte bilen*
yes we just go we take not the.car
"Yeah, we'll just go, we won't take the car."

B: *jo men det kan vi väl*
AGR but it can we MP.ADR
göra jag ska ju ha fatt
do 1S will MP.SPKR have get
i plantor till den blivande
in plants to the becoming
trädgårdsängen på
the.plant.bed at
i trädgården eller jag
in the.garden or 1S
menar blomsterängen i trädgårn
mean the.flower.bed in the.garden
"OK, but why don't we, I'm looking for plants for the planned plant bed at, in the garden, or, I mean the flower bed in the garden."

A: *kan vi göra allt på en gång[?]*
can we do everything at one time
"Can we do everything at once?"

B: *ja varför skulle vi inte kunna*
yes why should we not know
göra det
do it
du vet det är inte lång
2S know it is not long
tid innan blåklockorna försvinner […]
time before bluebells disappear
"Yeah, why shouldn't we be able to do that? You know it's not long before the bluebells will disappear […]"

Example (1) illustrates the dialogical characteristics of spoken discourse, *viz.* a shared context, a high level of interactiveness, and personal expressiveness (see Leech, 1998; Leech and Svartvik, 2002). These are reflected in the occurrence of egophoric pronouns (*jag,* 1S; *du,* 2S; *vi,*1P), question-answer pairs, modal particles (*väl, ju*), and formulaic expressions targeting the perspectives of the speaker and the addressee (*jag menar*, "I mean," *du vet* "you know"). The exchanges align the perspectives of the speech-act participants by mitigating differences of opinion and estimations of possibility/probability; it is a cooperative endeavor, whose component parts cannot be analyzed by focusing on isolated utterances in absence of discourse context. Example (1) does not constitute an exceptional instance of language use but is typical of casual talk between acquainted people, who talk for the sake of talking.

A prevalent feature of such discourse is "dialogic parallelism" (Du Bois, 2014), which signals the simultaneous perspectives of the speech-act participants and their respective positions vis-à-vis talked-about events, across turns. An example of a dialogic parallelism is in Example (2):

(2) B: *det känns nog inte som*
it feel.PRS MP.nog NEG like
du har föreställt dej
2S have imagine.PRF 2O
*det **tror** **jag** inte*
it think 1S NEG
"It doesn't feel like you have imagined it, I don't think."

A: ***tror** **du** inte det*
think 2S NEG it
"You don't think so?"

B: *nä man kan nog aldrig fatta*
no one can MP.nog never understand
hur det känns egentligen
how it feels really
"No, you could probably never understand how it feels, really."

The first position evaluation of Speaker B is qualified by the phrase *det tror jag inte* ("that, I don't think"), which is paralleled in the subsequent turn by Speaker A, who utters *tror du inte det?* ("you don't think so?"). Dialogic parallelisms have bearing on the analysis of how epistemic forms (e.g., *tro*) are distributed in spoken discourse and how they refer to the epistemic perspectives of the speaker and the addressee. They are introduced in section The Dialogic Nature of Grammar along with the related notion of "diagraph" (Du Bois, 2014, p. 362).

The present paper explores the perspectivizing function of pronouns, modal particles, mental verbs, and sentence-type in egophoric contexts and argues for the importance of dialogicity for their description. It seeks to address the epistemic function of forms from the point of view of how knowledge is negotiated in discourse, in contrast to traditional descriptions of such forms, which focus on the mental representation of events of a solitary speaker. This paper shows that contexts involving the speech-act participants may be formally differentiated depending on which participant is in focus. Second-person contexts predominantly feature non-declarative clauses, the modal particle *väl* ("right?"), and formulaic expressions that mitigate statements that affect the addressee, as well as expressions that prompt the addressee to consider and sometimes agree with the point of view of the speaker (*du vet*). By contrast, first-person contexts are characterized by the predominance of declarative clauses, the modal particle *ju* ("of course"/"obviously"), and expressions that mitigate/attenuate and qualify statements that involve the speaker.

An original contribution of the present paper lies in demonstrating the distribution and frequency of epistemic forms (i.e., modal particles and mental verbs) as dependent on perspectivizing, indexical forms like egophoric pronouns and sentence-type distinctions (see section Perspectivizing Constructs and Their Distribution, below). Such patterns of distribution and frequency are analyzed from the functionality and meaning of the investigated forms, as portrayed in the literature (see Dahl, 2000; Bergqvist, 2020). The presentation of the results is qualitatively formulated and void of any advanced statistical modeling since the

corpus data and the ensuing analysis must be regarded as explorative and preliminary with respect to the role of the observed patterns for a more detailed analysis of the studied forms.

The background for the ensuing data analysis details distinct notions such as dialogicity, stance, and egophoricity (Dahl, 2000; cf. San Roque et al., 2018), which serve as a theoretical backdrop for analyzing the central characteristics of the investigated forms. The following sections go on to explore egophoricity in spoken Swedish with special attention to the co-distribution of subject pronouns, modal particles, mental verbs, and sentence-type distinctions. These diverse resources signal the shared and private perspectives of the speech-act participants in distinct ways, as discussed in the concluding section of the paper.

THE DIALOGIC NATURE OF GRAMMAR

While the dialogic characteristics of language have been largely overlooked in descriptive, typologically oriented, linguistic research during the last half-century, early (modern) grammarians like Otto Jespersen have emphasized this aspect of language. He said, "The essence of language is human activity—activity on the part of one individual to make himself understood by another, activity on the part of that other to understand what was in the mind of the first…" (Jespersen, 1924, p. 17). Although contemporary, functionalist theories of grammar such as Role and Reference Grammar (Van Valin and LaPolla, 1997) and Functional Discourse Grammar (Hengeveld and Mackenzie, 2008) build on similar insights, typological-descriptive linguistics does not emphasize the importance of dialogic aspects of grammar. Additionally, the well-attested under-specification of meaning in linguistic form has had surprisingly little impact on descriptivist-typological research, possibly because taking it at face value would require more attention to be paid to pragmatics, which is a field of study with a partly different research agenda from mainstream, descriptive linguistics. Since a large part of the context relevant to an utterance rests on interactional and socio-cultural criteria, paying attention to the pragmatics of the utterance means paying attention to the speaker and the addressee. These two main players occupy inter-changeable roles, something that is reflected everywhere in grammar and in language use (see e.g., Couper-Kuhlen and Selting, 2018). These shortcomings of traditional descriptive-typological research have motivated John Du Bois and colleagues to formulate the theory of "Dialogic Syntax" (e.g., Du Bois, 2014) in order to explicitly account for dialogic structures within current functionalist investigations of grammar and language. Du Bois (2014) explores dialogicity in syntactic analysis, focusing on dialogic parallelisms and the partial reproduction of utterances by a previous speaker as supra-sentential structures that a fully accountable syntactic theory should include. He uses the following example to illustrate such parallelisms (Du Bois, 2014, p. 361):

(3) A: *It's kind of like you, Ken!*
B: *That's not at all like me, Joanne.*

Example (3) illustrates a dialogic parallelism by means of a partial reproduction by Speaker B of the previous utterance of Speaker A. The repetition of *like* in the utterance of B along with the co-referential function of pronouns (i.e., *you* and *me* refer to the same participant in the exchange) are the component parts of a diagraph (Du Bois, 2014, p. 362, see also Du Bois, 2007), which includes both positions of the turn (i.e., first and second positions). In the South Papuan language Nen, Evans (2012) accounts for the "assentive," which is a grammaticalized form expressing assent to a foregoing imperative utterance. Evans draws parallels to question-answer pairs, interrogative-demonstrative exchanges, and egophoric marking (see section Egophoricity and Epistemic Authority, below) in order to situate this phenomenon[2]. Example (4) shows how the assentive (zero) morpheme occurs in "perfective, singular, positive responses to imperatives" (Evans, 2012, p. 176):

Nen
(4) A: *Bm ombte nu t-z-ø-ø*
2S hot water 3S-cook-PFV.IMP-2S
"Boil the hot water!" (put it over the fire, boil it from the start).
B: *Ẽ, bä d-z-ø*
yes IMM.FUT 3S.-cook-PFV.1S.ASS
"Yes, I'll boil it." (from scratch) (Evans, 2012, p. 175, my adjusted glossing).

Assentives cannot be accounted for without making explicit reference to the preceding utterance, which makes up part of the immediate context. Assentives in Nen exemplify grammaticalized diagraphs, as defined by Du Bois, and Evans argues that typologically oriented research on morpho-syntax should include attention to structures of dialogic coordination in order to be able to capture similar phenomena. Assentives in Nen and parallelisms in American English are examples of distinct cross-turn structures and the present paper aims to contribute to the investigation of dialogicity in grammar by accounting for egophoricity in spoken Swedish in terms of the frequency and distribution of egophoric pronouns, modal particles, and sentence-type. These resources display clear dialogical functions that motivate comparisons to e.g., egophoric marking, as one instance of a diagraph.

STATUS, STANCE, AND TERRITORIES OF INFORMATION

"Stance" is a term that has figured in the literature on modality and evidentiality (e.g., Biber and Finegan, 1989; Mushin, 2001; Cornillie, 2009; *inter alia*). In these contexts, stance is used to target the motivations speakers have for using certain epistemic forms, signaling the attitude of the speakers toward a talked-about event. In the context of the present paper, the stance concept provides an interaction-focused framework for analyzing distinct linguistic resources

[2]Evans calls egophoric marking "conjunct/disjunct", which is a commonly used label for referring to such systems in the literature (e.g., Hale, 1980).

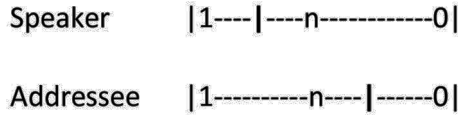


FIGURE 1 | Illustration of the territory of information of the speaker (after Kamio, 1997, p. 17).

in spoken Swedish as primarily belonging to the perspective of one of the speech-act participants. The contents of this section motivate the subsequent analysis of forms as signaling the relative placement of epistemic authority with the speaker and/or the addressee.

While stance has been regarded, by many, as a vague concept with too many uses and no clear differentiation from related concepts, such as "viewpoint" and "subjectivity" (e.g., Kockelman, 2004), recent developments in conversation analytical research by John Heritage and colleagues (e.g., Heritage and Raymond, 2005; Heritage, 2012, 2013) have moved toward an operationalization of the stance concept to focus on how speakers position themselves with respect to the allocation of epistemic authority in talked-about events (cf. Goffman, 1981 and the notion of "principal" as a speaker-role). A central issue for Heritage (2012) is accounting for the relationship between sentence types vis-à-vis speech acts in spoken American English. The fact that speakers routinely use declaratives to formulate polar questions (see Stivers and Rossano, 2010) is a long-standing problem in pragmatics, and Heritage proposes a solution to this problem by introducing the two concepts: "epistemic status" and "epistemic stance." Epistemic status designates a (more or less) stable property of the relation between the speaker and certain events, such as being married, being hungry, exercising a professional occupation, etc. (see directly below for more details; cf. Labov and Fanshel, 1977). Epistemic stance is the moment-to-moment positioning performed by the speaker with respect to his/her epistemic status, as well as that of other speech-act participants (Heritage, 2012). The relationship between epistemic status and stance produces a dynamic interplay, where a speaker can present himself/herself as an authority (stance) despite not being an authority based on his/her status, and *vice versa*. Both concepts (i.e., status and stance) may constitute grounds for claiming epistemic authority, which may also be placed with the addressee, or a third party, depending on context. Although the epistemic status of a speaker may warrant authority by default if it concerns the personal domain of the speaker, the epistemic stance adopted by the same speech-act participant may align or misalign with this epistemic status.

Kamio (1997) coins the concept "territories of information" to account for the factors that determine the preference for certain sentence types as reflecting the claim of knowledge of the speaker. Kamio observes how speakers of Japanese talk about events that belong to the respective territories of the speech-act participants and demonstrates the ungrammaticality associated with utterances targeting events that are outside the territory of the speaker without signaling this in the appropriate way by means of sentence-final particles and inflections (Kamio, 1997, p. 41). Territories of information, as identified by Kamio, are (1) information obtained from internal or direct experience; (2) detailed professional knowledge/expertise; (3) reliable information that one of the speech-act participants commits to; and (4) information about persons, objects, events, and facts close to one of the speech-act participants, including personal information (Kamio, 1997, p. 39). Kamio provides a scalar, linear representation of the psychological states of the speaker and the addressee and uses the value of 1 to indicate an event as closely belonging to the speaker and the value of 0 to signal one that does not belong to the speaker. The addressee has a corresponding scale and if an event (signaled by a bar in **Figure 1**, below) is judged to be close to 1 on the scale of the speaker and close to 0 on the scale of the addressee, then such an event belongs to the territory of the speaker. The letter *n* represents a neutral point on the scale, between 1 and 0.

Figure 1 may apply to many kinds of utterances, one of which could be, *I am running a fever.* When uttering this proposition, the speaker can assume an authoritative position with respect to his/her addressee, since a bodily state like fever automatically belongs to the affected person. However, such default positioning of authority may be challenged when seeing a doctor, who by profession can override the sensations of a patient (Heritage and Raymond, 2005)[3]. Kamio considers the theory to be, in principle, applicable to any language, although different grammatical resources are used to express this territorial difference, including evidentials (Kamio, 1997, p. 173). Heritage (2012) partly builds on the theory proposed by Kamio and uses a similar model for signaling the relative knowledgeability of the speaker and the addressee. Heritage's view of epistemic territories as they pertain to his exploration of epistemic status and stance is summarized in the following quote, "epistemic territories embrace what is known, how it is known, and persons' rights and responsibilities to know it" (Heritage, 2012, p. 5). The value of the insights provided by Heritage and Kamio, for the purposes of the present study, is to bring to light the simultaneous positions of the speech-act participants as part of a (potential) dialogic exchange, thereby extending the study of epistemic positioning to also include the perspective of the addressee.

Egophoricity and Epistemic Authority

Dahl (2000) defines egophoricity as reference to the speech-act participants (i.e., speaker and addressee) in discourse. This definition subsumes first person, second person, generic (*you/one*, Swe. *man*), and logophoric pronouns[4]. Egophoric

[3]The exemplified conditions for claiming epistemic authority based on the epistemic status of the speaker, target states of some kind. Given the definition by Heritage of epistemic status as revolving around states of knowing, internal sensation, and (personal) identity, this may appear to be a definitional trait. However, the epistemic status of one of the speech-act participants does not have to rely on having knowledge, or talking about personal sphere of a person. Epistemic status can be produced by the involvement of a speaker in some event without necessarily drawing on personal standing and previous experiences, although such factors, of course, color most interactions.

[4]Logophoric pronouns signal co-reference between subjects in main and subordinate clauses, e.g., *David$_i$ said he$_i$*[logophoric] *will go to the party*, and have been attested for not only many African languages (e.g., Dimmendaal, 2001)

reference in spoken Swedish as defined by Dahl focuses on reference to the speech-act participants (speaker/addressee) in terms of argument identity, but egophoricity also houses an epistemic aspect that concerns the speech-act participants' rights to knowledge. As suggested by the theory of epistemic status and stance proposed by Heritage, the speaker is in a privileged position to make assertions about events that primarily involve the speaker. Likewise, the speaker is required to concede authority to the addressee when the involvement in an event is focused on the addressee. This alteration of authority is grammaticalized in egophoric marking, a categorical expression that has been described for a small number of languages in different parts of the world (see Creissels, 2008; San Roque et al., 2018, for an overview). Example (5) illustrates the basic distribution of the egophoric marker (long vowel, ā) in Kathmandu Newar:

Kathmandu Newar

(5) a. *Ji ana wanā*
1S there go.EGO
"I went there."

b. *Cha ana wanā lā*
2S there go.EGO INTERR
"Did you go there?"

(Hale, 1980, p. 95).

The egophoric marker is found with first-person subjects in statements and with second-person subjects in questions, marking an alignment between the syntactic subject of the clause and the speech-act participant who is charged with epistemic authority. This kind of alignment usually requires a specific kind of (interchangeable) involvement on behalf of the speaker and the addressee, notably in terms of agency, control, and voluntary action. However, there is a high degree of attested variation in the type of involvement that conditions egophoric marking (see Bergqvist and Knuchel, 2017; cf. Creissels, 2008 for a discussion). Some languages restrict egophoric marking to occur with voluntary actions performed by one of the speech-act participants (e.g., Newar, given above), while other languages permit the egophoric marker to occur with any event that the speaker has authoritative knowledge of, including events that affect the speaker, as exemplified by Example (6) from Awa Pit, where *-s* is the egophoric marker:

Awa Pit

(6) *pɨna alu kɨ-ma-ti-s*
very rain do-COMP-PST-EGO
"It rained heavily [on me]." (Curnow, 2002, p. 620).

The notion of (conscious) involvement, as a defining feature of egophoricity, encompasses not only volition but also affectedness and attitude as grounds for signaling epistemic authority. This feature is also what links egophoricity to evidentiality, where "performative" and "participatory" evidentials have been attested as constituting part of evidential paradigms in Amerindian and Papuan languages (Oswalt, 1986 for Pomoan; San Roque and Loughnane, 2012 for Papuan; cf. Bergqvist and Kittilä, 2017 for a discussion from the point of view of person marking). Using this broader conceptualization of involvement permits a comparison between disparate systems, assuming that they belong to the same functional domain (Givón, 1981, 2001), a view that is representative of the present investigation, which aims to show how epistemic aspects of egophoricity converge differently on the speaker and the addressee in spoken Swedish. The identified characteristics of face-to-face conversation (see section Introduction) serve as a guide to identify relevant properties of those parts of grammar that ground the interactions of the speaking subjects.

EGOPHORICITY IN GRAMMAR: EXAMPLES FROM SPOKEN SWEDISH

As stated in the Introduction, the main argument of the present paper is that pronouns, modal particles, mental verbs, and sentence-type distinctions index aspects of the epistemic perspective of the speech-act participants. Their use in spoken discourse is expected to display discernible patterns that may contribute to the analysis of the investigated forms and the exploration of epistemic marking in language, more generally. The speech-act participants are especially prominent in discourse (see directly below), a fact that aligns with the relevance of dialogicity for analyzing grammar. Thus, egophoricity is an obvious starting point for exploratory work on epistemic marking in discourse.

Dahl (2000) investigates the occurrence of the egophoric arguments *jag* (1S), *du* (2S), and *man* (Gen) according to predicate type and valency and finds that the majority of all animate arguments in a corpus of Swedish conversations are egophoric. With mental verbs (e.g., *veta* "know," *tänka* "think," *hoppas* "hope"), the percentage of egophoric subjects is over 80% of all animate subjects, and with transitive verbs, egophoric subjects amount to 61%. Only with copular verbs (*vara* "be," *bli* "become," *heta* "be called," *finnas* "be/exist") is the percentage below 50%. With such verbs, "allophoric" reference (i.e., third person) is more common (Dahl, 2000, p. 47). Dahl finds support for the generalizability of these percentages in data sets of spoken English and Spanish, which reflect this distribution of egophoric pronouns. The main findings recorded by Dahl are that the majority of all animate subjects in spoken Swedish are egophoric, which is a fact that is restricted to conversations in spoken Swedish (in written Swedish, the percentage of egophoric arguments is substantially lower). The percentage of egophoric arguments is higher when an argument can only be animate (e.g., with mental verbs). Dahl concludes that the distribution of egophoric arguments in conversations cannot be accounted for by drawing on notions such as topicality and/or viewpoint but simply reflects how people talk and what they talk about.

The present paper is based on a corpus of spoken Swedish called "Conversations in Gothenburg" (Sv. *Samtal i Göteborg*). The corpus consists of 497,677 words, segmented into around 50,700 lines, which approximates turns. This is the same corpus

but also in languages of the Circum-Baltic area, e.g., Eastern Vidzeme and Leivu Estonian (Wälchli, 2015).

that Dahl (2000) used to extract a sub-corpus of some 65,000 words that he called "the G corpus" (Dahl, 2000, p.41). The "Conversations in Gothenburg"-corpus was collected as part of a sociolinguistic project where speakers residing in the Gothenburg area were asked to record themselves while conversing with another person (see Löfström, 1988, for details). This method of data collection is unusual for spoken language corpora but produces highly naturalistic language use. Dahl tagged his sub-corpus of 65,000 words in order to be able to classify all arguments and predicates therein, whereas the full word corpus with half a million words used in the present investigation is not tagged and therefore does not permit a comparable precision search for relevant forms. The choice to use the "Conversations in Gothenburg" corpus was motivated not only by the ease of accessibility and previous familiarity but also by the naturalistic character of the spoken language data (see also Bergqvist, 2020).

In this section, four interrelated linguistic resources are discussed with respect to patterns of co-distribution and the egophoric properties of forms. These are subject pronouns (*jag/du/man*), modal particles (*ju/väl*), mental verbs (*tro, veta, hoppas*), and sentence-type (SV/VS, argument-predicate order). Combinations of these forms were extracted by hand with search strings of the kind, *jag tror* ("I think") and *tror jag* ("[Do] I think"), for mental predicates (according to the SV/VS alternation). Modal particles are placed after the finite verb, which means that searches for subject pronoun and modal particle combinations are slightly less straightforward (see section Modal Particles and Subject Person, below; Bergqvist, 2020, for details). While a manual search of this kind is rudimentary, it was sufficient to uncover the patterns reported in this section and in the subsequent discussion summarizing the results. For the analysis of the investigated forms, the author relied on his native speaker competence and his ability to discern the illocutionary status of utterances and their effects in the investigated corpus (see directly below).

While egophoric contexts (i.e., ones that make reference to the perspective of the speaker and/or the addressee) display distributional and frequency characteristics that may be contrasted with allophoric contexts, there are also discernible differences between the linguistic representation of the involvement of the speaker and that of the addressee. As stated in the section directly above, the speaker may directly address their own beliefs and evaluations; but the same is not true for ones that belong to the addressee. This asymmetry has direct consequences for the distribution and frequency of constructs that reflect egophoricity.

Modal Particles and Subject Person

Modal particles have been attested for all Germanic languages and display formal and semantic characteristics that separate them from discourse particles (e.g., Zimmermann, 2011). Prominent syntactic and semantic features associated with modal particles in e.g., German, are a syntactic placement in the middle field, following the subject and finite predicate of the clause, and intersubjective semantics (e.g., Abraham and Leiss, 2012). Gast (2008) investigates the German modal particles *ja, wohl, doch*, and *etwa* and proposes a systematic classification of these as constituting "a system of oppositions with pairs of minimally contrasting elements" (Gast, 2008, p. 1). Gast identifies two semanto-pragmatic parameters, namely (i) strength of assertion and (ii) consistency with the relevant context, and argues that the forms differ in terms of how they align with some aspect of the referential context, and whether the utterance constitutes a strong or weak assertion and could be viewed as a fact, or as a hypothesis (Gast, 2008, p. 5). In accordance with this analytical model, Gast argues that *ja* and *wohl* both align with the relevant context (one aspect of which is shared knowledge/access) but that they contrast in terms of strength of assertion; *ja* marks factive utterances and *wohl* signals a hypothetical assertion.

Aijmer (1977) notes for the Swedish modal particles *ju* ("as you know") and *väl* ("right?"), that the modal component contained in the term used to denote both particles is subordinate to their pragmatic function, *viz.* how an utterance should be interpreted by the addressee. Aijmer argues that the communicative function of *ju* and *väl* may be regarded as their primary meaning (Aijmer, 1977, p. 206), while also discussing the modal components of both forms, namely that *ju* marks propositions as fact and *väl* signals possibility. This characterization is comparable to the one subsequently offered by Gast for German, where the cognate particles may be contrasted in terms of strength of assertion (Aijmer, 1977, p. 207; cf. Gast, 2008, p. 5). Bergqvist (2020) picks up on the pragmatic analysis of *ju* and *väl* as signaling the epistemic authority of the speaker (*ju*) and the addressee (*väl*) in marking a proposition that is regarded as accessible to both speech-act participants (i.e., shared access/knowledge). The formal and distributional characteristics of *ju* and *väl* motivate an analysis of these forms as a closed paradigm within the (slightly) larger group of modal particles, which also includes *nog* and *nu* (see Bergqvist, 2020; cf. Lindström, 2008). In contrast to Aijmer and Gast, Bergqvist argues that the "modal" component inherent to the respective form may be analyzed in terms of "engagement" (Evans et al., 2018) rather than epistemic modality. The non-defeasible meaning feature of *ju* and *väl* can be stated in terms of shared accessibility (engagement), whereas the modal analysis of Aijmer of both forms in terms of fact/possibility is produced by the claim, or deferral of epistemic authority, according to Bergqvist. It is not the subjective attitude of the speaker toward a proposition in terms of certainty that is at stake; rather, it is the speaker's claim of knowledge of events that involve the speaker that motivates the use of *ju* and the speakers deferral to the addressee when referring to events that primarily involve the addressee, which motivates the use of *väl.* Both are exemplified in (7) and (8):

(7) *Man* *får* *ju* *också* *se* *på*
one get MP.SPKR also look at
priset
the.price
"Of course, one should also look at the price."
(Teleman et al., 1999, p. 114).

TABLE 1 | Relativized comparison between *jag/man/du* and *ju/väl* (after Bergqvist, 2020, p. 490)[a].

	***Jag* (1S)**	***Man* (GEN)**	***Du* (2S)**	**Total MPs**
Ju	312 (37.5%)	404 (49%)	113 (13.5%)	829
Väl	62 (28%)	70 (32%)	86 (39%)	218
Total EGO PRNs	374	474	199	1,047

[a]*MP stands for modal particle.*

(8) *Du* *har* *väl* *hört* *vad* *som* *hänt*
2S have MP.ADR hear.PRF what that happen.PRF
"You heard what happened, right?" (Aijmer, 1977, p. 212).

A syntactic property of Swedish modal particles is that they only occur in declarative sentences. Despite this fact, sentences with *väl* usually feature a question mark in written Swedish, in contrast to *ju,* which always is accompanied by a full stop. The syntactic status of sentences with *ju* and *väl* is identical, however, as seen in examples (7) and (8), directly above. This separates Swedish modal particles from German modal particles, where the cognate *wohl* may be used in questions, as well. We will return to the issue of implicit sentence-type in the modal particle *väl* in section Shared Perspective.

The primary function of *ju* and *väl* is, according to Bergqvist (2020), to signal the epistemic authority of the speaker and the addressee. This is visible in how they co-occur with subject pronouns in a corpus of spoken Swedish. *Ju* is predominantly used in contexts with *jag* (1S) and *man* (Generic), and *väl* is mostly found with *du* (2S). This tendency is strongest with *man* and *du* in terms of how they are co-distributed with *ju* and *väl.* **Table 1** shows the co-distribution between egophoric pronouns and *ju/väl* in terms of percentages for respective combinations.

The numbers in **Table 1** should be viewed against the total number of egophoric pronouns in the corpus, where *jag* (13,768) is roughly twice as common as *du* (7,248) and almost three times as common as *man* (4,805). Out of all the combinations of *ju/väl* with egophoric pronouns, *man* stands out with 49% of all occurrences with *ju* and 32% with *väl*, despite being much less frequent in number compared to *jag* and *du*. *Du+väl* accounts for 39% of all combinations of egophoric pronouns and *väl,* whereas *jag+väl* has the lowest percentage of all such combinations, with 28%. *Man* frequently combines with both *ju* and *väl* in accordance with its generic meaning, which aligns with the notion of shared access featured in *ju/väl* (Bergqvist, 2020, p. 491). The frequent combinations of *man+ju/väl* and *du+väl* are remarkable given the comparatively lower number of *man/du* pronouns in the corpus. It is evident that subject person co-varies with modal particles according to the demonstrated connection between egophoric reference and the respective epistemic territories of the speech-act participants. If this were not the case, then a more random distribution of forms would be attested. Instead, we see a clear patterning with egophoric pronouns and *ju/väl* that suggests a grounding of perspective with the speaker for *ju* and the addressee for *väl.* Both modal particles allocate epistemic authority in a context of shared access since this is a semantic feature of both forms. With mental predicates, this allocation is made differently, although there are functional overlaps between these and modal particles like *ju/väl* (see also section Shared Perspective, below).

Mental Verbs and Subject Person

This section starts from Dahl's (2000) observations regarding egophoricity in spoken Swedish discourse, where mental verbs are more likely to have egophoric arguments in comparison to copular verbs, which commonly have allophoric arguments. Dahl notes that egophoric arguments are differently distributed, where second person (*du*) is more common than first person (*jag*) with some mental predicates such as *se* ("see"), *veta* ("know"), *vilja* ("want"), and *förstå* ("understand"), whereas, propositional attitude verbs like *tänka* ("think"), *tro* ("believe"), and *tycka* ("think/feel") are much more common with first person (*jag*) than second person (*du*). He speculates that mental verbs with complements containing shared knowledge favor first-person subjects, but he does not discuss this possibility in any detail. The role of shared vs. private perspectives in the distribution of mental verbs, such as *tro* ("believe") and *veta* ("know"), is discussed in section Perspectivizing Constructs and Their Distribution, below.

Another factor that determines the distribution of egophoric arguments is syntax, specifically SV and VS predicate-argument order. The VS-order produces either an interrogative sentence or the fronting of a non-subject constituent. Dahl (2000) notes that parenthetical mental verbs (e.g., *förstår du,* "you see"; Dahl, 2000, p. 56) make up a large number of such (non-interrogative) cases. He also notes that *du* occurs much more often with VS order (302 out of 546 cases, i.e., 55%) than *jag* (170 out of 720 cases, i.e., 24%; see Dahl, 2000, p. 56). To what degree such instances are interrogative is not discussed in detail, neither are examples provided of syntactic fronting. Given the untagged nature of the corpus used in this investigation, it requires some manual counting in order to arrive at an answer to this question, and some instances remain difficult to classify. Despite these challenges, there are some telling patterns that can be observed in the corpus, as detailed in the sections directly below. The following sections only discuss present tense forms. These present tense forms are much higher in number (by an order of ten) than past tense forms, and given the stated focus on egophoric reference in the immediate discourse context, present tense forms are a natural starting point.

Tro (Think/Believe)

The majority of the 164 instances of the construction *tror du* ("[do] you think") in the corpus are interrogative. Whether they convey polar questions, or if they request a more general response from the addressee, has not been determined for all instances given that the count is made from a text that does not indicate intonation and other prosodic cues relevant to making such distinctions. Out of 100 random examples of the VS-construction *tror du* ("[do] you think"), only three instances

TABLE 2 | Co-distribution of egophoric pronouns and *tror*.

	Numbers	Percentages	SV	VS
Jag (1S)+*tror*	974	72%	553	421
Man (Gen)+*tror*	15	1%	15	–
Du (2S)+*tror*	188	14%	24	164
Other+*tror*	179	13%		
Total	1,356	100%		

were of syntactic fronting (e.g., *det tror du inte,* "you don't think so"). The remaining 97 cases were all interrogative, requesting a response from the addressee. For the SV-construction *du tror* ("you think") of which there are only 24 instances in the corpus, almost half convey a question despite being declarative by form. This aligns with previously discussed observations regarding the form-function correspondence in polar questions (see section Status, Stance, and Territories of Information, above). Examples of "declarative questions" in the corpus are *så du tror att* ("so you think that") and *du tror inte* ("you don't think").

The situation is altogether different for first-person *jag,* where the SV-construction *jag tror* occurs 553 times in the corpus and the VS-construction *tror jag* occurs 421 times. Out of one hundred random examples of the SV construction, no instances of declarative interrogatives were found. For the VS-construction *tror jag,* 160 out of 421 instances are examples of the phrase *det tror jag* (i.e., syntactic fronting). No instances of interrogative sentences were found in one hundred randomly selected examples of *tror jag.* Examples of SV and VS-constructions with first-person *jag* are shown in example (9):

(9) A: *joo hemmadjur hundar hund*
Well pets dogs dog
jag tror *vi har haft*
1S think 1P have have.PRF
allt utom hund här
everything but dog here
"Yeah, house animals, dogs, dog, I think we have had everything but a dog here."
B: *jajamän* ***det tror jag med***
that.is.right it think 1S too
"Yup, I think so too."

In addition to the syntactic difference between the first-position utterance, *jag tror,* in (Example 9; line 1) and the response turn, *tror jag,* in (line 2), we may view these as composing a single diagraph, where *jag tror* and the following response, *det tror jag med,* make up a unit that exists across the turn. Similar diagraphs are prevalent in spoken Swedish and they are suggestive of the interactional motivations for the use of epistemic expressions like *jag tror* that go beyond the subjective epistemic evaluation of the speaker (see section The Dialogic Nature of Grammar; Du Bois, 2014). The differences between first-person *jag* and second-person *du* in combination with a mental verb like *tror* may be stated in both absolute and relative frequencies. Out of 1,356 counted instances of *tror*, the egophoric pronouns (*jag, du, man*) account for 1,177 combinations, corresponding to 87% of all instances of *tror.* First- and second-person plural forms make up < 10 instances, so these are grouped with third-person referents, for sake of simplicity. **Table 2** shows how these combinations are divided according to subject person and sentence-type.

The numbers and percentages in **Table 2** should be viewed against the total number of egophoric pronouns in the corpus, which, as stated in section Modal Particles and Subject Person, are 13,768 for *jag,* 7,248 for *du,* and 4,805 for *man.* As one might expect, first-person *jag+tror* is, by far, the most common, amounting to 72% of all combinations of *tror* and any argument; *jag+tror* is five times as common as *du+tror. Man+tror* only amounts to 1% of all instances. The very low co-occurrence of *man* and *tror* is most likely due to the generic function of *man* to presuppose a measure of generic-ness with respect to an event. The private character of *tro* does not permit genericity to be a prominent feature of utterances containing this verb (cf. Bergqvist, 2020). In fact, all mental predicates are infrequent with *man,* but to varying degrees. Sensory and cognitive predicates like *se* ("see") and *veta* ("know") have a higher percentage of arguments with *man*, as do volitional predicates like *vilja* ("want"). Reasons for these differences in frequency are discussed in section Private Perspective, below. Returning to the difference between the total number of egophoric pronouns in the corpus and their co-distribution with *tro,* the numbers in **Table 2** indicate that their combination is primarily a means to signal the perspective of the speaker with respect to some event. This view is reinforced by the complete lack of interrogative sentences with the VS-construction (*tror jag*). All counted instances of this construction are examples of syntactic fronting (e.g., *det tror jag med*). The opposite holds for SV and VS-constructions with second-person *du.* Only a handful of non-interrogative phrases were found even with SV constituent order (*du tror*), which furthermore amount to <2% of all instances. This means that the overwhelming majority of all utterances that target the mental state of the addressee using the epistemic verb *tror*, do so by asking for the perspective of the addressee and their epistemic estimation of some talked-about event. With mental predicates, sentence-type has a clear perspectivizing function, which is predictable from their co-occurrence with egophoric arguments.

Veta (Know)

The mental verb *veta* ("know") occurs with egophoric arguments in 76% of all cases, which is slightly lower than *tro*, which, as stated, combines with egophoric arguments in 87% of all sentences (see **Table 2**, above). The distribution of *veta* with first- and second-person pronouns differs in that *jag+vet* amounts to almost 36% of all cases and *du+vet* makes up just under 40% of all cases, making second-person *du* the most frequent argument with *vet.* The distribution of arguments with *vet* is in **Table 3** and may be compared to the numbers in **Table 2**, above.

More than half of the 236 instances of the VS-construction *vet+jag* consist of *det vet jag,* i.e., an instance of syntactic fronting (128 of 236). No interrogative examples were found, mirroring the results for *tror+jag* in **Table 2**, above. All investigated instances of *vet+jag* are examples of syntactic fronting. In comparison, the VS-construction with second person, *vet du,*

TABLE 3 | Co-distribution of egophoric pronouns and *vet*.

	Numbers	Percentages	SV	VS
Jag (1S)+*vet*	1,001	35.9%	765	236
Man (Gen)+*vet*	84	3%	41	43
Du (2S)+*vet*	1,104	39.5%	318	786
Other+*vet*	603	21.6%		
Total	2,792	100%		

occurs 786 times, and only 63 instances consist of *det vet du* ("know you that"), suggesting that few of these cases are instances of the syntactic fronting VS-construction. However, *vet du* also functions as a discourse marker that occurs at the beginning and the end of utterances:

(10) A: så det är så svårt å
So it be.PRS so hard to
träffa henne **vet du** om
meet 3.O you know if
man ringer till henne på
Gen call.PRS to 3O.F on
morron **vet du** ja då är
morning you know yes then be.PRS
hon ute med hunden
3S.F out with the.dog
"So, it's so hard to see her, you know, if you call her in the morning, you know, well then she is out walking her dog."
B: ja då så är hon inte hemma
Yes then so be.PRS 3S.F not home
"Yeah, then she's not at home."

Vet du/vettu as a discourse marking expression introduces new information into the conversation, with an aim to establish a shared perspective between the speaker and the addressee (see Lindström, 2008, p. 65). This means that although the expression contains a second person marker (*du*), it signals the perspective of the speaker, albeit with an invitation for the addressee to share this perspective. The discourse regulating function of *vet du/vettu* to establish a shared perspective between the speech-act participants is further discussed in the subsequent sections, below.

The SV-construction *jag vet* occurs 765 times and is over three times as common as the VS-construction *vet jag* (236 instances). A total of 478 of these are constituted by *jag vet inte* ("I don't know"), which literally means that the speaker is unaware, or ignorant, of some event. However, the fact that this expression makes up 62% of the total number of instances of *jag vet* suggests its function as a discourse-marking expression:

(11) B: hur har du det med
how have.PRS 2S it with
kärleken nu MÄRTA
love now PN
"How is your love life these days, Martha?"
A: ja **jag vet inte** nu ska dom
Well I don't know now will.PRS 3PL.S
åka till Norge
go to Norway
"Well, I don't know, now they're going to Norway."

The expression *jag vet inte* in example (11) is uttered in response to a question, but it does not express the ignorance of the speaker as much as it signals their uncertainty with respect to answering the question. This function of *jag vet inte* to signal the epistemic uncertainty/ignorance of the speaker can be seen in (12), where *jag vet inte* is followed by *jag tror inte* in the same utterance by Speaker A. The preceding utterance by Speaker B is not a question or a request, so the response of Speaker A is not prompted by Speaker B but qualifies the knowledge of speaker A on the event talked about:

(12) A: ja just det
Yes right it
nä dom hade
no 3P.S have.PST
väl klarat labben
MP.ADR manage.PRF laboration
men dom hade inte
but 3P.S have.PST NEG
klarat tentan
manage.PRF exam
"Yeah, right, they managed the laboration but they didn't manage the exam, right?"
B: jaha
Oh really!
A: **jag vet inte** om dom har vart
I don't know if 3P.S have.PRS be.PRF
uppe igen
up again
jag tror inte dom har
1s think NEG 3P.S have.PRS
"I don't know if they have written the exam again, I don't think that they have."

The relatively high number of *vet* combined with egophoric pronouns is connected to the semi-grammatical status of *vet du* and *jag vet inte* as discourse markers. Out of seven investigated mental predicates, *vet* occurs with the highest frequency (see also **Table 3**, section The Epistemic Perspective of the Addressee, below). Lindström (2008) mentions other discourse marking expressions such as *tro* ("think") and *vetja* ("I know"), but there are relatively few instances of these expressions in the investigated corpus. The native-speaker intuition of the author suggests that *vetja,* and possibly *tro* as well, are somewhat archaic expressions, which are not used by speakers of Standard Swedish under 40 years of age.

Hoppas

The verb *hoppas* ("hope") is not very frequent in the corpus and does not show any signs of developing into a discourse marking expression. Almost all of the 103 examples of *hoppas* are uttered from the perspective of the speaker, as indicated by either the first-person subject pronoun *jag*, the

generic *man*, or the first-person plural pronoun *vi.* Many instances are also exclamative with pronoun-ellipsis, such as in example (13):

(13) B: ja, hoppas det, det kanhända
yes hope.PRS it it could.be
det inte är så
it NEG be.PRS so
långt in, jag vet inte riktigt
far in 1S know NEG really
"Yes, hopefully, it could be that it's not so far in, I don't really know."

The expressions *vi får hoppas/får vi hoppas/vi hoppas/hoppas vi* ("we should hope"/"let's hope") occur 24 times altogether, and given the low number of occurrences for *vi* with mental predicates (and in the corpus as a whole), this is a noticeable increase in frequency. In the corpus, this appears to be a formulaic expression, but not one has become grammaticalized like *vet du/vettu* [see section Veta (Know), above]. There are no instances of *du* ("you") with *hoppas*, and there is only one instance of non-egophoric subject marking with *hoppas* in the corpus. This example is in (14), below:

(14) B: dom väntar ju på externa
3P.S wait.PRS SPKR on external
men dom hoppas
but 3P.S hope.PRS
ju få tag på nån då som
SPKR get take on someone then that
har hållit på
have.PRS keep.PRF on
med KITS-programmering i tjugo år
with KITS.programming in twenty years
"They're obviously waiting for external, but they are hoping to find someone, then, who has been KITS-programming for 20 years."

The distribution of *hoppas* with egophoric pronouns is special in the sense that it never occurs with *du,* but it combines with *vi* to a relatively higher degree than other mental predicates. The combination of *vi* and *hoppas* suggests a shared perspective that includes the addressee, permitting a functional comparison to *vet du* [see section Veta (Know), above]. Although second-person *du* is entirely absent with *hoppas,* this mental verb is not only exclusively used to express the positive expectations of the speaker but also those of the addressee.

The Epistemic Perspective of the Addressee

Mental verbs express various aspects of the beliefs, expectations, and desires of the speaker. As we have seen directly above, however, mental verbs display differences when it comes to attributing such states to the addressee. Some mental verbs, such as *hoppas* ("hope"), are almost never used with second-person subject pronouns, reflecting the reluctance of the speaker to exclusively attribute such internal states to the addressee. The preferred way to attribute "hoping" to the addressee is by using the first-person plural form *vi. Tro* refers to the private estimation of one of the speech-act participants and is almost always declarative with first-person *jag* and almost always interrogative with second-person *du.* The exclusive perspective of the speaker is, by far, most frequently referred to in the corpus. A verb like *vet* ("know"), on the other hand, occurs with second-person pronouns more often than with first-person pronouns. The combination of second-person *du* with *vet* mostly produces interrogative clauses, or instantiates the discourse particle *vet du/vettu*. This is in contrast with first-person pronouns, which mostly occur in declarative clauses and almost always convey statements. *Vet* in combination with *du* also has resulted in a discourse marker that indicates how the addressee should interpret an utterance. As such, it invites the addressee to share the estimation of the speaker of a novel proposition (see section Mental Verbs and Subject Person, above). There is thus a strong orientation toward the perspective of the addressee in the use of *vet* with second-person *du*.

These facts suggest that mental verbs may be arranged along a cline where the (explicit) attribution of primary perspective to the addressee goes from restricted to unrestricted. The notion of restricted attribution may be based on relative frequencies (percentages) and the predominance of a certain sentence-type in utterances with a given mental predicate. **Table 4** features the number of instances of seven mental verbs occurring in the corpus. These are listed in the leftmost column. The co-distribution percentages with the egophoric pronouns *jag* and *du* (columns 3 and 6) are used to calculate the attribution number in the rightmost column. This column indicates the ratio between *jag* and *du* in combination with a given verb. For example, the verb *tycka* ("think"/"opine") with *du* only amounts to one-tenth of the cases of *tycka+jag.* By contrast, the verb *se* ("see") is two and a half times more common with *du* than it is with *jag.* The verb *hoppas* ("hope") is not attested with *du* and therefore has no ratio value.

The figures in **Table 4** indicate the preference of the speaker to attribute various aspects of belief, opinion, expectation, and sensory access to the perspective of the addressee, depending on the meaning of the lexemes. From the point of view of egophoricity, the perspective of the speech-act participants is non-randomly mapped onto different kinds of mental verbs, and although the list is far from complete, it suggests the attributability of mental states, as indicated by verbs of emotion, cognition, and perception. The numbers in **Table 4** indicate that a perceptual verb like *se* ("see") is commonly used to talk about the perception of the addressee. The same is true for *veta* ("know"), which targets the addressee's knowledge in interrogative clauses, or in the form of the discourse marker *vet du/vettu*. Interestingly, grammaticalized, direct evidential markers sometimes originate with verbs of seeing, and there are attestations of semantic overlap between seeing and knowing in such forms, which is in agreement with well-known metaphorical mappings such as "I see what you mean" (e.g., Matlock, 1989; cf. Aikhenvald, 2004). This semantic overlap is thus reflected in the distributional characteristics of *se* and *veta* in spoken Swedish.

TABLE 4 | Distribution of mental and sensory verbs with *jag/du* and SV/VS constituent order.

Mental verb	Instances	Verb+*jag*	SV+*jag*	VS+*jag*	Verb+*du*	SV+*du*	VS+*du*	Attr
Hoppas (hope)	103	76%	61	18	–	–	–	–
Tycka (think)	1,255	71%	445	443	8%	25	81	0.11
Tro (believe)	1,356	72%	553	421	14%	24	164	0.19
vilja (want)	592	32%	133	54	12%	28	44	0.38
Känna (feel)	205	34%	49	21	15%	5	25	0.44
Veta (know)	2,792	36%	765	236	40%	318	786	1.11
Se (see)	657	10%	51	16	25%	36	130	2.5

PERSPECTIVIZING CONSTRUCTS AND THEIR DISTRIBUTION

Modal particles and mental verbs offer complementary views on how the perspective of the speaker and the addressee is expressed in language. Both resources map onto the subject person in ways that point to the pervasiveness of egophoricity and the special status of the perspective of the addressee. The main difference in perspective coneveyed by modal particles and mental verbs concerns shared and non-shared/private access. The modal particles *ju/väl* signal shared access to events and, at the same time, place epistemic authority with the speaker and the addressee, respectively. Mental verbs, on the other hand, are by definition private to the speaking subject to which they are referentially linked. In order to allocate the private perspective of belief, opinion, and expectation with the addressee, the speaker draws on resources such as sentence-type to signal that these private states are outside the domain of the speaker and belong to another speaking subject. The development of some combinations of mental verbs and subject pronouns into discourse markers, e.g., *vet du* (see section Mental verbs and Subject Person, above), is another way that the speaker can address the perspective of the addressee with the speaker maintaining their own perspective in the use of such forms (see section Hoppas, above).

Shared Perspective

When analyzing epistemic evaluations semantically, in terms of e.g., certainty, doubt, or ignorance, such expressions are almost always viewed from the perspective of the speaking subject. This analytical approach rests on a long philosophical tradition that treats the perspective of the other with caution, and in some parts of this tradition, the perspective of the speaker is even regarded as the only one that exists. Seen from the point of view of language use, however, it might not be warranted to assume that an exclusive/private speaker-stance is default when it comes to qualifying knowledge of events. There are in fact data that suggest otherwise, namely that knowledge is regarded as non-exclusive in casual conversations and that any claim of exclusive knowledge is marked in such contexts (see e.g., Zariquiey, 2015, for a discussion of genre in Kakataibo). The frequent use of *ju* and *väl* in spoken Swedish supports the view that shared access to events is commonplace in spontaneous conversations. However, the distribution of *ju* and *väl* in spoken Swedish also suggests that shared access to events does not equate symmetrical access to such events. In terms of frequency, *ju* is more common, which is a fact that coincides with the prevalence of the perspective of the speaker in spoken discourse as indicated by the high number of first-person subject pronouns (*jag*) and the generic pronoun *man*. This means that speakers explicitly mention themselves more often than their addressees and that assertions containing the modal particle *ju* align with this predominance of speaker perspective. Bergqvist (2020) argues that *ju* encodes shared access to events along with the epistemic authority of the speaker, and *väl* encodes shared access and the epistemic authority of the addressee. The allocation of epistemic authority to reside with the addressee is predominantly with reference to events that involve the addressee as an actor, or affected party. This is reflected by the high number of co-occurrences of *du* and *väl*. The study conducted by Bergqvist (2020) also suggests that the placement of epistemic authority with the addressee commonly is conceptualized as speaker uncertainty or estimated probability (see e.g., Gast, 2008, for the German cognate *wohl*) but that these concepts are defeasible and that the epistemic authority of the addressee is not. The quasi-interrogative function of *väl* to indicate a question, despite always occurring in declarative clauses, is also produced by the semantic feature of addressee authority. This is an expected result since the primary function of sentence-type is to request an assertion by the addressee based on his/her perspective, especially with events and actions that involve the addressee. The assertion of the speaker, while included in the use of *väl*, is by definition subordinate to the perspective of the addressee (assumed).

Private Perspective

The analysis of the mental verbs *tro* ("believe"), *veta* ("know"), and *hoppas* ("hope") above shows that these occur with different frequencies and that they are differently distributed with respect to egophoric pronouns (see section The Epistemic Perspective of the Addressee). Based on frequency and distribution, the verb *veta* is more readily available for talking about the epistemic perspective of the addressee than *tro*. While *tro* displays the highest number of combinations with egophoric arguments, amounting to 87%, only 14% consist of *tro+du*. The relatively restricted number of occurrences of *tro+du* along with the predominance of this combination in interrogative contexts (VS, see **Table 1**, above) suggests that the speaker prefers to address the epistemic territory of the addressee (i.e., their

beliefs) by way of placing epistemic authority with the same. This allocation of epistemic authority is achieved by asking the addressee for their beliefs by means of formally interrogative utterances (e.g., *tror du inte dom skulle vilja göra det då?*), or by declarative questions (cf. Stivers and Rossano, 2010). Talking about the beliefs of the addressee is not preferred unless these are addressed in the form of a question. This observation confirms the theory of stance proposed by Heritage (2012; see section Status, Stance, and Territories of Information), where an alignment between the epistemic status and stance of a speaker is unmarked in conversation. Utterances targeting the knowledge of the addressee by using *veta* are less restricted, and combinations of *veta+du* are also much more common than *tro+du*. Almost 40% of all instances of *veta* combine with *du*, and VS-combinations of *veta+du* amount to twice the number of SV-combinations. Most of the former are interrogative rather than instances of syntactic fronting; the opposite is true for combinations of *veta* and *jag*. *Veta* is also found in two frequently occurring discourse markers: *vet du/vettu* and *jag vet inte*. *Vet du/vettu* marks new information with an invitation for the addressee to share the perspective of the speaker, a function that bridges a private predicate like *veta* to modal particles like *ju* and *väl*. Based on frequency and distribution, a predicate like *veta* is less private than *tro,* suggesting the connection between knowing and a verifiable fact, compared to belief and subjectively estimated possibility.

CONCLUDING DISCUSSION AND SUMMARY

Perspective-taking is pervasive in grammar, and various forms of epistemic marking (modal, evidential, or egophoric) convey epistemic perspectives in terms of authority, either by claiming it, disclaiming it, or assigning it to another speech-act participant (see Bergqvist and Kittilä, 2020). Important aspects of assuming a perspective can be seen in dialogic exchanges between the speech-act participants, where parallelisms and the partial reproduction of previous utterances are used in the negotiation of epistemic authority. A diagraph that features the mental verb *tro* is seen in Example (2), repeated here. The two phrases, namely *det tror jag inte* (line 1) and *tror du inte det* (line 2) illustrate a pattern of epistemic alignment where the distribution of the epistemic modal verb *tror* depends on the subject person, sentence-type, and turn position:

(2) B: *det* *känns* *nog* *inte* *som* *du*
it feel.PRS MP.nog NEG like 2S
har *föreställt* *dej*
have imagine.PRF 2O
det ***tror*** ***jag*** *inte*
it think 1S NEG
"It doesn't feel like you have imagined it, I don't think."
A: ***tror*** ***du*** *inte* *det*
think 2S NEG it
"You don't think so?"
B: *nä* *man* *kan* *nog* *aldrig* *fatta*
no one can MP.nog never understand
hur *det* *känns* *egentligen*
how it feels really
"No, you could probably never understand how it feels, really."

Speaker B addresses the (future) emotional state of Speaker A by qualifying his/her statement with the modal particle *nog* ("probably") and the phrase *det tror jag inte*. Speaker A focuses on the belief of Speaker B by uttering *tror du inte det?,* prompting Speaker B to elaborate on their position, which they do in the subsequent line. The perspectivizing property of epistemic markers is thus visible in parallelisms, such as in (2), where the expectations and considerations relevant to the respective positions of the speech-act participants are reflected in the use of mental verbs like *tro*. In the present paper, three formally distinct perspectivizing resources, namely modal particles, mental predicates, and sentence-type are analyzed from how they situate the perspective of the speaker and the addressee with respect to talked-about events. Patterns that emerge from the co-distribution of such forms with egophoric subject pronouns (*jag/man/du*) indicate the epistemic perspectives implicit in these pronouns. The modal particles *ju* and *väl* occur with *jag/man/du* in non-random ways in accordance with the proposed semantics of both forms, i.e., shared access from the perspective of the speaker and the addressee, respectively. Complement-taking mental predicates are differently distributed against *jag/du* and can be analyzed from their capacity to reflect consideration of the perspective of the addressee. Two of the most common mental verbs are *vet* and *tror*, both of which occur with egophoric pronouns more than 75% of the time but display differences with regard to their co-distribution with *jag/du*. *Tror/vet+du* conveys a question, whereas *tror/vet+jag* almost never does. This patterning strongly suggests the epistemic function of sentence-type in egophoric contexts, i.e., when events involving the addressee are addressed. In addition, *vet+du* has developed into a discourse marker with intersubjective connotations and occurs with a much higher frequency than *tror+du*. These distributional patterns and resulting grammaticalization paths contribute analytical aspects to the study of epistemic marking that is amiss in traditional descriptive accounts of such markers.

DATA AVAILABILITY STATEMENT

Requests to access these datasets should be directed to HB, henrik.bergqvist@ling.su.se.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This study was supported by Swedish Research Council (dnr. 2017-01969) and the Marcus and Amalia Wallenbergs Minnesfond (MAW 2017.0081).

REFERENCES

Abraham, W., and Leiss, E. (eds.). (2012). *Modality and Theory of Mind Elements Across Languages (Trends in Linguistics Studies and Monographs 243)*. Berlin: Walter de Gruyter. doi: 10.1515/9783110271072

Aijmer, K. (1977). Partiklarna ju och väl [The particles ju and väl]. *Nysvenska Studier* 57, 205–216.

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford: Oxford University Press.

Bergqvist, H. (2015). "Epistemic marking and multiple perspective: an introduction. Special Issue: Epistemic marking in typological perspective," in *Language Typology and Universals (STUF)*, H. Bergqvist, and L. S. Roque, 123–141.

Bergqvist, H. (2020). Swedish modal particles as markers of engagement: evidence from distribution and frequency. *Folia Linguist.* 54, 469–496. doi: 10.1515/flin-2020-2047

Bergqvist, H., and Kittilä, S. (2017). Person and knowledge: introduction. *Open Linguist.* 3, 18–30. doi: 10.1515/opli-2017-0002

Bergqvist, H., and Kittilä, S. (eds.). (2020). "Epistemic perspectives: evidentiality, egophoricity, and engagement," in *Evidentiality, Egophoricity, and Engagement (Studies in Diversity Linguistics 99)* (Berlin: Language Science Press), 1–21.

Bergqvist, H., and Knuchel, D. (2017). Complexity in egophoric marking: from agents to attitude holders. *Open Linguist.* 3, 359–377. doi: 10.1515/opli-2017-0018

Biber, D., and Finegan, E. (1989). Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text* 9, 93–124. doi: 10.1515/text.1.1989.9.1.93

Cornillie, B. (2009). Evidentiality and epistemic modality: on the close relationship of two different categories. *Funct. Lang.* 16, 32–44. doi: 10.1075/fol.16.1.04cor

Couper-Kuhlen, E., and Selting, M. (2018). *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge: Cambridge University Press.

Creissels, D. (2008). Person variation in Akhvakh verb morphology: functional motivation and origin of an uncommon pattern. *STUF* 61, 309–325. doi: 10.1524/stuf.2008.0027

Curnow, T. J. (2002). Conjunct/disjunct marking in Awa Pit. *Linguistics* 40, 611–627. doi: 10.1515/ling.2002.025

Dahl, Ö. (2000). Egophoricity in discourse and grammar. *Funct. Lang.* 7, 37–77. doi: 10.1075/fol.7.1.03dah

Dimmendaal, J. G. (2001). Logophoric marking and represented speech in African Languages as evidential hedging strategies. *Aust. J. Linguist.* 21, 131–157. doi: 10.1080/07268600120042499

Du Bois, J. (2007). "The stance triangle," in *Stancetaking in Discourse,* ed R. Engelbretson (Amsterdam: John Benjamins), 139–182. doi: 10.1075/pbns.164.07du

Du Bois, J. (2014). Towards a dialogic syntax. *Cogn. Linguist.* 25, 359–410. doi: 10.1515/cog-2014-0024

Evans, N. (2006). View with a view: Towards a typology of multiple perspective. *Berk. Linguist. Soc.* 32, 93–120. doi: 10.3765/bls.v31i1.3429

Evans, N. (2012). "Nen assentives and the problem of dyadic parallelisms," in *Practical Theories and Empirical Practice: Facets of a Complex Interaction*, ed A. C. Schalley (Amsterdam: John Benjamins), 159–83. doi: 10.1075/hcp.40.07eva

Evans, N., Bergqvist, H., and San Roque, L. (2018). The grammar of engagement I: framework and initial exemplification. *Lang. Cogn.* 10, 110–140. doi: 10.1017/langcog.2017.21

Gast, V. (2008). "Modal particles and context updating: the functions of German ja, doch, wohl and etwa," in *Modalität und Grammatikalisierung,* eds O. Letnes, E. Maagerø, and H. Vater (Trier: Wissenschaftlicher Verlag), 153–177.

Givón, T. (1981). Typology and functional domains. *Stud. Lang.* 5, 163–193. doi: 10.1075/sl.5.2.03giv

Givón, T. (2001). *Syntax: An Introduction,* Vol 2. Philadelphia, PA: John Benjamins Publishing Company.

Goffman, E. (1981). *Forms of Talk.* Philadelphia, PA: University of Pennsylvania Press.

Grzech, K., Schultze-Berndt, E., and Bergqvist, H. (eds.). (2020). Knowing in interaction: empirical approaches to epistemicity and intersubjectivity in language. *Folia Linguistica* 54, 281–315.

Hale, A. (1980). "Person markers: finite conjunct and disjunct forms in Newari," in *Papers in Southeast Asian Linguistics* 7, ed R. Trail (Canberra, ACT: Pacific Linguistics), 95–106.

Hengeveld, K., and Mackenzie, L. J. (2008). *Functional Discourse Grammar: A Typologically-Based Theory of Language Structure*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199544004.013.0015

Heritage, J. (2012). Epistemics in action: action formation and territories of knowledge. *Res. Lang. Soc. Interact.* 45, 1–29. doi: 10.1080/08351813.2012.646684

Heritage, J. (2013). "Epistemics in conversation," in *Handbook of Conversation Analysis*, eds J. Sidnell and T. Stivers (Oxford: Wiley-Blackwell), 370–94. doi: 10.1002/9781118325001.ch18

Heritage, J., and Raymond, G. (2005). The terms of agreement: indexing epistemic authority and subordination in assessment sequences. *Soc. Psychol. Q.* 68, 15–38. doi: 10.1177/019027250506800103

Jespersen, O. (1924). *The Philosophy of Grammar*. Chicago, IL: University of Chicago Press.

Kamio, A. (1997). *Territories of Information*. Amsterdam: John Benjamins. doi: 10.1075/pbns.48

Kockelman, P. (2004). Stance and Subjectivity. *J. Linguist. Anthropol.* 14, 127–150. doi: 10.1525/jlin.2004.14.2.127

Labov, W., and Fanshel, D. (1977). *Theraputic Discourse: Psychotherapy as Conversation*. New York, NY: Academic Press.

Leech, G. (1998). "English grammar in conversation," in *Proceedings of the Chemnitz Symposium 20-21 February 1998*, eds J. Schmied and A. Wilson (Chemnitz University of Technology).

Leech, G., and Svartvik, J. (2002). *A Communicative Grammar of English*. London: Routledge.

Lindström, J. (2008). *Tur Och Ordning: Introduktion Till Svensk Samtalsgrammatik*. Stockholm: Norstedts.

Löfström, J. (1988). *Repliker Utan Gränser: Till Studiet av Syntaktisk Struktur i Samtal [Boundless Conversational Exchanges: A Contribution to the Study of Syntactic Structure in Conversation]*. Gothenburg: University of Gothenburg dissertation.

Matlock, T. (1989). "Metaphor and the grammaticalization of evidentials," in *Proceedings of the Fifteenth Annual Meeting of the BLS (Berkeley Linguistics Society.)*, 215–225.

Mushin, I. (2001). *Evidentiality and Epistemological Stance: Narrative Retelling*. Amsterdam: John Benjamins. doi: 10.1075/pbns.87

Oswalt, L. R. (1986). "The evidential system of kashaya," in *Evidentiality: the Linguistic Coding of Epistemology*, eds W. Chafe and J. Nichols (Norwood, NJ: Ablex Pub), 29–45.

San Roque, L., Floyd, S., and Norcliffe, E. (2018). *Egophoricity*. Amsterdam: John Benjamins. doi: 10.1075/tsl.118.01san

San Roque, L., and Loughnane, R. (2012). The New Guinea Highlands evidentiality area. *Linguist. Typol.* 16, 111–167. doi: 10.1515/lity-2012-0003

Stivers, T., and Rossano, F. (2010). Mobilizing response. *Res. Lang. Soc. Interact.* 43, 3–31. doi: 10.1080/08351810903471258

Teleman, U., Hellberg, S., and Andersson, E. (1999). *Svenska Akademiens Grammatik [Grammar of the Swedish Academy],* Vol. 4. Stockholm: Norstedts.

Van Valin, R. D., and LaPolla, R. J. (1997). *Syntax: Structure, Meaning and Function.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139166799

Wälchli, B. (2015). Logophoricity in eastern vidzeme: the literary latvian idiolect of andrievs niedra and leivu estonian. *Baltic Linguist.* 6, 141–192. doi: 10.32798/bl.397

Zariquiey, R. (2015). The encoding of addressee's perspective in Kakataibo (Panoan, Peru). *Lang. Typol. Univers.* 68, 143–164. doi: 10.1515/stuf-2015-0008

Zimmermann, M. (2011). "Discourse Particles," in *Handbook of Semantics,* eds P. Portner, C. Maienborn, and K. von Heusinger (Berlin: Mouton de Gruyter), 2011–2038. doi: 10.1515/9783110255072.2012

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.

Copyright © 2021 Bergqvist. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Communication

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fcomm.2021.611187

The Benefits and Obstacles to Perspective Getting

Debby Damen, Monique M. H. Pollmann and Teri-Louise Grassow*

TiCC, Department of Communication and Cognition, Tilburg University, Tilburg, Netherlands

In general, people tend to rely on egocentric projection when predicting others' emotions, attitudes, and preferences. However, this strategy is less effective than the more obvious strategy of directly asking others what they feel, think, or desire ('perspective getting'). In three experimental studies, we investigated how likely people are to ask for others' perspectives, whether it leads to better predictions, and what factors impede perspective getting. In the first study, we let participants predict how happy another person would be with different money distributions. Only 26% of all people engaged in perspective getting, and it did not lead to better predictions. In the second study, we let people predict how expensive another person would think certain products are. The majority of people engaged in some form of perspective getting, but only 23% of all people did this thoroughly. Perspective getting did lead to better predictions. In the final study, we let people predict another person's attitudes about a wide range of topics. Here, 70% of the people engaged in perspective getting and 12.5% did so thoroughly. Again, perspective getting led to better predictions. We found that confidence acted as a barrier for perspective getting. We also tested whether pointing out that perspective getting is the best strategy would increase perspective getting. We do not find a positive effect of this intervention. We discuss possible other interventions to increase people's tendency to get rather than take perspective.

Keywords: perspective getting, perspective taking, egocentric projection, interpersonal accuracy, overconfidence, experimental studies

OPEN ACCESS

Edited by:
Jorrig Vogels,
University of Groningen, Netherlands

Reviewed by:
Margaret Echelbarger,
University of Michigan, United States
Carla Contemori,
The University of Texas at El Paso, United States

****Correspondence:***
Debby Damen
D.J.Damen@tilburguniversity.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *28 September 2020*
Accepted: *12 May 2021*
Published: *28 May 2021*

Citation:
Damen D, Pollmann MMH and Grassow T-L (2021) The Benefits and Obstacles to Perspective Getting. Front. Commun. 6:611187. doi: 10.3389/fcomm.2021.611187

INTRODUCTION

How well do you know your partner or your friends? Do you think that you truly know their attitudes and their emotional reactions? Did you ask them about these things? Chances are that you did not and that you still feel confident about your ability to predict their beliefs and feelings. Oftentimes, however, your confidence will be misplaced. Your predictions will most likely be somewhat inaccurate because they are based on projection and too little on direct information.

Studies have repeatedly shown that we frequently overestimate our ability to infer what other people think, feel or desire, especially when we feel we truly know the other person (e.g., Swann and Gill, 1997; Scheibehenne et al., 2011; Eyal et al., 2018). Even though actively imagining what other people are thinking or experiencing should increase our interpersonal understanding of what goes on inside other people's minds, research shows that we are rarely accurate about the inferences we make (e.g., Realo et al., 2003; Ames and Kammrath, 2004; Eyal et al., 2018). One of the reasons is that the ease by which our own beliefs and perceptions are accessible clouds our ability to acknowledge another person's unique vantage point. Often, this causes us to overestimate the extent to which others view and evaluate the world similarly (e.g., Damen, et al., 2019a; Damen, et al., 2019b; Damen

et al., 2020a; Keysar, et al., 1998a; Gilovich et al., 2000; Keysar et al., 2000; Keysar et al., 2003; Ross et al., 1977; Krueger and Clement, 1994).

Taking another person's perspective is, therefore, a strategy of inferencing that is more prone to errors than other, more obvious, strategies such as actively inquiring about other peoples' mental states. Strikingly, even though asking about other's beliefs, feelings, and desires seems to be an obvious way to gain insight into other's perspective, people do not feel more confident that getting their interlocutor's perspective through conversation allows them to assess this person's perspective better than when they take it (Eyal et al., 2018; see also; Zhou et al., 2017). Apparently, people do not see the added value of asking others about their perspectives. We wonder why this is the case. In this study, we set out to investigate how likely people are to get their interlocutor's perspective. We investigate this question by examining people's propensity to ask their interlocutors about their feelings (study 1), beliefs (study 2), and attitudes (study 3) before judging them. Additionally, we examine the obstacles people face when deciding to adopt a perspective getting strategy. In particular, we examine the extent to which people's explicit knowledge that their perspective is not shared by others (study 2) and their confidence in their predictive accuracy (study 3) incites them to get rather than take perspective. We also investigate whether pointing out to people that perspective getting is the best strategy increases its use (study 3), and we explore the benefits this strategy has compared to perspective-taking on acquiring interpersonal accuracy.

Gaining Interpersonal Accuracy

Our ability to ascribe mental states onto others allows us to understand that our own view of the world might not be shared by those with whom we interact. This realization helps us to adjust our communication accordingly and helps us to collaborate, to empathize, and to build and sustain relationships. Yet, accurately perceiving another person's thoughts and feelings is an extremely difficult task (Van Boven et al., 2000). The most obvious reason for this is that we cannot directly perceive the internal thoughts of others (Myers and Hodge, 2009). In fact, to guess what goes on inside other minds, we regularly employ different "tools" to gain this insight (for an overview, see Ames, 2005). For instance, we might employ evidence-based techniques, such as perceiving and interpreting others' behavioral and emotional displays in context. In the absence of this direct behavioral "evidence", we might use stored, stereotypical or idiosyncratic information about social categories to infer what others might think, feel, or do in a particular situation (e.g., Gopnik and Wellman, 1994; Galinsky et al., 2005; Kunda and Thagard, 1996). Alternatively, we might rely on our own beliefs and experiences to explain those of others, especially when we feel the object of our attention is highly similar to us (Kenny and Acitelli, 2001; Lerouge and Warlop, 2006). In other words, perceivers use their interpretations of others' behaviors and emotions to infer other people's mental states, or they use information about social categories (stereotyping), or about their own mental states (egocentric or social projection) to attribute to others (Ames, 2004). There is no clear consensus of which type of mentalizing strategy is used when and how. The only consensus there seems to be is that, whichever strategy is adopted, people generally strive to reach an accurate understanding of what drives other people (Ames, 2005).

One solution might be to use more reliable sources of information to guide our prediction of another person's mental states (see also Ickes et al., 1990). Ample studies have evidenced that the more reliable (and relevant) information people receive about the other, the more reliable their inferences about this person's perspective will be (Zhou et al., 2017). The most reliable source of information is almost always the target of interest. Even when people try to deceive, their deception can clue to others what actually transpires inside their minds. This means that people who directly experience or acquire information about another person's actual thoughts, feelings, beliefs, or preferences are able to build a more reliable frame of reference that helps them to update their initial beliefs, and consequently, helps them to judge this person's perspective more accurately (e.g., Camerer et al., 1989; Eyal et al., 2018; Marangoni et al., 1995; Porter et al., 2000; Thompson and De Harpport, 1994; Weingartner and Klin, 2005; West, 1996; see also volume of information in; Marangoni et al., 1995, and simulation in; Zhou et al., 2017). The most recent study showing the added value of this individuated "feedback" on interpersonal accuracy is the study by Eyal et al. (2018), who instructed married couples to get their partner's perspective by conversing about each other's attitudes before estimating them. The authors compared the interpersonal accuracy of these couples against the accuracy of the couples who were instructed to take their partner's perspective by imagining themselves in their partner's situation, and against a baseline condition in which couples did not receive any formal instructions. As expected, couples who talked about their attitudes were able to assess their partner's perspective more accurately than those who took their partner's perspective or participated in the control condition. Interestingly, couples who used the most effective inferencing strategy were not more confident about their predictive accuracy than others. In other words, couples overestimated the effectiveness of perspective taking compared to perspective getting, and, thus, had a misbelief about what the best strategy was to predict the attitudes of their partner. This misbelief could prevent people from selecting the most optimal strategy.

Obstacles to Perspective Getting

How likely are people to ask others for their views? Unfortunately, Eyal et al. (2018) did not test whether people spontaneously engage in perspective getting, nor were we able to find any other studies that did. The amount of information seeking in conversations may depend on the conversational setting, the conversational and interlocutors' communicative goals, the number of conversational partners, and interlocutors' non-verbal and verbal communicative reciprocity. Several fields show that people are able to extract information from another person for a certain explicit goal. For example, if a person is interested in the quality of a certain product, they may ask others what they think about it. If a person

needs to negotiate a deal with another person, they may ask about their preferences (e.g., Schei et al., 2006). If a person is asked to find a way through a maze based on directions given by another person, they may ask clarification questions (Mills, 2014). In short, people do ask others for information. Although, even in these situations, the actual rate of questions is surprisingly low. What we are interested in are situations in which the only goal is to get a better understanding of what is on the other's mind. Do people directly ask others about their views to increase interpersonal accuracy? Research related to this question actually shows that in that case, people are not very likely to ask others for information. Below, we discuss several lines of research that give a first indication of whether people are inclined to seek information and, thus, to get rather than take perspective in conversation, even though this enumeration is, of course, not extensive.

First, research on negotiations shows the benefit of information sharing between interlocutors. When negotiators ask about their counterpart's perspective, and their counterpart answers these queries truthfully, negotiators learn more about both the compatible and incompatible interests that are at stake, positively affecting the negotiation process and the integrativeness of the negotiated outcome (e.g., Pruit and Lewis, 1975; Thompson and Hastie, 1990; Weingart et al., 1993; Butler, 1999; Hyder et al., 2000; Adair and Brett, 2005; Schei et al., 2006; Elfenbein et al., 2010). However, dispositional and situational factors may affect the amount of information that is shared between negotiators. For instance, low perceived feelings of interpersonal trust (Butler, 1995; Butler, 1999), willingness to provide truthful information (Harsanyi, 1962; Bacharach and Lawler, 1981), personal accountability (De Dreu et al., 2006), egoistic motives (De Dreu et al., 2000), and the perception of power (Schei et al., 2006) all affect the extent to which negotiators seek and process information. These moderating factors cause that negotiators are not always inclined to seek or to provide relevant information, and this is especially unfortunate as information sharing is considered to be a reciprocal process (Thompson and Hastie, 1990; Weingart et al., 1993). It should be noted that these studies all focus on interlocutors with partly opposing interests. In our studies, we will focus on interlocutors whose task it is to predict the other's perspective. So, the motives that hinder free information exchange in negotiations should play a smaller role here.

Another line of studies directs us to the doctor-patient communication setting. In these interactions, it should be obvious that the doctor asks their patient what is on their mind. However, several studies on how physicians talk to their patients show that not all patients are given the opportunity to explain their concerns, either because the physician does not start with an open question or because they interrupt their patient (on average within 23 s) (Beckman and Frankel, 1984; Marvel et al., 1999). Research has also shown that clinicians often do not check for understanding or assess their patient's beliefs by asking whether the patient has more questions (White et al., 1994) and that doctors do not sufficiently acknowledge their patients' emotional or social clues during an outpatient visit (Levinson et al., 2000). Most patients, therefore, are not able to voice their worries or their ideas, leading to misunderstanding (Barry et al., 2000). Hence, even in situations in which the other person is the best source of information, physicians do not engage in full perspective getting. In situations in which one's own view is a better proxy, like when predicting the attitudes of a friend, perspective getting will probably even be less common.

The third set of studies can be found in the gift-giving literature and hints toward possible obstacles to perspective getting. In gift-giving, perspective getting would mean to ask the other person explicitly what kind of present they would like to receive. This seems like a good strategy, but people mistakenly believe that gifts chosen from a gift registry will be appreciated less than gifts the giver came up with themselves. Giving requested gifts is seen as less thoughtful and considerate by the giver (Gino and Flynn, 2011), and apparently, giving a gift from a registry undermines the giver's motivation to show with the gift how well they know the other person (Ward and Broniarczyk, 2016). These findings show that in cases where predicting the other person's perspective is seen as a challenge, people may feel like they should not engage in perspective getting because they feel it is not the right way to show their empathic skills.

Another challenge that might prevent people from getting another person's perspective is their unawareness that their (private) perspective differs from someone else's. One theoretical account in the perspective taking literature, for instance, claims that people who are aware that significant differences between themselves and the other exist are less likely to project their perspective onto this other person (e.g., Damen et al., 2019a; Decety and Sommerville, 2003; Mitchell, 2009; Mitchell et al., 2006; Mussweiler, 2003; Santiesteban et al., 2012; Savitsky et al., 2011; Todd et al., 2011; Simpson and Todd, 2017). Hence, people who operate on a false belief of similarity might not see the need for perspective getting, whereas those who are explicitly aware that their perspective is not shared by others might.

Finally, the perspective taking literature shows that people are often overconfident in their ability to predict others' perspectives and that this overconfidence undermines their ability to learn and to improve (e.g., Einhorn and Hogarth, 1978; Eyal et al., 2018; Epley and Eyal, 2019; Ickes, 2003; Jacoby et al., 1994; Oskamp, 1965; Realo et al., 2003; Ryback, 1967; Swann and Gill, 1997; Thomas et al., 1997; see overconfidence effect in; Adams and Adams, 1961). People who are (over) confident in their perspective taking ability might, therefore, not feel the need to ask others for clarification or confirmation, making perspective getting less likely.

In general, the literature suggests that not all people will spontaneously engage in perspective getting. In the current paper, we present three studies in which we aim to test how likely people are to acquire information from the person they were trying to understand. We not only examined whether and the extent to which people inquired about their interlocutor's feelings (study 1), thoughts (study 2), and attitudes (study 3), but also whether this perspective getting strategy led to more accurate predictions. We additionally focus on two proposed obstacles to perspective getting: people's knowledge that their perspective is

not shared by others (study 2), and their confidence in their predictive accuracy (study 3). We investigate the extent to which these obstacles influence the selection of the perspective getting strategy. In the final study, we also test whether pointing out to interlocutors that perspective getting is the best strategy increases its use.

STUDY 1

The first study was designed to investigate how many people use the option to ask the other person about their feelings and whether doing so is related to interpersonal accuracy. To this end, we created situations in which people differed in perspectives. Specifically, we asked people to predict how another person felt when this person received more or less money than themselves. There were four conditions in this study: perspective taking, receiving feedback, the option for perspective getting, and control. For this paper, we are only interested in how many people use the option for perspective getting, and whether this perspective getting increases interpersonal accuracy. We therefore only report the findings of this option for perspective getting condition and present the comparisons of the other conditions in the supplement material.

Method

Participants and Sample Size

The required sample size for a medium effect size was calculated for the full study (including all four conditions) using G*Power (version 3.1.9.2). This calculation showed that we would require at least 45 participants per experimental condition to obtain an alpha err probability of .05 and a power of 80. In total, 52 participants were randomly allocated to the option for perspective getting condition. However, 6 participants were excluded from the analyses because they experienced technical difficulties during the experiment. Our analyses are thus based on 46 participants in the option for perspective getting condition (26 women, 20 men, M_{age} = 22.52, age-range 18-42). Participants were recruited from the university campus, and they received course credits for their participation. The majority of participants was Dutch (N = 33, 71.7%). The 13 non-Dutch participants came either from other countries in Europe (N = 7, 58.3%), or from India (N = 1), Colombia (N = 1), Iran (N = 1), Vietnam (N = 1) and Thailand (N = 1). The majority had received a High School diploma (N = 24, 52.2%), followed by a Bachelor's (N = 18, 39.1%) or Master's degree (N = 4, 8.7%). The study was conducted in English, and all participants were proficient in the language of the study.

Procedure

Participants were invited to partake in a study that investigated their ability to predict another person's feelings. Upon entering the lab, participants were seated in soundproof cubicles where they logged in into a computer to meet their partner via a synchronous, text-based chat (ChatPlat.com). This partner was a confederate of the study and followed a scripted text when responding to all participants' queries. The dyads were not able to see each other, but could only communicate via the chat-box that was pulled up next to a questionnaire. The chat-box filled half of the screen and was, therefore, saliently present for participants to use to chat with their partner. Participants were deceived into believing that their partner was another student partaking in the study. The experimental leader instructed participants to start the experiment once all participants were set up. At the start of the experiment, participants were explicitly informed they should introduce themselves to their partner and that they could communicate with their partner at any time during the study.

The experiment was administered in Qualtrics™. First, participants answered demographic questions (age, gender, nationality, education level), followed by a practice round to introduce them to the procedure of the experiment. In the questionnaire, participants were exposed to thirty different scenarios that appeared in a randomized order in which the computer divided fictitious amounts of money between them and their partner. To elicit different emotions, the fictitious distributions were either advantageous for the participants (e.g., 7/3), for their partner (e.g., 3/7), or were equally beneficial for both (5/5). For each distribution, participants rated how happy they thought their partner was with the money distribution (1 = extremely unhappy, 7 = extremely happy), how confident they were in their prediction of the other's happiness (1 = extremely unconfident, 7 = extremely confident), and how happy they personally were with the money distribution (1 = extremely unhappy, 7 = extremely happy). These scenarios were not tied in any way to the compensation participants got for participating, nor were participants led to believe they would.

Confederate Responses and Interpersonal Accuracy

The target's (confederate) actual happiness ratings were collected prior to the study from 26 different participants who rated their own happiness score for each of the 30 money distributions (1 = extremely unhappy, 7 = extremely happy). For each distribution, participants also described in their own words how happy they were personally with the distribution. The answers to this question were used to produce a script the confederate could use in answering the participants' queries in the main experiment. For each distribution, we averaged participants' responses to construct a baseline score of happiness. Interpersonal accuracy was measured as the absolute mean difference between participants' predicted ratings and these baseline scores, with smaller absolute differences representing larger interpersonal accuracy.

RESULTS AND DISCUSSION

Only 12 out of the 46 participants (26%) engaged in some form of perspective getting. Out of this total, six people asked a general question, for example, whether the other person would be happier with an equal split or a distribution in which the other person got more. Four people asked about 1 or 2 specific distributions, and four people used a combination of those two strategies.

Those who engaged in perspective getting had an accuracy score of M = 0.93 (SD = 0.48) and those who did not had an accuracy score of M = 0.77 (SD = 0.32). The small sample size makes statistical testing unreliable, but even from the direction of the means it becomes clear that perspective getting did not increase accuracy in this case. We did run a bootstrapped t-test to be complete in our reporting, $t(14.61) = 1.07$, $p = .302$ [−0.14; 0.46], which showed that there was no statistical difference in accuracy between those who engaged in perspective getting and those who did not.

These results illustrate that people do not easily engage in perspective getting. In this particular setting, in which people also experienced themselves how the different money distributions make one feel, people may feel confident about using projection as a strategy. Projection is actually a good strategy in cases in which there is no reason to believe that the target is different from the predictor (Ames, 2005), so there was no urgent need to engage in perspective getting. In the next study, we therefore introduced a factor that would create a difference between predictor and target, namely the presence of the predictor's explicit privileged knowledge.

Study 2

We made several changes in our second study. First, to test whether perspective getting leads to more interpersonal accuracy, we included a condition in which participants were explicitly instructed to engage in perspective getting and a control condition. Second, we used a scenario in which participants predicted how expensive a target would think different products were. This means that the perspective predictions are no longer about a target's emotional reactions, but about a target's price estimations. This was inspired by the negotiations literature in which correctly predicting the other parties' ideas about the value of things is seen as an important factor to negation success (Chambers and De Dreu, 2014). Third, we introduced privileged knowledge, which is known to stand in the way of good predictions of naïve targets (e.g., Damen, et al., 2019b; Damen et al., 2020a; Epley et al., 2004; Keysar, 1994). In these studies, predictors with privileged knowledge find it hard to inhibit this knowledge during perspective-taking, causing them to overestimate the similarity between their own and the target's perspective. Feedback that highlights this difference in perspectives, however, seems to decrease predictors' overestimation of similarities in perspectives (Damen et al., 2021). Hence, we predicted that explicitly telling perceivers that their own knowledge was privileged to them and, thus, not known to the target, would increase the need for getting the target's perspective.

Method

Participants and Design

For this study we used a 3 (*Condition*: instructed perspective getting, option for perspective getting, control) x 2 (*Knowledge*: no privileged knowledge, privileged knowledge) between-subjects design. For a medium effect size, the G*Power calculation (version 3.1.9.2) showed that we would require a sample size of at least 20 participants per experimental condition to obtain an alpha err probability of .05 and a power of .80. In total, we recruited 196 participants from the university campus to participate in a study in which they predicted another person's beliefs (women = 126, men = 69, non-binary = 1; M_{age} = 22.80, age-range = 18–57). The majority of participants was Dutch (N = 136, 69.4%). The 60 non-Dutch participants came either from other countries in Europe (N = 55, 91.7%), or from Russian (N = 1), African (N = 1), Arabic (N = 1), Malay (N = 1), or Papiamentu (N = 1) speaking countries. The majority had received or followed a Bachelor's degree (N = 167, 85.2%), followed by those who were following or had finished a Master's (N = 17, 8.7%) or High School degree (N = 12, 6.1%). The study was conducted in English, and all participants were proficient in the language of the study. Participants were randomly distributed to one of the six conditions, resulting in 63 participants partaking in the instructed perspective getting conditions (30 received privileged knowledge), 66 in the option for perspective getting conditions (34 received privileged knowledge), and 67 in the control conditions (32 received privileged knowledge). Participants received course credits for their participation.

Procedure

Participants were invited to partake in a study in which they predicted their partner's (a confederate named "Kim") beliefs about the price of ten products that were supposedly up for sale in a web store. Kim is a gender-neutral name in the Netherlands, so by using this name we made sure that participants did not picture the other participant as having a certain gender. The products were presented to participants in a randomized order. Participants were seated in soundproof cubicles where they logged onto the computer to meet up with their partner for the first time via a synchronous, text-based chat (Chatplat.com). This chat box was pulled up next to the Qualtrics™ questionnaire during the entire duration of the study and filled half of the computer screen. Upon entering the questionnaire, participants received different instructions depending on the experimental condition to which they were assigned.

Perspective Getting Conditions

Participants in the instructed perspective getting conditions were told to contact their partner (the confederate) each time before estimating their partner's price rating. The exact instructions were, "After seeing each advertised product, you will be asked several questions. Among these questions is what you think your partner thinks the price of the product is. Before answering that question, you have to talk to your partner using the chat function available. It is very important that you talk to your partner before you answer this particular question. Please, do not answer this question before you have spoken with your partner. When you ask your partner questions, you cannot ask directly for the price. Ask how he or she thinks or feels about the product." Participants in the option for perspective getting conditions were explicitly instructed that they had the opportunity to converse with the other participant via the chat box. The exact instructions were, "You may contact the other participant. After seeing each product, you will be asked several questions about the

product. Among these questions is what you think the other participant thinks the price of the item is. Before answering that question, you may talk to that other participant using the available chat function." In both perspective getting conditions, participants also read: "When you ask something about a product, always mention which product you are talking about so you avoid miscommunication. You and your partner do not have to come to an agreement. Just ask what the other person thinks about the advertised product." At the start of the experiment, participants in the perspective getting conditions were first instructed to introduce themselves to the other participant via the chat-box. The chat box was not available to participants in the control condition. These participants were only told that they had to estimate their partner's price rating for several products.

Privileged Knowledge Conditions

In the privileged knowledge conditions, participants were additionally informed that they were going to receive inside information about the product's actual prices. The exact instruction was, "We only tell you what the actual price of the products is. For every item that is for sale, you will receive the actual price of that product. Please note that ONLY YOU will receive this information. The other participant does NOT know the actual price of the products. Make sure you take a careful look at the product before answering the questions that follow it." In addition, above each product, participants read again that the price presented above the item was privileged to them and not known to the other participant. All these additional instructions were absent in the condition in which participants did not receive this privileged information.

After having read the instructions, participants were asked to introduce themselves to their partner Kim, after which Kim introduced herself to the participants. The confederate introduced herself as a fellow student and informed the participants that her experience with estimating the prices of products was limited, and that she tended to look at the quality of the product before guessing its price. She ended her introduction by stating, "Sometimes the actual price matches my estimation, but sometimes it doesn't". Hereafter, participants were confronted with one practice trial to ensure all participants followed the condition-specific instructions. For each product, participants estimated their own and their partner's price rating (a maximum of three digits were allowed), and how confident they were in their predictions of their partner's perspective (1 = not confident, 11 = very confident).

Confederate Responses and Interpersonal Accuracy

The target's (confederate) actual price ratings were collected prior to the study from one female participant who guessed the price of 15 unique accessories, electronics, and gadgets. After seeing each product, the participant additionally described in her own words what she thought of each product. The participant's thoughts were used to create a script that the confederate used in answering participants' queries in the main experiment. After collecting the target ratings, we selected 10 unique products to use in the main experiment. The prices for these products ranged from 20 to 95. Participants' interpersonal accuracy was measured as the absolute mean difference between participants' predicted price rating and the target's actual price rating (smaller absolute differences representing larger accuracy).

Results and Discussion

Descriptive Statistics

The descriptive statistics are presented in **Table 1**. The accuracy scores were not normally distributed, and the variances were somewhat unequal. However, since there were at least 30 observations in every condition, the ANOVA should be fairly robust against these violations. To be sure, we also calculated the parameter estimates with robust standard errors to see whether the results are the same.

Frequency of Perspective Getting

First, we were interested in how many people in the optional perspective getting condition actually used the chat function. Of the 66 people in that condition, only 15 (23%) asked the confederate something about all items. 25 (38%) people did not use the function at all or only for one item. Whether or not someone engaged in perspective getting did not depend on the privileged knowledge condition, $\chi^2(1) = 0.20$, $p = .655$, nor did the number of items asked about depend on the privileged knowledge condition (Mann-Whitney $U = 567$, $p = .765$). Those who had privileged knowledge, and were expected to have more difficulty in taking the other person's perspective, did not engage in more perspective getting to compensate for the difference in perspectives.

The Effect of Perspective Getting on Accuracy

We first conducted a factorial ANOVA to test whether the perspective getting condition (instructed, optional, none) and privileged knowledge (yes, no) influenced the accuracy with which the target's price estimates were predicted. We were not interested in the privileged knowledge condition for this analysis, so we included this factor mainly to see whether there was an interaction effect or whether the effect of perspective getting can be tested independently. Both perspective getting ($F(2, 190) = 4.17$, $p = .017$), and privileged knowledge ($F(1, 190) = 8.61$, $p = .004$) had main effects, but there was no significant interaction between perspective getting and privileged knowledge, $F(2, 190) = 0.77$, $p = .46$ (also no interaction parameter with robust standard errors was significant). We therefore conducted the more focused one-way ANOVA next with perspective getting as the sole independent variable. Given the unequal variances, we report the Welch statistic. For the contrasts, the comparison group was the condition in which no perspective getting was possible. There was a significant overall effect of the perspective getting condition on accuracy ($F(2, 123.56) = 4.37$, $p = .015$). In the control condition, the predicted price estimation was on average 36.73 euros ($SD = 17.69$) away from the target's actual price estimation. In the instructed perspective getting condition, the difference was not significantly smaller ($M = 31.96$,

TABLE 1 | Means and standard deviations for accuracy scores (absolute mean difference score between predicted and actual price estimates) in Study 2.

Perspective getting	With privileged knowledge			No privileged knowledge		
	M	**SD**	**n**	**M**	***SD***	***n***
Instructed	27.3	9.52	30	36.20	19.52	33
Optional	27.81	8.93	32	30.43	13.88	35
Not possible	32.84	12.05	32	40.29	21.16	35

SD = 16.10, $t(127.87) = 1.61$, $p = .110$), but in the optional perspective getting condition it was (M = 29.08, SD = 11.58, $t(114.02) = 2.96$, $p = .004$). We did not formulate a hypothesis for the difference in accuracy between the instructed and optional perspective getting conditions, so we did not perform a contrast for this. A post-hoc test with Bonferroni correction showed that they did not significantly differ ($p = .864$).

It is somewhat surprising that people who were instructed to get the other person's perspective do not perform better than people in the control condition. It could be that the instructions felt artificial and that people were more comfortable with the option to get perspective. In our next study, we therefore use a more natural instruction. Instead of asking participants to get perspective for each item separately, we allow participants to discuss all items at once.

Next, we focused on the optional condition and tested whether engaging in more perspective getting (i.e. asking about more items) is related to accuracy. Because the number of items that was asked about was not normally distributed, we calculated the Spearman rank correlation. We found that the more items someone asked about, the more accurate that person was (there was a lower discrepancy between the predicted and the actual price estimates), $\rho = -0.277$, $p = 0.024$. The 95% bootstrapped confidence interval of the correlation coefficient ranged from -49 to -0.20. This is fairly wide due to the relatively large standard deviation in the data. Still, we can be quite confident that people who ask the target about their opinion about a product (not even the direct price estimation) are better atn predicting the target's price estimation than people who do not do this.

Study 3

In our final study, we made another set of changes. First, instead of emotions or thoughts, we now let people predict a target's attitudes. Second, instead of making predictions about an unknown target, the target was now a friend of the participant. Third, we tested whether increasing people's awareness of the benefit of using the perspective getting strategy to predict their friend's attitudes increases the use of this strategy. We expected this because even though people sometimes discount advice they receive, if it comes from a credible source they often follow it (Bonaccio and Dalal, 2006). To investigate people's propensity to engage in perspective getting, and to explore the strategy's impact on interpersonal accuracy, we conceptually replicated Eyal et al. (2018) design. Instead of romantic couples, we invited befriended dyads to partake in a study that was conducted online via an online video-conferencing tool. As in Eyal et al. (2018), we asked one party to predict their friend's attitude on 21 opinion statements (Hoch, 1987). Besides allocating dyads to an instructed perspective getting and control condition, we also allocated dyads to a condition in which they had the option to converse with one another before predicting the target's attitudes. Moreover, half of these dyads were made aware of the benefits of perspective getting over perspective-taking strategies. We expected that perspective getting would positively affect interpersonal accuracy. More specifically, we tested the hypothesis that people who engage in perspective getting (whether they are instructed or given the option to do so) will predict more statements accurately than people who cannot engage in perspective getting (control). Furthermore, we expect that making people aware of the benefits of perspective getting increases the use of perspective getting, and that, per consequence, people will predict more statements correctly. Finally, we test the hypothesis that people who are more confident that they can predict their friend's attitudes are less likely to engage in perspective getting.

Method

Participants and Sample Size

The effect sizes found in Eyal et al. (2018) varied between medium and high effects. Eyal et al. (2018) allocated 25 participants to each of the experimental conditions, and our G*Power calculation showed that we would require at least 19 (high effect) to 45 (medium effect) participants per experimental condition to obtain an alpha err probability of .05 and a power of 80. In total, we were able to invite 160 befriended dyads to partake in a study that was conducted online via the online video-conferencing tool Zoom (Predictors: 49 male, 111 female, M_{age} = 24.96, age-range 18-57; Targets: 63 male, 97 female, M_{age} = 25.38, age range 13-68). On average, predictors reported being friends with their partner for a period of 8 years and 2 months (SD = 7.77 years). When dyads signed up for the study, we randomly allocated them to one of the four conditions. Our analyses are based on 37 participants in the control, 41 in the instructed perspective getting condition, 40 in the option for perspective getting with awareness intervention condition, and 42 in the option for perspective getting without awareness intervention condition. This means that per condition, we allocated at least 17 to 20 persons more per experimental condition than Eyal et al. (2018). Participants were recruited through various social media channels of the researchers. The majority of participants was Dutch (N = 142, 88.8%). The 18 non-Dutch participants came either from other countries in Europe

(N = 15, 9.4%), or from Russian (N = 1), Indonesian (N = 1), or Vietnamese (N = 1) speaking countries. The majority had completed an academic education (N = 79, 49.4%), followed by a higher professional (N = 37, 23.1%), a vocational (N = 24, 15.0%), or a pre-university education (N = 16, 10%). Other education levels were senior general secondary school (N = 2), secondary school (N = 1), and undefined (N = 1). This study was conducted either in English or in Dutch, depending on the type of language participants felt most comfortable to converse in. To ensure participants had sufficient time to read the study's information and sign their consent, dyads received an information letter in which they were assigned the role of either the predictor or the target one day for their participation. This also ensured that the time the dyads were left alone in the video-call would not be used to read the instructions. Predictors were informed that the next day, during the experiment, they were going to estimate their friend's (the target) attitudes on 21 opinion statements (Hoch, 1987) that were attached to the email.

Procedure and Measurements

Dyads met each other and the experimenter in an online video-call that was administered in Zoom. Instead of conversing through a chat platform, dyads now conversed via video-call. In Zoom, the experimenter ensured that dyads could hear and see one another, that they were not seated in the same room, and that they were preferably logged on from another location before continuing the experiment. Then, the experimenter informed those participants who had been randomly assigned the role of the predictor that they were going to predict their friend's attitude on the 21 declarative statements that were attached to their invitation email. These statements covered a wide range of topics, including politics, personality, nutrition, and social issues that were taken from Hoch (1987), and adapted to students' current lifestyle and situation (e.g., "I like to pay contactless for everything I buy", "The corona crisis is the biggest pandemic in history around the world today"). Hereafter, each participant received a personalized link to Qualtrics™ that they opened after their sound was muted and their webcams were turned off in Zoom. In this questionnaire, dyads gave their informed consent for recording their audio in Zoom. Subsequently, we measured predictors' confidence in their perspective-taking ability by asking them to indicate: "how many of these statements do you think you will predict correctly?" (0–21). After that, they were instructed to return to the Zoom conversation.

Perspective Getting Conditions

Once back in the Zoom, dyads received further instructions based on the condition to which they had been randomly assigned. The instructions for both the instructed perspective getting and control conditions were directly replicated from Eyal et al. (2018). In the instructed perspective getting condition, predictors were explicitly told to take approximately 5 min to ask their friend about the extent to which they agreed or disagreed with each of the 21-opinion statements. Dyads received the following instructions, "In this experiment, we will examine your ability to accurately perceive your friend's opinions to the 21 opinion statements. Please take approximately 5 min to ask your friend about the extent to which they agree or disagree with each of the topics on the list, trying to get a sense of the range of your friend's opinions. Your partner might strongly agree with some statements, somewhat agree with others, and they may strongly disagree with others. Once you have done that, we would like you to use the information you got from your friend to predict the extent to which your friend would agree or disagree with these statements. Please open the list with the 21 opinion statements. I (the experimenter) will now leave the video room for 5 min and give you both time to discuss." Next, the experimenter left the video call by muting her audio and turning her camera off for approximately 5 min to give the participants the opportunity to discuss the statements and engage in a conversation with each other. After 5 min, the experimenter rejoined the video call and instructed the participants to mute their audio and turn their camera off to make sure that consultation between the participants was avoided when they re-entered the online questionnaire. The explicit instructions to discuss the target's opinion on each of the 21 statements were absent in the option for perspective getting conditions, and in the control condition. In the option for perspective conditions, the experimenter also left the video call by muting her audio and turning off her camera for approximately 5 min to give the friends the opportunity to engage in a conversation with each other. In the control condition, participants did not receive this 5-min opportunity to discuss the target's attitudes. Instead, dyads were informed that "In this experiment, we will examine your (the predictor's) ability to accurately perceive your friend's (the target's) opinions to the 21 opinion statements you already received in the information mail. We would like for you to use whatever strategy you think is best to predict your friend's opinion to each of the statements."

Awareness Intervention

The awareness intervention was administered only in the option for perspective getting conditions, and only half of the dyads received the intervention before they had the opportunity to discuss the target's attitudes. In both option for perspective getting conditions, participants were informed about the different perspective-taking strategies people generally use when estimating another person's opinion. They received the following instruction, "In this experiment, we will examine your (the predictor's) ability to accurately judge your friend's (the target's) attitude on 21 opinion statements. You can use different strategies to predict your friend's opinion. Let us summarize some of the strategies with an example using apples. Imagine you want to predict whether your friend prefers red apples over green ones. To infer your friend's opinion, you can either 1) Use your own preferences to predict those of your friend: "I prefer red apples over green apples, so I think my friend will also prefer the red ones over the green ones too"; 2) Apply general knowledge about what other people who are similar to your friend like: "Most people prefer red apples, so I think my friend is most likely to prefer red apples over green ones too"; 3) Let situational circumstances guide your prediction: "Most of the apples in the supermarket are red, so I think my friend will prefer the

red apples over the green ones"; or 4) Directly ask your friend which type of apples he or she prefers". In the awareness condition, participants were additionally informed of the benefit of perspective getting (strategy 4) over the other three, "Now, we know from previous research that not all of these strategies help us to understand our friend's perspective. The main reason for this is that, for the first three summarized strategies, we often overestimate the extent to which our friend thinks (or feels) like we do or like a stereotypical group does. Hence, if we use these strategies, we often fail to appreciate that our friend can have unique, different preferences. Therefore, research has shown that the most effective strategy to judge our friend's perspective is to directly ask our friend about their opinions before we try to predict them". After these instructions, the experimenter left the dyads alone for 5 min after saying and stressing they could use whatever strategy they thought was best to get a sense of their friend's attitudes.

After 5 min (or immediately after the general instruction in the control), participants were redirected to the questionnaire. The experimenter made sure that participants' audio was muted and that their webcams were turned off before they re-entered the questionnaire. The predictors predicted their friend's (the target) opinions on each of the 21 opinion statements (1 = strongly disagree, 7 = strongly agree), followed by how confident they were about the accuracy of their prediction (1 = not at all confident, 11 = extremely confident). After predicting their friend's attitudes, predictors reported their own attitude for each of the 21 statements in succession (1 = strongly disagree, 7 = strongly agree). At the end of the survey, predictors indicated: "how many of the statements do you think you predicted correctly?" (0–21), after which they indicated the length of their friendship with the target (in years and months), and filled in their demographics. Finally, participants were debriefed and thanked for their participation.

Coding of Perspective Getting

The conversations were transcribed and coded by three coders. For each statement, it was coded whether the predictor asked the target about their attitudes or not. All data were coded by two coders and disagreements were solved by discussion. There were no cases in which this could not be accomplished.

Interpersonal Accuracy

As in the previous studies, interpersonal accuracy was measured as the absolute mean difference between predictors' estimation of a target's attitude and the target's actual attitude, with smaller absolute mean differences representing larger accuracy.

Results and Discussion

Frequency of Getting Perspective

First, we were interested in how many people in the optional perspective getting conditions asked their friend about the statements. Of the 80 people in those conditions, 56 (70%) engaged in perspective getting by actively asking information about at least one statement. Ten people (12.5%) asked something about all statements. On average, participants asked about 8.75 of the 21 statements. Remember that it was made very obvious to people that they were allowed to talk about the statements, and they were left alone in the online meeting. It is therefore quite remarkable that still, 30% of the people did not start talking about the statements, and that, on average, not even half of the statements were discussed and checked.

Next, we tested whether the awareness intervention increased the use of perspective getting. We conducted a chi-square test with the intervention (yes/no) as independent variable, and perspective getting (yes/no) as dependent variable. There was no significant relationship between the intervention and whether people engaged in perspective getting, $\chi^2 1) = 0.95$, $p = 0.329$. We also conducted a *t*-test with the intervention (yes/no) as independent variable, and the number of statements asked about as dependent variable. The number of statements was not normally distributed, so the *p*-value may not be reliable, and we will rely on the bootstrapped confidence intervals to decide whether there is a significant difference. We found that people who were informed that perspective getting is the best strategy asked about 10.33 (SD = 7.92) items, and people who did not get that information asked about 7.18 (SD = 7.88) items. This difference was not significant ($t(78) = 1.78$, $p = 0.078$, $d = 0.40$), as the confidence interval ranged from −0.4 to 6.5. Thus, our intervention, in which we told participants that perspective getting is a good strategy, did not significantly increase perspective getting.

The Effect of Perspective Getting on Accuracy

The descriptive statistics of the accuracy scores are presented in **Table 2**. All assumptions for parametrical testing were met. We first conducted a one-way ANOVA to test whether the perspective getting condition (instructed, optional (with intervention), optional (without intervention), control) influenced the accuracy with which the friend's attitudes were predicted.

Based on our hypotheses, we conducted three planned contrasts: 1) the three perspective getting conditions against the control condition, 2) the instructed condition against the optional conditions, and 3) the optional condition with intervention against the optional condition without intervention. There was a significant overall effect of the type of condition on accuracy ($F(3, 156) = 12.53$, $p < .001$). The first contrast was significant, $t(156) = 4.81$, $p < .001$, ds = 0.90. People who could engage in perspective getting were more accurate than people who could not engage in perspective getting. The second contrast was also significant, $t(156) = 3.54$, $p = 0.001$, ds = 0.68. In line with our hypothesis, we found that people who were instructed to engage in perspective getting were more accurate than people who had the option to engage in perspective getting. The third contrast was not significant, $t(156) = 1.37$, $p = 0.173$, ds = 0.30, showing that there were no significant differences in interpersonal accuracy between the people who were made aware of the benefit of perspective getting and those who were not.

Next, we focused on the optional condition and checked whether engaging in more perspective getting is related to differences in accuracy. Because the number of statements that was asked about was not normally distributed to performed a Spearman rank correlation. We found that people are more accurate (the discrepancy between their prediction and their friend's actual attitude is smaller) if they ask about more

TABLE 2 | Means and standard deviations for accuracy scores (absolute mean difference score between predicted and actual attitudes) in Study 3.

	M	*SD*	*n*
Instructed	0.86	0.34	41
Optional (with intervention)	1.05	0.38	40
Optional (no intervention)	1.16	0.36	42
Not possible	1.35	0.37	37

statements, $\rho = -0.70$, $p < .001$. The 95% bootstrapped confidence interval of the correlation coefficient ranged from −0.78 to −0.60. This is a relatively narrow interval and the lower limit is far from zero. Based on this we can be quite confident that people who ask their friend about a larger number of their attitudes are better at predicting these attitudes than people who use this opportunity to a lesser extent.

Confidence as a Barrier to Perspective Getting?

Finally, we tested whether people who are more confident that they can predict their friend's attitudes are less likely to engage in perspective getting. Again, we only did this for people in the optional condition. We used the estimation given by the participant at the start of the study of how many items they would answer correctly as independent variable and the number of statements that they asked about as dependent variable. Because the latter variable was far from being normally distributed (with two peaks at 0 and 21), we conducted a Spearman's rank correlation test. We found a significant negative correlation, $\rho = -0.22$, $p = 0.047$, 95% CI [−0.44, −0.01], indicating that people who thought that they knew more of their friend's attitudes asked about fewer of the statements. Confidence was not related to accuracy, $\rho = 0.04$, $p = 0.70$, 95% CI [−0.19, 0.28], so the lower engagement in perspective getting was not justified. This null effect is intriguing because it suggests that there is another factor that is related to confidence and accuracy. If confidence is related to perspective getting, and perspective getting is related to accuracy, confidence should also be directly related to accuracy. There must be another variable that cancels out the effect. This could be, for example, actual knowledge. If actual knowledge is positively related to confidence and accuracy its effect would be opposite to the effect of perspective getting, which is negatively related to confidence and positively to accuracy. In future studies, it may therefore be interesting to include actual knowledge.

Insights From the Conversations

When coding the conversations, we noticed that there was a certain reluctance to engage in perspective getting. Some people thought that it was too straightforward to simply ask the other person about their attitudes, "The easiest now is that I ask all questions, but I think that is a little bit stupid." Other people thought perspective getting was redundant. "I knew all of this before you said it" or "Ok, do we need to discuss more or not? Because I think that I can predict the rest of it pretty well." Moreover, others claimed that they just did not want to know, but wanted to do the prediction on their own:

> P: "After the corona crisis we should keep working from home if possible." I was wondering, what would you think about this?
>
> T: Yes, do you want to know?
>
> P: Yes, I don't know actually. I don't want to know it, but it is a question that I can't . . . No, I don't want to know. I will answer it for you.
>
> P: Or are we supposed to. . .. Are we the typical case that we are so naive and think we can predict each other? That even if we sit together for 5 min and can ask all questions, that we still don't do it? That we still think we know better. I think that's what we are.

What we also noticed is that dyads took ample time to discuss the validity of each of the previously introduced techniques to gauge another person's perspective. In addition, even the targets were often very confident that their predictor would be able to take their perspective on their own:

> T: But you can accurately guess my opinions, right?
>
> P: Yes, absolutely. But, look, she (the experimenter) wants me to use a technique. Though, I think I can answer all those questions for you. But, I have to use a strategy, so I'm going to do it with one of those four techniques.
>
> T: But with the latest technique, then you can just ask me what I would do?
>
> P: Yes, I think so, but let's not do that.
>
> T: No.

Most predictors even used projection as a strategy while discussing the items with their friend:

> P: "I would rather live in a big city than a small town." Yes. "The university has provided good education for homeschooling during the corona crisis." That's an ambiguous question. It depends on the professors and everything. But, okay, in general.
>
> T: If you want to ask me, I could tell you what I think.
>
> P: I think you would agree.
>
> T: Alright. Okay.

Another Example

> P: "I easily make contact with other people." Yes, absolutely. "Information from advertisements helps me make a better purchasing decision." Ehm. . .
>
> T: Hey, I have to answer that.

These quotes show that there are many reasons why people do not engage in perspective getting, even after it was pointed out to them that it is the best strategy.

GENERAL DISCUSSION

In this study, we investigated the likelihood of people getting their interlocutor's perspective. We examined this by investigating people's proneness to ask about their interlocutor's feelings (study 1), beliefs (study 2), and attitudes (study 3) before predicting them. Furthermore, we explored whether the perspective getting strategy increases interpersonal accuracy and examined the barriers to the use of perspective getting. In this line of research, we only examined explicit perspective differences (study 2) and one's confidence (study 3) as possible reasons for people to (mis)select the most appropriate strategy to judge other people's perspectives. However, the list of obstacles to perspective getting might be extensive, and offers, therefore, an important starting point for future studies.

Frequency of Perspective Getting

Our findings showed that perspective getting is a technique that people do not readily use. Results from the first study indicate that when predicting another person's feelings about money distributions, the majority of people will rely on their own judgment rather than asking the other person. Moreover, those who did engage in some form of perspective getting did not do so thoroughly enough to be beneficial to interpersonal accuracy. Similar results were found in the second study. When predicting another person's price estimates, the majority of people did engage in some form of perspective getting, but only 23% of all people did this thoroughly by asking their interlocutor's perspective for all of the items. In the third study, participants predicted their friend's attitudes. In this final study, we found that, although the majority of people did engage in perspective getting, again only a small number did this thoroughly. On average, people inquired about their friend's opinion on less than half of all the statements they were able to discuss and check.

In general, our findings are in line with previous research suggesting that people will often choose their own intuition rather than other, non-intuitive alternatives when gauging others' interpretations (Denes-Raj and Epstein, 1994; Kirkpatrick and Epstein, 1992, see also "intuition bias" in; Simmons and Nelson, 2006). People seem to prefer using their own judgments rather than using more effective (and more obvious) methods of gaining interpersonal accuracy, such as getting the other person's perspective by asking about it.

But why is perspective getting not an attractive strategy? There are several explanations for our findings that people do not use it thoroughly. The first one lies in the methodology of the studies. In two of the studies, confederates were used to represent either an average target person (study 1) or an actual person (study 2) whose perspective participants were asked to predict. There are pros and cons to each method used. The disadvantage of the 'average response' is that the prediction is not about an actual person. We took away that concern by using an actual person in study 2, and the confederate in that study used that person's actual responses to respond to all participants' queries. Still, the setup in the first two studies is a bit artificial. Also because in study 2 participants were not allowed to ask for the price directly. That is why we used real friends and real conversations in the final study. It is still not a complete 'real-life' situation, because one rarely has to know another person's exact attitudes. This setup may have felt like a game to participants and perspective getting may have felt like cheating. To take away that concern, we stressed that perspective getting is a valid strategy, but one explanation for the incomplete use of it may still be that it did not 'feel right'.

The second reason people may be reluctant to use non-intuitive techniques could be that they are not aware of the benefits other strategies have on increasing interpersonal accuracy. Eyal et al. (2018) found that people overestimate the effectiveness of perspective taking compared to perspective getting and that they, therefore, have a misbelief about which strategy was better for predicting the attitudes of their partner. Eyal et al. (2018) suggest that this misconception may be preventing people from selecting the most optimal strategy–getting a person's perspective. We tested whether pointing out that perspective getting is the best strategy would increase its use, but we found no indication for this. Thus, people's reluctance toward using perspective getting cannot be overcome by simply informing people of its merits.

Interestingly, results from our study indicate that even when people do engage in perspective getting they do not do so in depth. This suggests there may be a limit on how willing people are to ask continuously about other people's perspectives. Fiske and Taylor (1984) suggest that people are, by design, "cognitive misers". Previous research shows that we naturally use 'cognitive shortcuts', and although these shortcuts sometimes lead us to wrong answers, they tend to provide useful answers often enough to be beneficial and utilized by default in most circumstances (Judd and Park, 1993). Although interpersonal understanding is an important concept in close relationships and a predictor of relationship satisfaction (Reis, et al., 2017), it is oftentimes not based on having actual knowledge about the other person (Pollmann and Finkenauer, 2009). It is often sufficient to give the partner the feeling that they are understood by making validating statements. Therefore, it may not always be necessary to try to achieve a perfect understanding of others' feelings and attitudes. In situations like the ones in our studies, perspective getting may take too much effort in relation to a relatively little gain, but in situations in which accurate understanding is more important, like in conflict situations, it may be more valuable. Future studies should investigate whether people are more likely to engage in perspective getting if more is at stake.

Another explanation for our findings may be found in the gift-giving literature. Asking what gift someone would like, or simply buying a gift from a registry, is not only thought of as less thoughtful by the giver, but also diminishes the giver's opportunity to show how well they know the other person (Gino and Flynn, 2011; Ward and Broniarczyk, 2016). Perhaps predicting another person's perspective is seen as a challenge, causing people to avoid perspective getting as they feel it is the wrong way to show their empathic skills. It could be suggested

that our findings reflect those found in gift-giving literature. In study 3, insights from the conversations showed that some people did not want to know their interlocutor's perspective, but wished to predict it on their own. It could be argued that the ability to guess someone's opinion accurately could be viewed as more beneficial to the relationship than simply asking - similar to the way people feel about predicting what gift would suit the receiver rather than asking what they want. Our research did not ask follow-up questions about why people used a specific method. Future research may wish to explore the reasons people give for not wishing to engage in perspective getting. After all, a good way to understand the mind of others is to simply ask. This could be done through qualitative research and investigating if there is a common theme or reason that people give when they do not ask their interlocutor about their perspective.

Finally, the setup of our study may have reduced perspective getting. Our first and second study were conducted in English, which was not the native language of our participants. Processing a second language is more effortful than processing a native language (e.g. Ardila, 2003). It may be that our participants' working memory capacity was reduced and there was less capacity left to engage in perspective getting. Relatedly, research on the foreign language effect shows that decisions made in a foreign language differ from those made in the native language (Keysar et al., 2012), although these findings remain somewhat unclear (Hayakawa et al., 2019). So the results of Study 1 and Study 2 may underestimate people's tendency to engage in perspective getting. It should be noted, however, that all our participants were fluent in English and they use it on a daily basis because their study program is English. Also, in the third Study participants were allowed to have the conversation in Dutch, and we still find low to moderate levels of perspective getting there. In any case, more studies are needed to investigate how likely people are to get perspective in different context and how we can stimulate it.

The Effects of Perspective Getting on Accuracy

The findings from our studies (2 and 3) showed that people who engaged in perspective getting were more accurate at predicting others' price estimates (study 2) and attitudes (study 3) than those who did not engage in perspective getting. Our findings thus support previous research proposing that perspective getting leads to more interpersonal accuracy than other intuitive techniques (Eyal et al., 2018).

We did find a different effect of instructed perspective getting on interpersonal accuracy between study 2 and 3. In study 2, instructed perspective getters were not more accurate in their predictions than those who were given the option to do so. This in contrast to study 3, where predictors instructed to get their friend's perspective for all statements were more accurate than those who could choose to discuss the statements. We speculate that this difference occurred because of the somewhat artificial chat-based perspective getting instructions in study 2. Whereas predictors were instructed to get the target's perspective for each item separately in study 2, dyads in study 3 could freely discuss each attitudinal statement at the start of the experiment. Of course, this assumption can only be tested by comparing different perspective getting procedures in one experiment.

In our study, interactions took place online via synchronized chat conversations and online video meeting platforms. This produced a unique insight into the effectiveness of different strategies to obtain interpersonal accuracy without face-to-face interactions. Our research shows that perspective getting is still effective in leading to more accurate results, even when communication is done via online modes, where the transmission of non-verbal and social cues (study 1 and study 2), or direct eye-contact (study 3) is limited. We encourage future researcher to examine how effects of perspective getting on accuracy are related to communicative settings in which interlocutors are able to rely on a variety of cues to build interpersonal report, and simultaneously or consequently, interpersonal understanding (e.g., Damen, et al., 2020b).

Confidence as a Barrier to Perspective Getting

In this study, we also tested whether people who are more confident that they could predict their friend's attitudes would be more or less likely to engage in perspective getting (study 3). As expected, the results showed that people who were confident they knew more of their friend's attitudes asked about fewer of the statements. However, confidence was not related to accuracy, so the lower engagement in perspective getting was not justified. This finding supports previous research that indicated that people are often overconfident in their ability to predict others' perspective, and that this overconfidence prevents the willingness to learn or to improve (Eyal et al., 2018; Epley and Eyal, 2019; Ickes, 2003; Jacoby et al., 1994; Oskamp, 1965; Realo et al., 2003; Ryback, 1967; Swann and Gill, 1997; Thomas et al., 1997; see overconfidence effect in; Adams and Adams, 1961). Therefore, this study shows that (over) confidence acts as a barrier to perspective getting, and the less people asked, the less accurate their inferences about their interlocutor's perspective were. What we still do not know is whether perspective getting reduces interlocutors' overconfidence once they learn more about their communicative partner's perspective. Further studies could examine how likely people are to get their interlocutor's perspective after receiving feedback that makes them lose confidence in their ability to accurately judge other minds.

CONCLUSION

This study has revealed that although directly asking one's interlocutor about their perspective is a good strategy to increase interpersonal accuracy, this perspective getting is not readily adopted. Most people do not ask others in

detail about others' perspectives, not even when they are clearly given the option to do so, or are explicitly advised on the benefits of doing so. This hesitation to use the perspective getting strategy highlights that there may be many barriers to perspective getting and interpersonal accuracy as a whole.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found here: https://osf.io/xt3n6/?view_only=df80ccd4ea3b4eccaaa2e97c0078e91f.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Review Board (ERB) of the Tilburg School of Humanities of Tilburg University. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed to writing the manuscript. DD planned the studies and supervised the data collection. MP analyzed the data. T-LG contributed to the data collection. All authors approved the final version of the paper.

ACKNOWLEDGMENTS

We would like to thank Giordano Tuvo, Saya Hoelasie, Tessa Guntlisbergen, Yvette Laros, Lotte Boone, and Kim Schneider for their significant input while conducting (parts of) the studies that are presented in this paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.611187/full#supplementary-material

REFERENCES

Adair, W. L., and Brett, J. M. (2005). The Negotiation Dance: Time, Culture, and Behavioral Sequences in Negotiation. *Organ. Sci.* 16, 33–51. doi:10.1287/orsc.1040.0102

Adams, J. K., and Adams, P. A. (1961). Realism of Confidence Judgments. *Psychol. Rev.* 68 (1), 33–45. doi:10.1037/h0040274

Ames, D. R. (2005). "Everyday Solutions to the Problem of Other Minds," in *Other Minds: How Humans Bridge the divide between Self and Others*. Editors B. F. Malle and S. D. Hodges (New York: Guilford Press).

Ames, D. R. (2004). Inside the Mind Reader's Tool Kit: Projection and Stereotyping in Mental State Inference. *J. Personal. Soc. Psychol.* 87 (3), 340–353. doi:10.1037/0022-3514.87.3.340

Ames, D. R., and Kammrath, L. K. (2004). Mind-reading and Metacognition: Narcissism, Not Actual Competence, Predicts Self-Estimated Ability. *J. Nonverbal Behav.* 28 (3), 187–209. doi:10.1023/B:JONB.0000039649.20015.0e

Ardila, A. (2003). Language Representation and Working Memory with Bilinguals. *J. Commun. Disord.* 36 (3), 233–240. doi:10.1016/S0021-9924(03)00022-4

Bacharach, S. B., and Lawler, E. J. (1981). Power and tactics in bargaining. *ILR Rev.* 34 (2), 219–233.

Barry, C. A., Bradley, C. P., Britten, N., Stevenson, F. A., and Barber, N. (2000). Patients' Unvoiced Agendas in General Practice Consultations: Qualitative Study. *Britisch Med. J.* 320 (7244), 1246–1250. doi:10.1136/bmj.320.7244.1246

Beckman, H. B., and Frankel, R. M. (1984). The Effect of Physician Behavior on the Collection of Data. *Ann. Intern. Med.* 101 (5), 692–696. doi:10.7326/0003-4819-101-5-692

Bonaccio, S., and Dalal, R. S. (2006). Advice Taking and Decision-Making: An Integrative Literature Review, and Implications for the Organizational Sciences. *Organizational Behav. Hum. Decis. Process.*, 101., 2006 Behaviors, trust, and goal achievement in a win-win negotiating role play, Group and Organization Management, 2199520486501, 127–151. doi:10.1016/j.obhdp.2006.07.001

Butler, J. K., Jr. (1995). Behaviors, trust, and goal achievement in a win-win negotiating role play. *Group Organizat. Manag.* 20 (4), 486–501. doi:10.1177/1059601195204006

Butler, J. K., Jr. (1999). Trust Expectations, Information Sharing, Climate of Trust, and Negotiation Effectiveness and Efficiency. *Group Organ. Management* 24 (2), 217–238. doi:10.1177/1059601199242005

Camerer, C., Loewenstein, G., and Weber, M. (1989). The Curse of Knowledge in Economic Settings: An Experimental Analysis. *J. Polit. Economy* 97 (5), 1232–1254. doi:10.1086/261651

Chambers, J. R., and De Dreu, C. K. (2014). Egocentrism drives misunderstanding in conflict and negotiation. *J. Experim. Social Psychol.* 51, 15–26. doi:10.1016/j.jesp.2013.11.001

Damen, D., Van Amelsvoort, M., Van der Wijst, P., and Krahmer, E. (2019a). Changing Views: the Effect of Explicit Perception-Focus Instructions on Perspective-Taking. *J. Cogn. Psychol.* 31 (3), 353–369. doi:10.1037/xlm000083010.1080/20445911.2019.1606000

Damen, D., Van Amelsvoort, M., Van der Wijst, P., Pollmann, M., and Krahmer, E. (2021). Lifting the Curse of Knowing: How Feedback Improves Perspective-Taking. *Q. J. Exp. Psychol.* 74, 1054–1069. doi:10.1177/1747021820987080

Damen, D., Van der Wijst, P., Van Amelsvoort, M., and Krahmer, E. (2020a). Can the Curse of Knowing Be Lifted? the Influence of Explicit Perspective-Focus Instructions on Readers' Perspective-Taking. *J. Exp. Psychol. Learn. Mem. Cogn.* 46 (8), 1407–1423. doi:10.1037/xlm0000830

Damen, D., Van der Wijst, P., Van Amelsvoort, M., and Krahmer, E. (2019b). Perspective-Taking in Referential Communication: Does Stimulated Attention to Addressees' Perspective Influence Speakers' Reference Production?. *J. Psycholinguist Res.* 48 (2), 257–288. doi:10.1007/s10936-018-9602-7

Damen, D., Van der Wijst, P., Van Amelsvoort, M., and Krahmer, E. (2020b). The Effect of Perspective-Taking on Trust and Understanding in Online and Face-To-Face Mediations. *Group Decis. Negot.* 29, 1121–1156. doi:10.1007/s10726-020-09698-8

Decety, J., and Sommerville, J. A. (2003). Shared Representations between Self and Other: A Social Cognitive Neuroscience View. *Trends Cogn. Sci.* 7 (12), 527–533. doi:10.1016/j.tics.2003.10.004

De Dreu, C. K., Beersma, B., Stroebe, K., and Euwema, M. C. (2006). Motivated information processing, strategic choice, and the quality of negotiated agreement. *J. Person. Soc. Psychol.* 90 (6), 927–943. doi:10.1037/0022-3514.90.6.927

De Dreu, C. K., Koole, S. L., and Steinel, W. (2000). Unfixing the fixed pie: a motivated information-processing approach to integrative negotiation. *J. Person. Soc. Psychol.* 79 (6), 975–987. doi:10.1037/0022-3514.79.6.975

Denes-Raj, V., and Epstein, S. (1994). Conflict between Intuitive and Rational Processing: When People Behave against Their Better Judgment. *J. Personal. Soc. Psychol.* 66 (5), 819–829. doi:10.1037/0022-3514.66.5.819

Einhorn, H. J., and Hogarth, R. M. (1978). Confidence in Judgment: Persistence of the Illusion of Validity. *Psychol. Rev.* 85 (5), 395–416. doi:10.1037/0033-295x.85.5.395

Elfenbein, H. A., Curhan, J. R., Eisenkraft, N., Shirako, A., and Brown, A. (2010). "Why are some negotiators better than others? Opening the black box of bargaining behaviors," in Paper presented at the 23rd Annual Meeting of the International Association for Conflict Management, Cambridge, MA. Available at: http://ssrn.com/abstract=1612491.

Epley, N., and Eyal, T. (2019). Through a Looking Glass, Darkly: Using Mechanisms of Mind Perception to Identify Accuracy, Overconfidence, and Underappreciated Means for Improvement. *Adv. Exp. Soc. Psychol.* 60, 65–120. doi:10.1016/bs.aesp.2019.04.002

Epley, N., Keysar, B., van Boven, L., and Gilovich, T. (2004). Perspective Taking as Egocentric Anchoring and Adjustment. *J. Personal. Soc. Psychol.* 87 (3), 327–339. doi:10.1037/0022-3514.87.3.327

Eyal, T., Steffel, M., and Epley, N. (2018). Perspective Mistaking: Accurately Understanding the Mind of Another Requires Getting Perspective, Not Taking Perspective. *J. Personal. Soc. Psychol.* 114 (4), 547–571. doi:10.1037/pspa0000115

Fiske, S. T., and Taylor, S. E. (1984). *Social Cognition*. Reading, MA: Addison-Wesley.

Galinsky, A. D., Ku, G., and Wang, C. S. (2005). Perspective-Taking and Self-Other Overlap: Fostering Social Bonds and Facilitating Social Coordination. *Group Process. Intergroup Relations* 8, 109–124. doi:10.1177/1368430205051060

Gilovich, T., Medvec, V. H., and Savitsky, K. (2000). The Spotlight Effect in Social Judgment: An Egocentric Bias in Estimates of the Salience of One's Own Actions and Appearance. *J. Personal. Soc. Psychol.* 78 (2), 211–222. doi:10.1037/0022-3514.78.2.211

Gino, F., and Flynn, F. J. (2011). Give Them what They Want: the Benefits of Explicitness in Gift Exchange. *J. Exp. Soc. Psychol.* 47 (5), 915–922. doi:10.1016/j.jesp.2011.03.015

Gopnik, A., and Wellman, H. M. (1994). "The Theory Theory," in *"The Theory Theory," in* Domain Specificity In Cognition And Culture. Editors L. Hirschfeld and S. Gelman (New York: Cambridge University Press), 257–293. doi:10.1017/cbo9780511752902.011

Harsanyi, J. C. (1962). Bargaining in ignorance of the opponent's utility function. *J. Conf. Resol.* 6 (1), 29–38.

Hayakawa, S., Lau, B. K. Y., Holtzmann, S., Costa, A., and Keysar, B. (2019). On the Reliability of the Foreign Language Effect on Risk-Taking. *Q. J. Exp. Psychol.* 72 (1), 29–40. doi:10.1177/1747021817742242

Hoch, S. J. (1987). Perceived Consensus and Predictive Accuracy: The Pros and Cons of Projection. *J. Personal. Soc. Psychol.* 53 (2), 221–234. doi:10.1037/0022-3514.53.2.221

Hyder, E. B., Prietula, M. J., and Weingart, L. R. (2000). Getting to best: efficiency versus optimality in negotiation. *Cognit. Sci.* 24 (2), 169–204. doi:10.1207/s15516709cog2402_1

Ickes, W. (2003). *Everyday Mind Reading: Understanding what Other People Think and Feel*. Amherst, NY: Prometheus Books.

Ickes, W., Stinson, L., Bissonnette, V., and Garcia, S. (1990). Naturalistic Social Cognition: Empathic Accuracy in Mixed-Sex Dyads. *J. Personal. Soc. Psychol.* 59 (4), 730–742. doi:10.1037/0022-3514.59.4.730

Jacoby, L. L., Bjork, R. A., and Kelley, C. M. (1994). "Illusions of Comprehensions and Competence,". *In Learning, Remembering, Believing: Enhancing Individual and Team Performance*. Editors D. Druckman and R. A. Bjork (Washington, DC: National Academy Press), 57–80.

Judd, C. M., and Park, B. (1993). Definition and Assessment of Accuracy in Social Stereotypes. *Psychol. Rev.* 100 (1), 109–128. doi:10.1037/0033-295X.100.1.109

Kenny, D. A., and Acitelli, L. K. (2001). Accuracy and Bias in the Perception of the Partner in a Close Relationship. *J. Personal. Soc. Psychol.* 80 (3), 439–448. doi:10.1037/0022-3514.80.3.439

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 (1), 32–38. doi:10.1111/1467-9280.00211

Keysar, B., Barr, D. J., and Horton, W. S. (1998a). The Egocentric Basis of Language Use. *Curr. Dir. Psychol. Sci.* 7 (2), 46–49. doi:10.1111/1467-8721.ep13175613

Keysar, B., Hayakawa, S. L., and An, S. G. (2012). The Foreign-Language Effect. *Psychol. Sci.* 23 (6), 661–668. doi:10.1177/0956797611432178

Keysar, B. (1994). The Illusory Transparency of Intention: Linguistic Perspective Taking in Text. *Cogn. Psychol.* 26, 165–208. doi:10.1006/cogp.1994.1006

Keysar, B., Lin, S., and Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition* 89 (1), 25–41.

Kirkpatrick, L. A., and Epstein, S. (1992). Cognitive-experiential Self-Theory and Subjective Probability: Further Evidence for Two Conceptual Systems. *J. Personal. Soc. Psychol.* 63 (4), 534–544. doi:10.1037/0022-3514.63.4.534

Krueger, J., and Clement, R. W. (1994). The Truly False Consensus Effect: an Ineradicable and Egocentric Bias in Social Perception. *J. Personal. Soc. Psychol.* 67 (4), 596–610. doi:10.1037/0022-3514.67.4.596

Kunda, Z., and Thagard, P. (1996). Forming Impressions from Stereotypes, Traits, and Behaviors: a Parallel-Constraint-Satisfaction Theory. *Psychol. Rev.* 103 (2), 284–308. doi:10.1037/0033-295x.103.2.284

Lerouge, D., and Warlop, L. (2006). Why it Is So Hard to Predict Our Partner's Product Preferences: the Effect of Target Familiarity on Prediction Accuracy. *J. Consum Res.* 33 (3), 393–402. doi:10.1086/508523

Levinson, W., Gorawara-Bhat, R., and Lamb, J. (2000). A Study of Patient Clues and Physician Responses in Primary Care and Surgical Settings. *JAMA* 284 (8), 1021–1027. doi:10.1001/jama.284.8.1021

Marangoni, C., Garcia, S., Ickes, W., and Teng, G. (1995). Empathic Accuracy in a Clinically Relevant Setting. *J. Personal. Soc. Psychol.* 68 (5), 854–869. doi:10.1037/0022-3514.68.5.854

Marvel, M. K., Epstein, R. M., Flowers, K., and Beckman, H. B. (1999). Soliciting the Patient's Agenda. *JAMA* 281 (3), 283–287. doi:10.1001/jama.281.3.283

Mills, G. J. (2014). Dialogue in joint activity: complementarity, convergence and conventionalization. *New Ideas Psychol.* 32, 158–173. doi:10.1016/j.newideapsych.2013.03.006

Mitchell, A. J. (2009). A Meta-Analysis of the Accuracy of the Mini-Mental State Examination in the Detection of Dementia and Mild Cognitive Impairment. *J. Psychiatr. Res.* 43 (4), 411–431. doi:10.1016/j.jpsychires.2008.04.014

Mitchell, J. P., Macrae, C. N., and Banaji, M. R. (2006). Dissociable Medial Prefrontal Contributions to Judgments of Similar and Dissimilar Others. *Neuron* 50 (4), 655–663. doi:10.1016/j.neuron.2006.03.040

Mussweiler, T. (2003). Comparison Processes in Social Judgment: Mechanisms and Consequences. *Psychol. Rev.* 110 (3), 472–489. doi:10.1037/0033-295X.110.3.472

Myers, M. W., and Hodges, S. D. (2009). "Making it up and Making Do: Simulation, Imagination, and Empathic Accuracy," in *Handbook of Imagination and Mental Simulation*. Editors K. Markman, W. Klein, and J. Suhr (New York, NY: Psychology Press).

Oskamp, S. (1965). Overconfidence in Case-Study Judgments. *J. Consulting Psychol.* 29 (3), 261–265. doi:10.1037/h0022125

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 (11), 1512–1527. doi:10.1177/0146167209342754

Porter, S., Woodworth, M., and Birt, A. R. (2000). Truth, Lies, and Videotape: An Investigation of the Ability of Federal Parole Officers to Detect Deception. *L. Hum. Behav.* 24 (6), 643–658. doi:10.1023/a:1005500219657

Pruitt, D. G., and Lewis, S. A. (1975). Development of Integrative Solutions in Bilateral Negotiation. *J. Personal. Soc. Psychol.* 31 (4), 621–633. doi:10.1037/0022-3514.31.4.621

Realo, A., Allik, J., Nõlvak, A., Valk, R., Ruus, T., Schmidt, M., et al. (2003). Mind-reading Ability: Beliefs and Performance. *J. Res. Personal.* 37 (5), 420–445. doi:10.1016/s0092-6566(03)00021-7

Reis, H. T., Lemay, E. P., Jr., and Finkenauer, C. (2017). Toward Understanding Understanding: The Importance of Feeling Understood in Relationships. *Social Personal. Psychol. Compass* 11 (3), e12308–22. doi:10.1111/spc3.12308

Ross, L., Greene, D., and House, P. (1977). The "false Consensus Effect": An Egocentric Bias in Social Perception and Attribution Processes. *J. Exp. Soc. Psychol.* 13 (3), 279–301. doi:10.1016/0022-1031(77)90049-X

Ryback, D. (1967). Confidence and Accuracy as a Function of Experience in Judgment-Making in the Absence of Systematic Feedback. *Percept Mot. Skills* 24 (1), 331–334. doi:10.2466/pms.1967.24.1.331

Santiesteban, I., White, S., Cook, J., Gilbert, S. J., Heyes, C., and Bird, G. (2012). Training Social Cognition: from Imitation to Theory of Mind. *Cognition* 122, 228–235. doi:10.1016/j.cognition.2011.11.004

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

Schei, V., Rognes, J. K., and Mykland, S. (2006). Thinking deeply may sometimes help: cognitive motivation and role effects in negotiation. *Appl. Psychol.* 55 (1), 73–90. doi:10.1111/j.1464-0597.2006.00240.x

Scheibehenne, B., Mata, J., and Todd, P. M. (2011). Older but Not Wiser-Predicting a Partner's Preferences Gets Worse with Age. *J. Consumer Psychol.* 21 (2), 184–191. doi:10.1016/j.jcps.2010.08.001

Simmons, J. P., and Nelson, L. D. (2006). Intuitive Confidence: Choosing between Intuitive and Nonintuitive Alternatives. *J. Exp. Psychol. Gen.* 135 (3), 409–428. doi:10.1037/0096-3445.135.3.409

Simpson, A. J., and Todd, A. R. (2017). Intergroup Visual Perspective-Taking: Shared Group Membership Impairs Self-Perspective Inhibition but May Facilitate Perspective Calculation. *Cognition* 166, 371–381. doi:10.1016/j.cognition.2017.06.003

Swann, W. B., Jr, and Gill, M. J. (1997). Confidence and Accuracy in Person Perception: Do We Know what We Think We Know about Our Relationship Partners?. *J. Personal. Soc. Psychol.* 73, 747–757. doi:10.1037/0022-3514.73.4.747

Thomas, G., Fletcher, G. J. O., and Lange, C. (1997). On-line Empathic Accuracy in Marital Interaction. *J. Personal. Soc. Psychol.* 72, 839–850. doi:10.1037/0022-3514.72.4.839

Thompson, L., and DeHarpport, T. (1994). Social Judgment, Feedback, and Interpersonal Learning in Negotiation. *Organizational Behav. Hum. Decis. Process.* 58 (3), 327–345. doi:10.1006/obhd.1994.1040

Thompson, L., and Hastie, R. (1990). Social perception in negotiation. *Organizat. Behav. Human Decis. Proc.* 47 (1), 98–123.

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., and Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *J. Person. Soc. Psychol.* 100 (6), 1027–1042. doi:10.1037/a0022308

Van Boven, L., Dunning, D., and Loewenstein, G. (2000). Egocentric Empathy Gaps between Owners and Buyers: Misperceptions of the Endowment Effect. *J. Personal. Soc. Psychol.* 79 (1), 66–76. doi:10.1037/0022-3514.79.1.66

Ward, M. K., and Broniarczyk, S. M. (2016). Ask and You Shall (Not) Receive: Close Friends Prioritize Relational Signaling over Recipient Preferences in Their Gift Choices. *J. Marketing Res.* 53 (6), 1001–1018. doi:10.1509/jmr.13.0537

Weingart, L. R., Bennett, R. J., and Brett, J. M. (1993). The Impact of Consideration of Issues and Motivational Orientation on Group Negotiation Process and Outcome. *J. Appl. Psychol.* 78 (3), 504–517. doi:10.1037/0021-9010.78.3.504

Weingartner, K. M., and Klin, C. M. (2005). Perspective Taking during reading: An On-Line Investigation of the Illusory Transparency of Intention. *Mem. Cogn.* 33 (1), 48–58. doi:10.3758/bf03195296

West, P. M. (1996). Predicting Preferences: an Examination of Agent Learning. *J. Consum Res.* 23 (1), 68–80. doi:10.1086/209467

White, J., Levinson, W., and Roter, D. (1994). Oh, by the Way. *J. Gen. Intern. Med.* 9 (1), 24–28. doi:10.1007/bf02599139

Zhou, H., Majka, E. A., and Epley, N. (2017). Inferring Perspective versus Getting Perspective: Underestimating the Value of Being in Another Person's Shoes. *Psychol. Sci.* 28 (4), 482–493. doi:10.1177/0956797616687124

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 June 2021
doi: 10.3389/fpsyg.2021.624604

Irony and Perspective-Taking in Children: The Roles of Norm Violations and Tone of Voice

Franziska Köder[1,2] *and Ingrid Lossius Falkum*[2,3]*

[1]*Department of Linguistics and Scandinavian Studies, Center for Multilingualism in Society Across the Lifespan, University of Oslo, Oslo, Norway,* [2]*Department of Philosophy, Classics, History of Arts and Ideas, University of Oslo, Oslo, Norway,* [3]*Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway*

In order to understand most, if not any communicative act, the listener needs to make inferences about what the speaker intends to convey. This perspective-taking process is especially challenging in the case of nonliteral uses of language such as verbal irony (e.g., "Thanks for your help!" uttered to someone who has not provided the expected support). Children have been shown to have difficulties with the comprehension of irony well into the school years, but the factors that hamper or facilitate children's perspective-taking in irony comprehension are not well understood. This study takes as its starting point the relevance-theoretic *echoic* analysis of verbal irony, and focuses on two of irony's distinctive features as defined by this theory: (i) the normative bias and (ii) the characteristic tone of voice. In this study, we investigated the comprehension of irony in children aged 3–8. We manipulated these two factors, namely, the violation of different types of norms and the use of different tones of voice – to see how they affected children's processing and interpretation of irony. Using an irony comprehension task that combined picture selection and eye-tracking, we found that the type of norm violation affected 4-to 5-year-olds' offline understanding of irony, with a better performance on moral compared with social norm violations. Tone of voice had an effect on gaze behavior in adults, but not children, although a parodic, pretense-oriented tone of voice tended to lead to more looks to the angry compared with the happy emoticon at the offset of the ironical utterance, potentially facilitating children's irony understanding. Our results show that the understanding of irony can be detected on explicit measures around age 6 – with the emergence of second-order perspective-taking abilities – but that a sensitivity to some of irony's features can be detected several years earlier. Finally, our study provides a novel input to the debate on the existence of a so-called literal stage in pragmatic development, in particular regarding 3-year-olds' differential performance on the offline and online measures of irony understanding, suggesting that they are not naively mistaking ironical utterances for "ordinary" literal ones.

Keywords: irony, development, pragmatics, relevance theory, tone of voice, norms

OPEN ACCESS

Edited by:
Dieter Hillert,
San Diego State University,
United States

Reviewed by:
Mikhail Kissine,
Université Libre de Bruxelles, Belgium
Paolo Rosso,
Universitat Politècnica de València,
Spain

***Correspondence:**
Ingrid Lossius Falkum
i.l.falkum@ifikk.uio.no

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

Received: *31 October 2020*
Accepted: *21 April 2021*
Published: *03 June 2021*

Citation:
Köder F and Falkum IL (2021) Irony and Perspective-Taking in Children: The Roles of Norm Violations and Tone of Voice. Front. Psychol. 12:624604. doi: 10.3389/fpsyg.2021.624604

INTRODUCTION

What Is Irony?

Although children's pragmatic abilities develop early, many studies have shown that until quite late in development, children have difficulties with so-called Gricean pragmatic inferences, which require them to go beyond the literal meaning of the linguistic form used to obtain the meaning intended by the speaker, as in implicature (e.g., Barner et al., 2011) and uses of figurative language (e.g., Winner, 1988/1997). This perspective-taking process is especially challenging in the case of verbal irony (1), where children have been shown to have difficulties well into the school years (Winner, 1988/1997; Creusere, 1999; Filippova and Astington, 2010; Glenwright and Pexman, 2010):

1. "Thanks for your help!" [*Uttered when someone has not provided the expected support*].

But the age of acquisition differs between studies, depending on the material and measures used to assess children's comprehension. While some studies have found only poor comprehension of irony in 8-year-old children (Massaro et al., 2013; Nicholson et al., 2013) or even 13-year-olds (Demorest et al., 1984), the lower age limit for irony understanding in experimental contexts appears to be around age 6 (Dews et al., 1996; Loukusa and Leinonen, 2008; Glenwright and Pexman, 2010). Some studies have even suggested that initial signs of irony comprehension may already be present in 3- and 4-year-old children (Loukusa and Leinonen, 2008; Reccia et al., 2010). However, the factors that hamper or facilitate children's perspective-taking in irony comprehension are not well understood (Falkum and Köder, 2020).

One proposed explanation for the relatively late development of irony comprehension is that unlike other types of nonliteral uses, such as metaphor, irony requires rather sophisticated mind-reading abilities, specifically the ability to understand second-order belief (Perner and Wimmer, 1985; Happé, 1993). Some studies have found developmental evidence of a positive correlation between the second-order theory of mind abilities and irony understanding (Happé, 1993; Massaro et al., 2013; but see Massaro et al., 2014), a finding that supports theoretical accounts of irony, which claim that it involves a speaker having a thought about a thought, more specifically dissociating herself from a thought attributed to someone else (Wilson and Sperber, 2012). However, links between theories of irony processing and developmental studies have been largely absent (Creusere, 1999; Wilson, 2013). Few developmental studies of irony have specifically tested claims made by theories of adults' understanding of verbal irony (but see, e.g., Keenan and Quigley, 1999). Moreover, a wide variety of operational definitions are used in the developmental literature, where irony is often understood in a very broad sense to include, for example, hyperbole, understatement, satire/parody, and rhetorical questions (see, e.g., Reccia et al., 2010). But is not clear that these phenomena all involve the same mechanism, and consequently, whether they should all be expected to follow the same developmental trajectory. Needless to say, how irony is conceptualized greatly influences hypotheses about the pragmatic and cognitive abilities required for its mastery, and about the underlying causes of children's apparent difficulties with it.

The aims of this study are as follows. We take as our starting point the relevance-theoretic *echoic* analysis of verbal irony (Wilson and Sperber, 2012), and focus on two of irony's distinctive features as defined by this theory, regarding (i) the normative bias involved in irony and (ii) the ironical tone of voice. Based on developmental data from children aged 3–8, we investigate how these two factors – the violation of different types of norms and the use of different tones of voice – affect children's processing and interpretation of irony. Finally, our study aims to address a general topic in the study of pragmatic development regarding whether children go through a so-called literal stage in figurative language acquisition.

Echoic Irony: Three Distinctive Features

Relevance theory (Sperber and Wilson, 1986/1995; Wilson and Sperber, 2004) is a theory of human communication that seeks to provide a psychologically realistic explanation of how hearers infer speaker meanings by combining contextual information with the linguistic evidence provided. The central claim is that utterances – including those that involve nonliteral uses of language – create expectations of relevance, which are precise and predictable enough to guide the addressee toward the speaker's meaning (see Wilson and Sperber, 2004; p. 608, for more detail). Relevance theory offers an alternative to the classical and Gricean views (Grice, 1975/1989) of irony as involving a deviation from a norm of literalness, which have been increasingly questioned (Gibbs, 1994; Glucksberg, 2001; Wilson and Sperber, 2012). Instead, the idea is that in verbal irony, the speaker is tacitly *echoing* a thought (a belief, an intention, a norm-based expectation) she attributes to another source (or to herself at a different time) and expressing a dismissive attitude to that thought. Thus, when Mary says "Thanks for your help" to someone who has not provided the expected support, she is expressing a dismissive attitude to the norm-based expectation that people should in general be helpful and supportive of each other. In this way, irony involves a "thought about a thought," that is, a second-order metarepresentation, and in order to grasp it, the hearer must be able (a) to understand the utterance as echoic and (b) to recognize the speaker's dismissive attitude to the attributed thought. Thus, the process of irony understanding places rather heavy demands on the child's perspective-taking abilities.

Wilson and Sperber (2012) discuss three distinctive features of irony that a theory of irony must explain: (a) the characteristic dismissive (e.g., mocking, skeptical, or contemptuous) attitude it expresses; (b) the normative bias it involves, usually pointing out that a state of affairs does not live up to some norm-based expectation; and (c) the ironical tone of voice, characterized by a flat or deadpan intonation. Further, they show how each of these features follows from their account of irony as an *echoic* use of language: first, irony is directly targeted at attributed thoughts and its purpose is to convey an attitude to those thoughts (a). Second, norms, in the sense of socially shared

ideas about how things should be, are always available to be ironically echoed when they are not satisfied (b). Third, the ironical tone of voice provides a cue to the particular type of dismissive attitude that the speaker intends to convey to the thought being echoed (c). When does the ability to understand irony, involving the three distinctive features in (a–c) above, emerge in development?

Echo or Pretense?

The main contemporary competitors to the echoic account of irony developed within relevance theory are so-called pretense theories (e.g., Clark and Gerrig, 1990; Kumon-Nakamura et al., 1995). Pretense accounts see the speaker of an ironical utterance as *pretending* to perform a speech act – in (2) below this would be an assertion – in order to express a dismissive attitude to the speech act itself, or anyone who would perform it or take it seriously.

2. A [*to a particularly clumsy person*]: you're so graceful!

A key issue in the literature is thus whether or not irony necessarily involves an element of pretense. Wilson and Sperber (2012) argue that echoing and pretense are distinct mechanisms; while all ironic utterances are echoic, only some of them involve pretense. Consider the ironical utterances in (3) below.

3. *John has just spilled a glass of wine on their new, white carpet.*

   a. Mary [*with a deadpan intonation*]: oh, that's great.
   b. Mary [with an exaggerated imitation of an enthusiastic tone of voice]: oh, that's GREAT!

While (3a) could be seen as a case of purely echoic irony, where the speaker is tacitly echoing an attributed thought and expressing a dismissive attitude to it, (3b) involves both echo and pretense and could be seen as an instance of "parodic" irony, where the speaker is imitating and thereby ridiculing the sort of person who would enthusiastically make such an exclamation. This example illustrates how the echoic and pretense accounts make different predictions about the ironical tone of voice (Sperber, 1984): on the pretense account, the speaker is expected to mimic the tone of voice of the person she is imitating. On the echoic account, the ironical speaker is not expected to leave her own voice behind, but to use instead a tone of voice designed to reflect her own dismissive attitude to the thought she is echoing.

Given children's ability to use pretense very early in development (i.e., their beginning to engage in pretending play roughly around the age of 18 months), we may hypothesize that the imitative, exaggerated tone of voice used in pretense-based forms of irony might make it easier for them to recognize that the speaker is distancing herself from the literal speech act she is performing, and thereby positively influence their understanding. If so, this would suggest that the distinct tones of voice used in the two forms of irony may be linked to different mechanisms, with "regular" irony involving echoing alone and "parodic" irony involving both echoing and pretense.

Although the role of pretense in irony is a matter of theoretical debate, irony clearly has an affinity with children's early-emerging ability for pretense. Further, irony is seen as closely related to humor and jokes (Gibbs et al., 2014), which also develop early (Hoicka, 2014). While the mechanisms underlying irony and jokes may be distinct (Wilson, 2017), there are some common aspects that children seem to relate to from an early age. Consider, for instance, the common parental practice of "reverse psychology." In a situation where a young child is unwilling to do something, the parent may tell the child, using a tone of voice that signals a joking attitude, *not* to perform the desired action (e.g., "Do not put on your shoes now…!"). And the child will typically disobey the parent's instruction with delight. In such cases, the child has recognized the parent's pretense, which signals a dissociation from the utterance's literal content. While these are not true cases of irony according to the standard echoic account, they share some of irony's features (characteristic tone of voice, dissociative attitude to a propositional content). Given this, we might expect some features of irony to be understandable to children younger than 6 years, which is considered to be the lower age limit for irony understanding in experimental settings.

Is There a Literal Stage in the Development of Nonliteral Uses of Language?

Children's early-emerging pragmatic competence is attested across a variety of studies and pragmatic tasks (see Tomasello, 2008, for an overview), including prelinguistic communication (Stephens and Matthews, 2014), word learning (Bloom, 2000), and referential communication (Matthews et al., 2006). In this light, children's apparent difficulties with nonliteral uses of language, that is, cases where they have to go beyond the conventional senses of the words and sentences (e.g., metonymy, metaphor, and irony), is increasingly seen as a puzzling feature of their pragmatic development. Some researchers have suggested that pre- and primary school children go through a *literal stage* in figurative language development (Asch and Nerlove, 1960; Winner et al., 1976; Winner, 1988/1997; Levorato and Cacciari, 2002), characterized by a decrease in their production of figurative language, for instance, early "metaphors" (Billow, 1981) and "metonyms" (Falkum, 2019), and a bias toward literal interpretations, before a more sophisticated level of figurative language competence is attained. For instance, Levorato and Caccari (2002; p. 129) have claimed that up to about age 7, a primitive type of processing is prevalent, which involves "a piece-by-piece elaboration of linguistic input; children process language literally even when it does not make sense in the linguistic context."

The idea of the existence of a literal stage has been criticized in connection with the growing evidence, gleaned from studies focusing on metaphor comprehension, that attests to the presence of a figurative language competence emerging as early as the preschool years (Deamer, 2013; Di Paola et al., 2020; Pouscoulous and Tomasello, 2020). However, in a recent study of metonymy comprehension using an offline picture selection task, Falkum et al. (2017) found a U-shaped

development, with 3-year-olds performing better than 4- and 5-year-olds, who tended to interpret metonymic uses literally (e.g., choosing a picture of a mustache instead of the man with the mustache for "*The mustache* sits down at the table"). In a replication of this study using a methodology that combined picture selection (offline) and eye-tracking (online), we observed the same U-shape in children aged 3–8 years, with performance starting to improve at age 6 (Köder and Falkum, 2020). However, results revealed a clear sensitivity to metonymic uses in the online eye-tracking data of *all* the participants, who preferred looking at the contextually appropriate metonymic referent, including those 4- to 5-year-old children who chose literal interpretations in the offline picture selection task. This suggests that there are properties of offline tasks that mask children's understanding – and which might lead them to choose literal interpretations of nonliteral communicative intentions – and that gaze data from eye-tracking could be more revealing of their actual pragmatic processing of the utterance.

Relating this to standard irony comprehension tasks, they often involve children being asked complex comprehension questions such as "Did X really mean that *p*?" or "Why did X say that *p*?" requiring quite developed verbal reasoning abilities. Recent research on children's irony comprehension, using more child-friendly tasks and implicit comprehension measures, suggests that children may not have a bias to access the literal interpretation first (Climie and Pexman, 2008; Whalen et al., 2020). Could the use of eye-tracking measures reveal that a sensitivity to certain types of ironical uses may emerge earlier than previously thought?

Developmental Hypotheses

In this paper, we link our developmental study to theoretical accounts of irony understanding and test children's processing and comprehension of irony, in the sense of the speaker tacitly dissociating herself from an *echoed* thought (cf. Wilson and Sperber, 2012). We focus on two of irony's distinctive features, as defined by Wilson and Sperber (2012): (i) the normative bias and (ii) the ironical tone of voice. Specifically, we investigate how different tones of voice and the violation of different types of norms affect children's processing and comprehension of irony. First, as mentioned above, irony is typically used when the speaker's norm-based expectations have been violated (Wilson and Sperber, 2012), such as in (1) that the addressee should be more helpful. A growing body of work attests to a strong disposition in children to infer, adhere to, and enforce norms (and conventions) across disparate domains, including language (Clark, 2007), pretense games (Rakoczy, 2007), and social behavior (Göckeritz et al., 2014). Children not only learn norms from direct instruction and prohibition, but also seek norms themselves: for instance, 3-year-olds have been shown to spontaneously infer a social norm from a single observation of adult intentional behavior (Schmidt et al., 2016). Further, the developmental literature shows that children are sensitive to different types of norms, distinguishing between moral and social norms from an early age (Nucci and Turiel, 1978; Nucci and Nucci, 1982). Children have been shown to perceive violations of moral norms as more serious and deserving of punishment than violations of social norms (Smetana, 1981). However, it is unclear whether the type of norm that is violated affects children's irony understanding, or whether children's attested early sensitivity to norms might help them in detecting the mismatch between the context and the speaker's utterance in irony. We hypothesize that irony is easier to understand when a moral norm compared with a social norm is violated (e.g., people violating an expectation to be helpful vs. violating an expectation to take off one's hat inside) as the discrepancy between expectation and reality, as well as the severity of the violation, is potentially more salient.

Second, irony is also typically associated with a characteristic tone of voice: a flat or deadpan intonation, slower tempo, lower pitch level, and greater intensity (Rockwell, 2000; Cheang and Pell, 2008) are important cues to a speaker's ironic intention. As we have seen, the two main theoretical accounts of irony, the echoic account (Wilson and Sperber, 2012) and the pretense account (Clark and Gerrig, 1990), make different predictions regarding the ironical tone of voice (Sperber, 1984): the pretense account expects the ironical speaker to mimic the tone of voice of the person she is imitating, leaving her own voice behind. The echoic account expects the speaker to use a tone of voice, which reflects her dismissive attitude to the echoed thought, and pretense elements are a possible additive, but not necessary ingredient. Given children's early familiarity with pretense, we hypothesize that a parodic, exaggerated tone of voice (pretense) might make an ironical utterance easier to understand than a deadpan tone of voice (echo). Theoretically, our goal was to test these predictions regarding the ironical tone of voice: if their developmental trajectories differ, that is, if one emerges before the other, this would suggest that the distinct tones of voice used in the two forms of irony may be linked to different mechanisms.

Finally, our study addresses the ongoing discussion regarding the existence of a "literal stage" in pragmatic development. While we expect that young children would choose mainly literal interpretations in the offline picture selection task, we hypothesize that, given irony's affinity with other pragmatic phenomena of which children show an early appreciation (pretense, jokes, humor), we would see some sensitivity to irony when including an online measure of eye-tracking, in line with our previous results for the comprehension of metonymy (Köder and Falkum, 2020). If so, it would suggest that young children, even if showing a literal preference on offline tasks, are not necessarily naively mistaking ironical utterances for "ordinary" literal ones.

MATERIALS AND METHODS

Participants

We tested 195 Norwegian-speaking children between 2 and 8 years and a control group of 20 adults. Since the accuracy data revealed that the task was too demanding for 2-year-old children, we excluded them from further analysis. In addition,

the data of one 7-year-old child was excluded due to experimenter error, leaving us with a total of 183 children (79 females) between 3 and 8 years (see **Table 1**). Written parental consent was obtained prior to the experiment. All children were tested individually in a separate room in their kindergarten or school on an SMI iView 250 RED mobile eye-tracker. They received a small reward (stickers) for participating.

Experimental Design

Participants listened to prerecorded stories, which were accompanied by pictures on a screen. This setup was chosen because of its high ecological validity, simulating children's common experience of picture-book reading. Before presenting the experimental items, we ensured that children knew that the two emoticons used in the experiment (see **Figure 1D**) represented a happy or angry emotion, respectively.

All experimental stories feature interactions between a parent and a child, a constellation that has previously been shown to facilitate irony understanding in children (Massaro et al., 2013). The stories are structured in the following way: first, the parent protagonist tells the child protagonist what she expects of him. For instance, in example story (1), mum is telling Emil to tidy up his messy room (see **Figure 1A**). With a comprehension question (see **Figure 1B**), we checked whether participants understood the parent's expectations and, in case they did not, repeated the story from the beginning. The story proceeds with the child protagonist either meeting or violating the parent protagonist's expectations, for instance, Emil tidying up his room or leaving it in a messy state (see **Figure 1C**). The parent reacts by either praising or criticizing the child, using one of three types of utterances: literal praise (e.g., "That's great! You have really tidied up"), literal criticism (e.g., "That's bad! You have not tidied up!"), or irony (e.g., "That's great! You have really tidied up!") uttered with either a deadpan or

TABLE 1 | Participants.

Age group	Mean age	Range	Number (f/m)
3	3.5	3.0–3.9	27 (14/13)
4	4.5	4.0–4.9	28 (9/19)
5	5.5	5.0–5.9	25 (7/18)
6	6.5	6.0–6.9	36 (17/19)
7	7.5	7.0–7.9	27 (11/16)
8	8.6	8.0–8.9	40 (21/19)
Adults	30.6	21–57	20 (10/10)

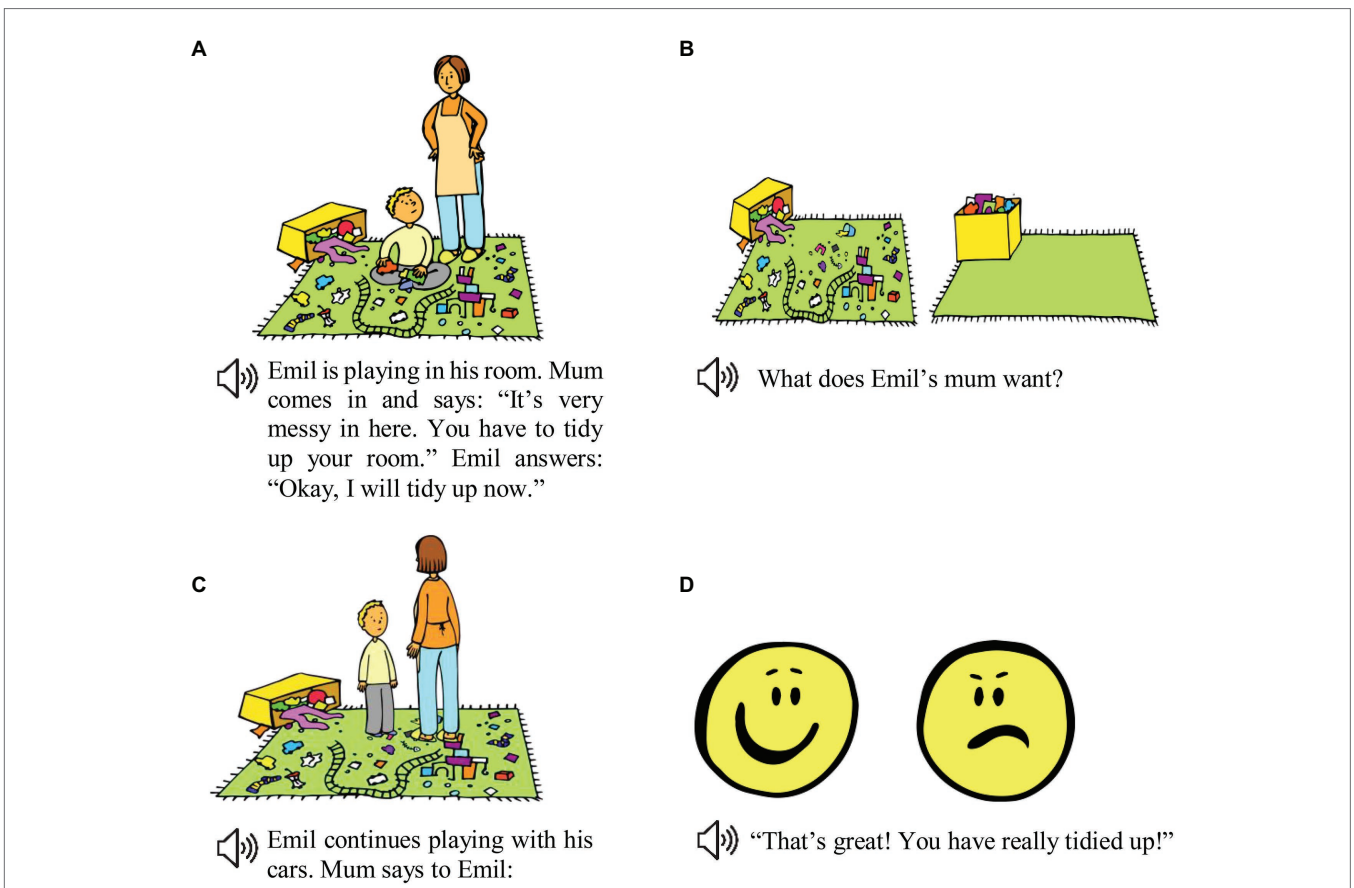


FIGURE 1 | Example of an experimental story, consisting of parts **(A-D)**. Stories are presented aurally, accompanied with pictures. Participants' gaze is measured during audio-visual presentation of **(D)**.

parodic tone of voice.[1] The literal or ironic target utterance starts 1,000 ms after the appearance of a happy and angry emoticon on the screen (see **Figure 1D**). Note that we decided to portray the story protagonists' emotions with emoticons rather than drawings of real faces to make clear that the protagonists' inner feelings are depicted, which are not necessarily reflected in facial expressions.

After each story, the experimenter asked the participant *How is mum/dad feeling inside? Is she/he happy* (experimenter pointing to the happy emoticon) *or angry* (experimenter pointing to the angry emoticon)? – which the participant could answer either verbally or by pointing to one of the emoticons. We counterbalanced the position of the emoticons (left, right) on the screen and in accordance also the order in which the two emotions are mentioned in the test question.

Each participant listened to 12 different stories, presented in random order. Six of them had a positive ending, that is, the child met the parent's expectations and received praise, and six had a negative ending, that is, the child violated the parent's expectations and was criticized either literally or ironically. All participants were randomly assigned to one of two tones of voice conditions (deadpan or parodic). Of the 12 stories, three involved moral norms (e.g., people should not hurt each other), three social norms (e.g., hands should be washed before eating), and three personal preferences (e.g., mum wants child to put on a nice dress).

Auditory Stimuli

The auditory stimuli for the experiment were recorded in a silent room with a high-quality recording device (H2 Zoom Handy Recorder). The recorded speaker, a male native speaker of Norwegian, received extensive training beforehand so that he could produce the two distinct types of ironic tone of voice naturally and consistently. For the deadpan stimuli, he was instructed to produce an ironical utterance with a flat, monotonous intonation. For the parodic stimuli, he was instructed to pretend to be happy while simultaneously expressing a dismissive attitude. From several recordings, we selected the ones that exhibited the specific auditory characteristics of a deadpan and a parodic tone of voice most clearly. The same speaker also produced utterances for the two literal control conditions. For literal praise, he was instructed to use a sincerely happy tone of voice and for the literal criticism an angry tone of voice.

Acoustic Analysis

All auditory stimuli were acoustically analyzed with the PRAAT software (version 6.1.15). As **Table 2** shows, the four types of utterances exhibit specific acoustic characteristics, with literal praise having, for instance, a higher fundamental frequency than literal criticism. In the statistical analysis, we focus on comparing deadpan and parodic irony, using *t*-tests with Bonferroni correction to correct for multiple comparisons. Comparing the two types of ironical tones of voice, we find that deadpan irony is uttered with lower intensity ($p < 0.001$), and its mean fundamental frequency is lower, even though not significantly different from that of parodic irony ($p = 0.32$).

TABLE 2 | Acoustical analyses for different types of utterances: mean duration, mean intensity, and mean F0.

	Deadpan irony	Parodic irony	Literal praise	Literal criticism
Duration (s)	2.2 (0.4)	2.4 (0.4)	2.2 (0.3)	2.3 (0.4)
Intensity (dB)	57.0 (2.0)	66.4 (2.2)	62.9 (1.7)	58.5 (4.5)
F0 (Hz)	132.1 (52.5)	158.6 (29.4)	201.3 (47.9)	128.9 (22.3)

Perceptual Evaluation

To determine whether the tone of voice of our auditory stimuli is a successful cue to the speaker's ironical intent, we conducted a perception study with 35 adult native speakers of Norwegian (mean age: 35, age range: 20–73). The participants listened to 36 recorded utterances of the type deadpan irony ($n = 12$), parodic irony ($n = 12$), and literal praise (n=12) that were presented in a random order without any contextual support. All utterance types have identical wording but differ in their acoustic properties. Participants were instructed to rate, based on the speaker's tone of voice, whether they agreed or disagreed that the speaker is being ironical/sarcastic (1 meaning strongly agree, 5 meaning strongly disagree). The results show that both deadpan (*mean* = 2.05, *SD* = 0.92) and parodic utterances (*mean* = 2.12, *SD* = 1.09) were perceived as ironical, while literal praise utterances were not (*mean* = 4.59, *SD* = 0.68). Pairwise *t*-tests with Bonferroni adjustments indicate that utterances of the types deadpan and parodic irony differed significantly from literal praise ($p < 0.001$), but not from each other ($p = 0.825$).

RESULTS

Picture Selection Results

Figure 2 shows the results from the picture selection task, in which participants were asked to choose the emoticon (happy/angry) that matched the feeling of the story protagonist best. Participants of all age groups performed well above chance in the literal praise and literal criticism condition. In the irony condition, 3-year-olds performed below chance, showing a bias toward literal interpretations of ironical utterances, while 4- and 5-year-old children performed at chance level. From the age of 6, children were mostly able to correctly assess the emotion of the ironical speaker, with 7- and 8-year-old children approaching adult-like performance.

We analyzed participants' accuracy of picture selection with generalized linear mixed-effects modeling (GLMM), using the "lme4 package" in the software R (version 3.6.1). Stepwise, random effects and fixed effects were added to a logistic regression model that improved the model fit significantly (decrease of AIC by more than 2; Akaike, 1974). The final

[1]Note that we did not include so-called ironic compliments, i.e., ironically praising someone (e.g., "Bad job!" uttered when someone succeeded), since irony is typically used to express a dismissive attitude (Wilson and Sperber, 2012).

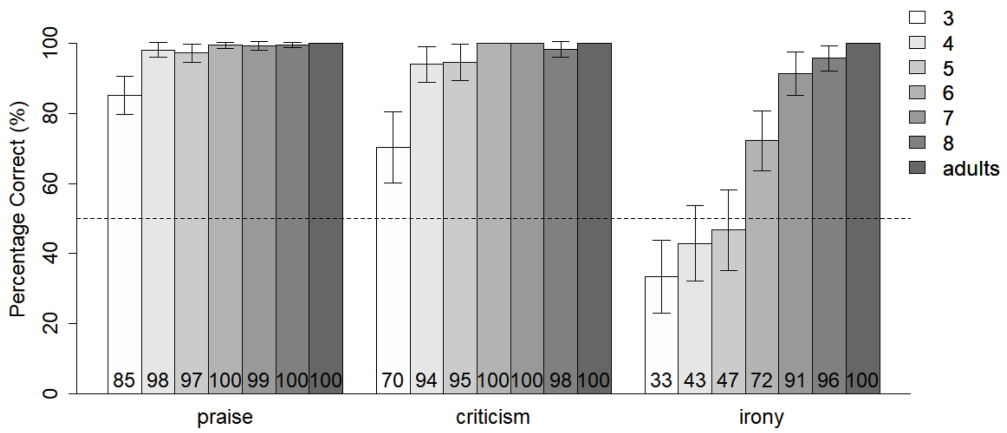


FIGURE 2 | Percentage of correct picture choice for different utterance types (literal praise, literal criticism, and irony) and age groups (3–8, adults). Error bars represent 95% confidence intervals. Dotted line indicates chance level.

model includes random intercepts for subjects and items and fixed effects for utterance type (praise, criticism, and irony) and age.[2] Based on the model, participants were significantly more often correct in their interpretation of literal praise (β = 4.13; z = 10.17; $p < 0.001$) or literal criticism (β = 3.15; z = 7.99; $p < 0.001$) compared with irony. Age has a positive effect on accuracy (β = 1.04; z = 10.13; $p < 0.001$). The type of norms (moral, social, and personal preference) addressed in the story, the tone of voice of the ironical utterances (deadpan and parodic), and the gender of participants (f/m) did not influence participants' picture choice. There was also no indication that participants' performance improved or deteriorated during the course of the experiment.

In the next step, we focused on the irony comprehension of 4- to 5-year-old children, who are in a transitional phase in irony acquisition, between a possible literal bias (age 3 and younger) and more robust irony understanding (age 6 and above). With generalized linear mixed effects models, we investigated whether the factors norm type and tone of voice influence irony comprehension in 4- and 5-year-old children. Our analysis reveals that tone of voice did not influence irony understanding, but that norm type had an effect: 4- to 5-year-olds tended to understand ironical utterances better when moral norms were violated as compared with personal preferences (β = 1.11; z = 2.13; p = 0.03) or social norms (β = 0.86; z = 1.68; p = 0.09), even though the latter was not statistically significant. When we combined the categories "social norms" and "personal preferences," we found that 4- to 5-year-olds performed better on irony understanding in stories where moral norms rather than nonmoral expectations (i.e., social norms and personal preferences) were violated (β = 0.98; z = 2.18; p = 0.03).[3]

Gaze Results

We analyzed participants' gaze to the two emoticons during the processing of the target utterances (irony, criticism, and praise), with looks to the angry emoticon compared with looks to the happy emoticon as a dependent variable. The analyzed time window spans from the onset of the target utterance, 1,000 ms after the image of the two emoticons appeared on the screen, until the average offset of the target utterance at 3,300 ms, adding 500 ms after the offset to take into account that young children need more time to initiate a saccade (Yang et al., 2002; Irving et al., 2006). We are interested in the looking behavior of three age groups, based on the accuracy analysis: younger children (3–5 years), who are still at chance level or below in their irony interpretation; older children (6–8 years), who are on average able to interpret ironical utterances correctly; and adults.

We analyzed the gaze data with generalized additive mixed modeling (GAMM) in R, using the R package "mgcv" (version 1.8–28; Wood, 2017). A GAMM analysis has the advantage that it allows for the modeling of nonlinear time-course effects, typical for gaze data in the visual world paradigm (Porretta et al., 2017). Stepwise, we included parametric and nonparametric factors, using the compareML function from the R package "itsadug" (van Rij et al., 2017) for model comparison. To investigate the interpretation of different types of utterances (irony, criticism, and praise) across different age groups (younger children, older children, and adults), we created the combined factor "Utterance type-Age group" with nine levels, such as "irony-adults" or "criticism-younger children." Our final model, presented in **Table 3**, includes the parametric predictor Utterance type-Age group, a time by-Utterance type-Age group smooth, and random intercepts for items.[4]

To visualize the model, we plotted the summed effects, using the "plot_smooth" function from the R package "itsadug"

[2]Note that when utterance type by participant random slopes were included, the model failed to converge.

[3]Model: Norm_type_combined + (1|Participant), Note that the model failed to converge when Norm_type_combined by participant random slopes were included.

[4]Note that random smooths for participants and items did not improve the model fit significantly.

TABLE 3 | Generalized additive mixed model reporting parametric coefficients (Part A), and smooth terms and random effects (Part B).

Part A				
Parametric coefficients	**Estimate**	**Std. Error**	***z* value**	***p* value**
(Intercept)	0.884	0.063	13.835	<0.001
Criticism-Adults	−0.096	0.107	−0.902	0.367
Criticism-Older	−0.533	0.104	5.121	<0.001
Criticism-Younger	−0.621	0.104	−5.964	<0.001
Irony-Older	−0.690	0.033	−20.959	<0.001
Irony-Younger	−0.768	0.032	−23.685	<0.001
Praise-Adults	−2.069	0.105	−19.660	<0.001
Praise-Older	−1.281	0.104	−12.379	<0.001
Praise-Younger	−1.102	0.104	−10.636	<0.001
Part B				
Smooth terms	**edf**	**Ref.df**	**Chi.sq**	***p* value**
s(Time): Irony-Adults	6.357	7.509	49.50	<0.001
s(Time): Criticism-Adults	7.902	8.685	54.24	<0.001
s(Time): Criticism-Older	8.169	8.812	197.14	<0.001
s(Time): Criticism-Younger	8.128	8.792	187.93	<0.001
s(Time): Irony-Older	7.284	8.284	63.63	<0.001
s(Time): Irony-Younger	8.014	8.737	87.23	<0.001
s(Time): Praise-Adults	7.120	8.163	393.43	<0.001
s(Time): Praise-Older	7.091	8.143	277.28	<0.001
s(Time): Praise-Younger	8.073	8.768	217.57	<0.001
Random intercepts for items	44.063	45.00	2120.23	<0.001

(van Rij et al., 2017; see **Figure 3**). If the graph is above 0.5, this means that participants looked more to the angry compared with the happy emoticon; if it is below 0.5, participants looked more to the happy compared with the angry emoticon; if it is equal to 0.5, participants looked equally often to the happy and angry emoticon.

The gaze data of the adults (**Figure 3C**) reveal clearly distinct looking patterns for literal praise on the one hand – with more looks to the happy emoticon (graph below 0.5) – and irony and literal criticism on the other hand – with more looks to the angry emoticon (graphs above 0.5). Similar to adults, older children (**Figure 3B**) showed distinct looking patterns for ironical utterances compared with literal praise utterances throughout the whole time frame. However, for 6-to 8-year-olds also, the looking patterns for ironical utterances and literal criticism differed significantly: from 3,375 ms onward, they looked less at the angry emoticon when listening to ironical utterances, with an equal amount of looks to both emoticons in the irony condition from around 3,300 ms. Younger children (**Figure 3A**) also showed a clear tendency to look more at the angry emoticon when listening to ironical utterances compared with literal praise utterances. However, from 3,347 ms, the graphs for irony and literal praise do not differ significantly anymore. Around the same time (3,375 ms), younger children's looking patterns in the irony and literal criticism condition start to differ from each other.[5] **Figure 3** also reveals an age-related trend: the older the participants, the bigger the difference in "angry vs. happy looks" between literal praise, on the one hand, and the two forms of criticism (irony and literal criticism), on the other hand, visible as the spatial distance between the graphs.

Next, we analyzed whether tone of voice and norm type influence participants' looks in the irony condition, using generalized linear mixed models. For the complete time frame (1,000–3,800 ms), we did not find an effect of tone of voice or norm type for neither adults nor children. However, when focusing on the time window after the offset of the target utterance (3,300–3,800 ms), which in the GAMM analysis showed the most distinct looking patterns for the different age groups, we can notice interesting effects. Adults looked significantly more to the angry compared with the happy emoticon when irony was produced with a deadpan compared with a parodic tone of voice (β = 12.60; z = 1.96; p = 0.05).[6] For children, we found a tendency toward the reverse effect, with the parodic tone of voice leading to more looks to the angry compared with the happy emoticon (β = 1.77; z = 2.05; p = 0.04).[7] However, when we included random intercepts for items in the model, this effect was not significant any more (β = 1.83; z = 1.25; p = 0.21). The type of norm did not affect eye gaze in children and adults.

DISCUSSION

We investigated children's irony processing and understanding by combining the offline measure of picture selection with the online measure of eye-tracking. The data from the picture selection task show a clear improvement of irony understanding with age: 3-year-olds were below chance in the interpretation of irony, while 4- and 5-year-old children were at chance level. From the age of 6, children have already a good understanding of verbal irony, which improves further in the years after. The fact that even the youngest children we tested, the 3-year-olds, performed well in the two literal control conditions confirms that the task was age-appropriate. The gaze data provide novel information about the processing of ironical utterances compared with utterances with the same wording (literal praise), and utterances with the same context and similar negative valence (literal criticism). In adults, we can see a two-way distinction between the processing of utterances with positive and negative valence, with no significant difference in looking behavior between irony and literal criticism. Older children, aged 6–8, exhibit a similar distinction between the processing of literal praise on the one hand and irony and literal criticism on the other hand. However, toward the end of an ironical utterance, these older children looked equally often to the happy as to the angry emoticon. This could be an indication that the literal meaning is more prominent during children's than during adults' processing of irony. The below-chance performance of 3-year-olds' in the picture selection task suggests that they have a strong bias to interpret ironical utterances literally.

[5]The values on when differences between conditions are significant were calculated using the "plot_diff" function from the R package "itsadug" (van Rij et al., 2017).

[6]Model: Utterance_type + (1 | Participant) + (1 | Stimulus), data: adults, irony data, time window: 3,300–3,800 ms.

[7]Model: Utterance_type + (1 | Participant), data: children aged 3–8, irony data, time window: 3,300–3,800 ms.

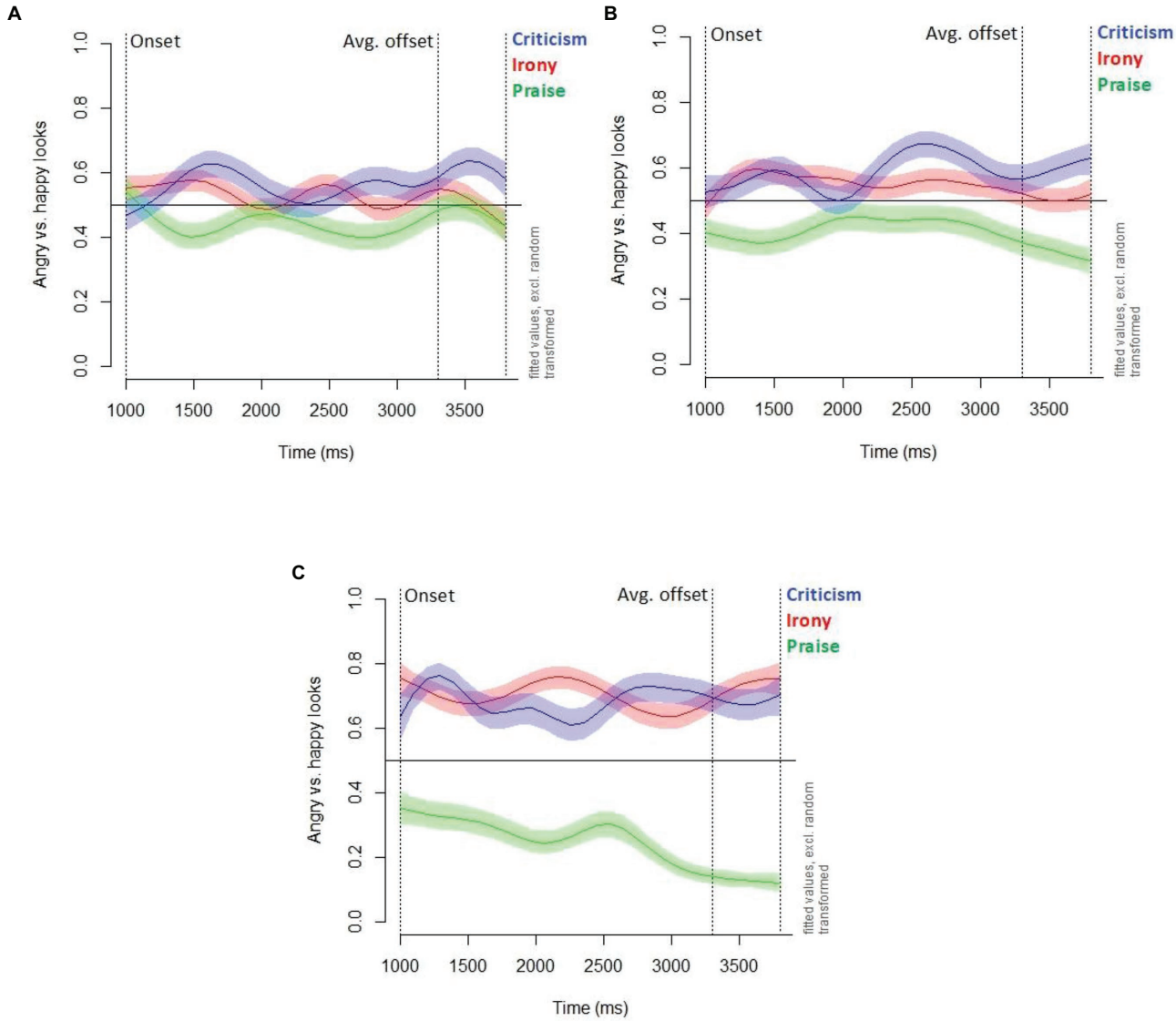


FIGURE 3 | Estimated difference in proportion of looks to angry vs. happy emoticon for different utterance types (criticism, irony, praise) in **(A)** Younger children aged 3-5 years, **(B)** Older children aged 6-8 years, and **(C)** Adults. Estimations are based on the GAMM presented in **Table 3**.

However, the gaze data show that ironical utterances led to significantly more looks to the angry compared with the happy emoticon during most of the utterance, suggesting that 3- to 5-year-olds are sensitive to cues such as context or tone of voice, which are indications that the speaker is likely to have a negative attitude. However, once 3- to 5-year-olds have analyzed the complete utterance, the gaze pattern for ironical utterances resembles that of literal praise utterances.

Norm Type

Although the developmental literature shows that children distinguish between moral and social norms from an early age, we found only partial support for the hypothesis that the type of norm violated influences children's irony understanding. Interestingly, children aged 4–5, who are on the verge of understanding irony, tended to perform better in the picture selection task when a moral norm such as "do not hurt others" was violated than when a social norm (e.g., "wash your hands before dinner") or personal preference (e.g., mother's clothing preference) was violated. If children perceive moral norm violations as more severe and punishable than violations of social norm (Smetana, 1981; or violations of personal preferences), this might have made the discrepancy between expectation and reality more salient and thereby reduced the likelihood that the ironical utterance was meant literally. This, in turn, might have made it easier for children to notice the dismissive attitude of the ironical speaker and discard the literal interpretation. However, for children aged 6 and older, who in general performed well on irony comprehension, we did not find an effect of norm type violated. This might be because once the perspective-taking

abilities necessary for irony comprehension are mastered, it does not matter so much what sort of norm violation underlies the ironical utterance. If anything, it might be slightly less appropriate to use irony in a moral norm violation setting, because the perceived severity of the violation may not lend itself naturally to the mockery that is typically involved in ironical uses. However, we did not find evidence of this in our data.

Tone of Voice

We also investigated whether the tone of voice used with the ironical utterance influences children's understanding, comparing a deadpan (flat, slower tempo, lower pitch level, and greater intensity) and a parodic (imitative and exaggerated) intonation. As mentioned above, the two competing accounts of irony, the *echoic* account (Wilson and Sperber, 2012) and the *pretense* account (Clark and Gerrig, 1990), make different predictions about the ironical tone of voice (Sperber, 1984): while on the pretense account, the speaker is expected to mimic the tone of voice of the person she is imitating, on the echoic account, the ironical speaker is not expected to leave her own voice behind, but to use instead a tone of voice designed to reflect her own dismissive attitude to the thought she is echoing. Our hypothesis was that the imitative, exaggerated tone of voice used in pretense-based forms of irony might make it easier for children, who are already familiar with pretend situations, to recognize that the speaker is distancing herself from the literal speech act she is performing, and thereby positively influence their understanding.

While we did find support for an effect of tone of voice in our data, this was subtler than predicted. In the gaze data, adults looked more at the angry emoticon during the final phase of the utterance when a deadpan tone of voice was used. This could indicate that for adults a deadpan tone of voice is the default tone of voice for verbal irony and therefore more strongly linked to a negative attitude, a finding that would support the echoic account of irony. By contrast, in the same time window, there was an indication for the opposite effect in children. They tended to look more at the angry emoticon when irony was uttered with a parodic tone of voice, suggesting that the imitative, exaggerated tone of voice used in pretense-based irony made it easier for them to recognize the speaker's nonliteral intention. Such a tone of voice is already familiar to children from other types of pretense contexts, including humor, jokes, and "reverse psychology," which share some of irony's attributes (e.g., characteristic tone of voice, dissociative attitude to a propositional content). We take our results to provide some support for the claim that the distinct tones of voice used in the two forms of irony may be linked to different mechanisms, with "regular" irony involving echoing alone and "parodic" irony involving both echoing and pretense. However, the lack of clear processing differences between ironic utterances produced with a deadpan compared with a parodic tone of voice could also indicate that tone of voice is potentially a less reliable ironic cue for children than cues such as context and facial expressions (cf. Bryant and Fox Tree, 2005; Deliens et al., 2018). Another possibility, as one reviewer pointed out, might be that from an acoustic point of view the stimuli with deadpan intonation were not sufficiently different from those with a parodic intonation to allow the children to discriminate between them (although adults clearly did).

All in all, the effects of both tone of voice and norm type were only evident for certain age groups, certain time windows or certain types of data. This is why these findings need to be interpreted with caution and require replication in future studies.

A "Literal" Bias in Interpretation?

One clear advantage of including an online measure of eye-tracking in our study of irony comprehension was that it allowed us to explore what is going on when children misinterpret irony. Do they interpret ironical utterances as if they were literal, positive utterances? For instance, in (1) above, would they actually think that the speaker is praising the addressee for helping her, even if the context should clearly suggest otherwise? In the offline picture selection task, we expected young children to choose mainly literal interpretations, a prediction that was borne out by the data. In fact, the below-chance performance of 3-year-olds in the picture selection task suggests that they have a strong bias to interpret ironical utterances literally, similar to how they interpret cases of literal praise. The gaze data, however, provide a more refined picture. As we have seen, ironical utterances led to significantly more looks to the angry compared with the happy emoticon during most of the utterance. This shows that these young children are not naively mistaking irony for sincere praise, but that they are sensitive to cues such as context and/or tone of voice, which are indications that the speaker is likely to have a negative attitude. However, once 3- to 5-year-olds have analyzed the complete utterance, the gaze pattern for ironical utterances resembles that of literal praise utterances. This suggests that at this age, the semantic content of the ironical utterance is prioritized over other types of information such as context and tone of voice.

Our results have implications for the discussion of a possible "literal stage" in pragmatic development. First, they do not mesh well with the claim that children process utterances literally, piece-by-piece, even if it does not make sense in the context (cf. Levorato and Cacciari, 2002). If this were true, we should expect children to perform consistently below chance on the picture selection measure in our irony comprehension task. While this was the case for the 3-year-olds in our study, their eye-tracking behavior clearly suggests that they take context and/or tone of voice into account when processing ironical utterances, although they ultimately end up with a "wrong," literal interpretation. Second, our results corroborate evidence from other tasks combining picture selection and eye-tracking to test figurative language comprehension (e.g., metonymy; Köder and Falkum, 2020). These show that the literal preference that young children tend to show on offline measures such as picture selection is discontinuous with their more "pragmatically appropriate" gaze behavior on eye-tracking measures, and that the latter might be more revealing not only of their actual pragmatic processing of the utterance but also of their pragmatic competence more generally.

One possible explanation for the 3- to 5-year-olds' poor performance on the picture selection task could be the rather

heavy demands that irony comprehension places on their still developing perspective-taking abilities, and that children in this age group are simply unable to grasp that a speaker is tacitly dissociating herself from a thought she attributes to someone else. While this is likely to be one side of the explanation, we take our results to suggest that the situation is more nuanced than this, with young children being sensitive to some of irony's features well before they show any understanding on offline measures. Also, as we have discussed above, irony has an affinity with other pragmatic phenomena of which children show an early appreciation, such as pretense, jokes, and humor, some of which might also – in certain cases, at least – require some rather complex perspective-taking. More research is needed to pin down the contribution of several different factors – including children's perspective-taking abilities – in the development of irony comprehension.

CONCLUSION

In this study, we investigated the comprehension of irony in children aged 3–8 years. We took as our starting point the relevance-theoretic account of verbal irony, focusing on two of irony's distinctive features as defined by this theory: (i) the normative bias and (ii) the characteristic tone of voice. We manipulated these two factors, namely, the violation of different types of norms and the use of different tones of voice, to see how they affected children's processing and interpretation of irony. While the type of norm violation affected 4- to 5-year-olds' offline understanding of irony, with a better performance on moral norm violations, tone of voice did not have a significant effect on children's online gaze behavior, although the parodic tone of voice tended to lead to more looks to the angry compared with the happy emoticon at the offset of the ironical utterance.

All our experimental items were cases of "echoic" irony in the sense of Wilson and Sperber (2012), where the speaker is tacitly dissociating herself from an attributed thought (belief, norm-based expectation). Our results show that the understanding of irony in this sense can be detected on explicit measures around age 6 – around the same age as second-order perspective-taking abilities emerge – but that a sensitivity to several of irony's features can be seen in both offline and online measures several years earlier. With links between developmental studies and theoretical accounts of irony understanding (in adults) hitherto being largely absent, our study can be seen as one step toward connecting these two strands of research and thereby contributing to building a more coherent account of the development of irony understanding.

Finally, our study provides novel input to the debate on the existence of a so-called literal stage in pragmatic development, in particular regarding 3-year-olds' differential performance on the offline and online measures of irony understanding, suggesting that they are not naively mistaking ironical utterances for "ordinary" literal ones.

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 NSD - Norwegian Centre for Research Data. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

FK and IF designed and conducted the study in collaboration. FK wrote the methodology and results sections, and IF wrote the introduction and discussion sections. Both the authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Research Council of Norway (project numbers 240324 and 302083).

ACKNOWLEDGMENTS

We would like to thank all the children and adults who participated in our study, as well as the parents and staff at the kindergartens and schools for their collaboration; this research would not have been possible without them. We are grateful to Deirdre Wilson, who has provided us with inspiration for this study, and support and encouragement throughout the process. We thank Tanja Russita for the illustrations of our experimental stories, Bjørn Torgrim Ramberg for help with the audio-recordings, Ane Theimann for help with the data collection, the audience of various conferences and workshops, Francisco Pons, and the two reviewers Mikhail Kissine and Paolo Rosso for helpful comments and suggestions.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Asch, S., and Nerlove, H. (1960). "The development of double function terms in children: an exploratory investigation," in *Perspectives in Psychological Theory: Essays in Honor of Heinz Werner*. eds. B. Kaplan and S. Wapner (New York: International Universities Press), 47–60.

Barner, D., Brooks, N., and Bale, A. (2011). Assessing the unsaid: the role of scalar alternatives in children's pragmatic inference. *Cognition* 118, 84–93. doi: 10.1016/j.cognition.2010.10.010

Billow, R. M. (1981). Observing spontaneous metaphor in children. *J. Exp. Child Psychol.* 31, 430–455. doi: 10.1016/0022-0965(81)90028-X

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge, MA: The MIT Press.

Bryant, G. A., and Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Lang. Speech* 48, 257–277. doi: 10.1177/00238309050480030101

Cheang, H. S., and Pell, M. D. (2008). The sound of sarcasm. *Speech Comm.* 50, 366–381. doi: 10.1016/j.specom.2007.11.003

Clark, E. V. (2007). Conventionality and contrast in language and language acquisition. *New Dir. Child Adolesc. Dev.* 115, 11–23. doi: 10.1002/cad.179

Clark, H., and Gerrig, R. (1990). On the pretense theory of irony. *J. Exp. Psychol. Gen.* 113, 121–126.

Climie, E. A., and Pexman, P. M. (2008). Eye gaze provides a window on children's understanding of verbal irony. *J. Cogn. Dev.* 9, 257–285. doi: 10.1080/15248370802247939

Creusere, M. A. (1999). Theories of adults' understanding and use of irony and sarcasm: applications and evidence from research with children. *Dev. Rev.* 19, 213–262. doi: 10.1006/drev.1998.0474

Deamer, F. (2013). An Investigation into the Processes and Mechanisms Underlying the Comprehension of Metaphor and Hyperbole. PhD dissertation: University College London.

Deliens, G., Antoniou, K., Clin, E., Ostashchenko, E., and Kissine, M. (2018). Context, facial expression and prosody in irony processing. *J. Mem. Lang.* 99, 35–48. doi: 10.1016/j.jml.2017.10.001

Demorest, A., Meyer, C., Phelps, E., Gardner, H., and Winner, E. (1984). Words speak louder than actions: understanding deliberately false remarks. *Child Dev.* 55, 1527–1534. doi: 10.2307/1130022

Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., et al. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Dev.* 67, 3071–3085. doi: 10.2307/1131767

Di Paola, S., Domaneschi, F., and Pouscoulous, N. (2020). Metaphorical developing minds: the role of multiple factors in the development of metaphor comprehension. *J. Pragmat.* 156, 235–251. doi: 10.1016/j.pragma.2019.08.008

Falkum, I. L. (2019). "Metaphor and metonymy in acquisition: a relevance-theoretic perspective," in *Relevance, Pragmatics and Interpretation*. eds. K. Scott, B. Clark and R. Carston (Cambridge: Cambridge University Press), 205–217.

Falkum, I. L., and Köder, F. (2020). The acquisition of figurative meanings. *J. Pragmat.* 164, 18–24. doi: 10.1016/j.pragma.2020.04.007

Falkum, I. L., Recasens, M., and Clark, E. V. (2017). "The moustache sits down first": on the acquisition of metonymy. *J. Child Lang.* 44, 87–119. doi: 10.1017/S0305000915000720

Filippova, E., and Astington, J. W. (2010). Children's understanding of social-cognitive and social-communicative aspects of discourse irony. *Child Dev.* 81, 913–928. doi: 10.1111/j.1467-8624.2010.01442.x

Gibbs, R. W. (1994). *The Poetics of Mind*. Cambridge: Cambridge University Press.

Gibbs, R. W., Bryant, G. A., and Colston, H. L. (2014). Where is the humor in verbal irony? *Humor* 27, 575–595. doi: 10.1515/humor-2014-0106

Glenwright, M., and Pexman, P. (2010). Development of children's ability to distinguish sarcasm and verbal irony. *J. Child Lang.* 37, 429–451. doi: 10.1017/S0305000909009520

Glucksberg, S. (2001). *Understanding Figurative Language: From Metaphors to Idioms*. New York: Oxford University Press.

Göckeritz, S., Schmidt, M. F. H., and Tomasello, M. (2014). Young children's creation and transmission of social norms. *Cogn. Dev.* 30, 81–95. doi: 10.1016/j.cogdev.2014.01.003

Grice, H. P. (1975/1989). "Logic and conversation," in *Studies in the Way of Words*. Cambridge, Massachusetts: Harvard University Press, 22–40.

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition* 48, 101–119. doi: 10.1016/0010-0277(93)90026-R

Hoicka, E. (2014). "The pragmatic development of humour," in *Pragmatic Development in First Language Acquisition*. ed. D. Matthews (Amsterdam: John Benjamins), 219–238.

Irving, E. L., Steinbach, M. J., Lillakas, L., Babu, R. J., and Hutchings, N. (2006). Horizontal saccade dynamics across the human life span. *Investig. Ophthalmol. Vis. Sci.* 47, 2478–2484. doi: 10.1167/iovs.05-1311

Keenan, T., and Quigley, K. (1999). A test of echoic mention theory. *Br. J. Dev. Psychol.* 17, 83–96. doi: 10.1348/026151099165168

Köder, F., and Falkum, I. L. (2020). Children's metonymy comprehension: evidence from eye-tracking and picture selection. *J. Pragmat.* 156, 191–205. doi: 10.1016/j.pragma.2019.07.007

Kumon-Nakamura, S., Glucksberg, S., and Brown, C. A. (1995). How about another piece of pie: the allusional pretence theory. *J. Exp. Psychol. Gen.* 124, 3–21. doi: 10.1037/0096-3445.124.1.3

Levorato, M. C., and Cacciari, C. (2002). The creation of new figurative expressions: psycholinguistic evidence in Italian children, adolescents and adults. *J. Child Lang.* 29, 127–150. doi: 10.1017/S0305000901004950

Loukusa, S., and Leinonen, R. (2008). Development of comprehension of ironic utterances in 3- to 9-year-old Finnish-speaking children. *Psychol. Lang. Commun.* 12, 55–69. doi: 10.2478/v10057-008-0003-0

Massaro, D., Valle, A., and Marchetti, A. (2013). Irony and second-order false belief in children: what changes when mothers rather than siblings speak? *Eur. J. Dev. Psychol.* 10, 301–317. doi: 10.1080/17405629.2012.672272

Massaro, D., Valle, A., and Marchetti, A. (2014). Do social norms, false belief understanding, and metacognitive vocabulary influence irony comprehension? A study of five- and seven-year-old children. *Eur. J. Dev. Psychol.* 11, 292–304. doi: 10.1080/17405629.2013.821407

Matthews, D., Lieven, E., Theakston, A., and Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Appl. Psycholinguist.* 27, 403–422. doi: 10.1017/S0142716406060334

Nicholson, A., Whalen, J. M., and Pexman, P. M. (2013). Children's processing of emotion in ironic language. *Front. Psychol.* 4:691. doi: 10.3389/fpsyg.2013.00691

Nucci, L., and Nucci, M. (1982). Children's responses to moral and social conventional transgressions in free-play settings. *Child Dev.* 53, 1337–1342. doi: 10.2307/1129024

Nucci, L., and Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Dev.* 49, 400–407. doi: 10.2307/1128704

Perner, J., and Wimmer, H. (1985). "John thinks that Mary thinks that …" attribution of second-order beliefs by 5- to 10-year-old children. *J. Exp. Child Psychol.* 39, 437–471. doi: 10.1016/0022-0965(85)90051-7

Porretta, V., Kyröläinen, A.-J., van Rij, J., and Järvikivi, J. (2017). "Visual world paradigm data: from preprocessing to nonlinear time-course analysis," in *Intelligent Decision Technologies 2017 Proceedings of the 9th KES International Conference on Intelligent Decision Technologies (KES-IDT 2017) – Part II*; June 21–23, 2017. *Vol. 15.* eds. I. Czarnowski, R. J. Howlett and L. C. Jain, 268–277.

Pouscoulous, N., and Tomasello, M. (2020). Early birds: metaphor understanding in 3-year-olds. *J. Pragmat.* 156, 160–167. doi: 10.1016/j.pragma.2019.05.021

Rakoczy, H. (2007). Play, games, and the development of collective intentionality. *New Dir. Child Adolesc. Dev.* 115, 53–67. doi: 10.1002/cd.182

Reccia, H., Howe, N., Ross, H., and Alexander, S. (2010). Children's understanding and production of verbal irony in family conversations. *Br. J. Dev. Psychol.* 28, 255–274. doi: 10.1348/026151008X401903

Rockwell, P. (2000). Lower, slower, louder: vocal cues of sarcasm. *J. Psycholinguist. Res.* 29, 483–495. doi: 10.1023/A:1005120109296

Schmidt, M. F. H., Butler, L. P., Heintz, J., and Tomasello, M. (2016). Young children see a single action and infer a social norm: promiscuous normativity in 3-year-olds. *Psychol. Sci.* 27, 1360–1370. doi: 10.1177/0956797616661182

Smetana, J. (1981). Preschool children's conceptions of moral and social rules. *Child Dev.* 52, 1333–1336. doi: 10.2307/1129527

Sperber, D. (1984). Verbal irony: pretence or echoic mention. *J. Exp. Psychol. Gen.* 113, 130–136. doi: 10.1037/0096-3445.113.1.130

Sperber, D., and Wilson, D. (1986/1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stephens, G., and Matthews, D. (2014). "The communicative infant from 0-18 months. The social-cognitive foundations of pragmatic development," in *Pragmatic Development in First Language Acquisition*. ed. D. Matthews (Amsterdam: John Benjamins), 13–35.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: The MIT Press.

van Rij, J., Wieling, M., Baayen, R., and van Rijn, H. (2017). Iitsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package version 2.3.

Whalen, J. M., Doyle, A., and Pexman, P. (2020). Sarcasm between siblings: children's use of relationship information in processing ironic remarks. *J. Pragmat.* 156, 149–159. doi: 10.1016/j.pragma.2019.05.005

Wilson, D. (2013). Irony comprehension: a developmental perspective. *J. Pragmat.* 59, 40–56. doi: 10.1016/j.pragma.2012.09.016

Wilson, D. (2017). "Irony, hyperbole, jokes and banter," in *Formal Models in the Study of Language: Applications in Interdisciplinary Contexts*. eds. J. Blochowiak, C. Grisot, S. Durriemann and C. Laenzlinger (Cham: Springer), 201–220.

Wilson, D., and Sperber, D. (2004). "Relevance theory," in *The Handbook of Pragmatics*. eds. L. R. Horn and G. L. Ward (Oxford: Blackwell), 607–632.

Wilson, D., and Sperber, D. (2012). "Explaining irony," in *Meaning and Relevance*. Cambridge: Cambridge University Press, 123–145.

Winner, E. (1988/1997). *The Point of Words: Children's Understanding of Metaphor and Irony*. Cambridge, MA: Harvard University Press.

Winner, E., Rosenstiel, A. K., and Gardner, H. (1976). The development of metaphoric understanding. *Dev. Psychol.* 12, 289–297. doi: 10.1037/0012-1649.12.4.289

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R. 2nd Edn*. Boca Raton: Chapman and Hall/CRC.

Yang, Q., Bucci, M. P., and Kapuola, S. (2002). The latency of saccades, vergence, and combined eye movements in children and in adults. *Investig. Ophthalmol. Vis. Sci.* 43, 2939–2949.

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 June 2021
doi: 10.3389/fpsyg.2021.610401

Children's Pronoun Interpretation Problems Are Related to Theory of Mind and Inhibition, But Not Working Memory

*Sanne J. M. Kuijper[1], Catharina A. Hartman[2] and Petra Hendriks[3]**

[1] Inclusive and Special Needs Education Unit, Department of Pedagogical and Educational Sciences, University of Groningen, Groningen, Netherlands, [2] Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), University of Groningen, University Medical Center Groningen, Groningen, Netherlands, [3] Center for Language and Cognition Groningen, University of Groningen, Groningen, Netherlands

In several languages, including English and Dutch, children's acquisition of the interpretation of object pronouns (e.g., *him*) is delayed compared to that of reflexives (e.g., *himself*). Various syntactic and pragmatic explanations have been proposed to account for this delay in children's acquisition of pronoun interpretation. This study aims to provide more insight into this delay by investigating potential cognitive mechanisms underlying this delay. Dutch-speaking children between 6 and 12 years old with autism spectrum disorder (ASD; n = 47), attention-deficit/hyperactivity disorder (ADHD; n = 36) or typical development (TD; n = 38) were tested on their interpretation and production of object pronouns and reflexives and on theory of mind, working memory, and response inhibition. It was found that all three groups of children had difficulty with pronoun interpretation and that their performance on pronoun interpretation was associated with theory of mind and inhibition. These findings support an explanation of object pronoun interpretation in terms of perspective taking, according to which listeners need to consider the speaker's perspective in order to block coreference between the object pronoun and the subject of the same sentence. Unlike what is predicted by alternative theoretical accounts, performance on pronoun interpretation was not associated with working memory, and the children made virtually no errors in their production of object pronouns. As the difficulties with pronoun interpretation were similar for children with ASD, children with ADHD and typically developing children, this suggests that certain types of perspective taking are unaffected in children with ASD and ADHD.

Keywords: ADHD, autism, inhibition, language acquisition, perspective taking, syntax, theory of mind, working memory

OPEN ACCESS

Edited by:
Julie Franck,
Université de Genève, Switzerland

Reviewed by:
Eniko Ladanyi,
Vanderbilt University Medical Center, United States
Cornelia Hamann,
University of Oldenburg, Germany
Geraldine Legendre,
Johns Hopkins University, United States

***Correspondence:**
Petra Hendriks
p.hendriks@rug.nl

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

Received: *25 September 2020*
Accepted: *04 May 2021*
Published: *04 June 2021*

Citation:
Kuijper SJM, Hartman CA and Hendriks P (2021) Children's Pronoun Interpretation Problems Are Related to Theory of Mind and Inhibition, But Not Working Memory. Front. Psychol. 12:610401. doi: 10.3389/fpsyg.2021.610401

INTRODUCTION

A fundamental aspect of children's language acquisition is learning what the linguistic expressions in their language refer to. Proper names (e.g., *John*) generally have a fixed reference. In contrast, personal pronouns (e.g., *he*, *she*, *him*, *her*) and reflexives (e.g., *himself*, *herself*) depend on other words in the sentence or the discourse for their interpretation. For instance, in the sentence "Paul got upset when John accidentally hit him" the object pronoun *him* refers back to the subject of

the previous clause, *Paul*. The fact that *him* cannot refer back to the subject of the same clause, *John*, indicates that not only the linguistic discourse, but also grammatical principles play a role. These grammatical principles also apply to reflexives, such as *himself*, which must refer back to the subject of the same clause and cannot refer back to the subject of a previous clause. The patterns of use and interpretation of pronouns and reflexives have been the focus of much theoretical work in linguistics, including Chomsky's syntactic binding theory (Chomsky, 1981), later revisions of binding theory such as Reinhart and Reuland's reflexivity account (Reinhart and Reuland, 1993) and Reuland's primitives of binding account (Reuland, 2011), and pragmatic alternatives to binding theory such as Levinson (1991, 2000).

Already early on it was realized that language acquisition research can inform linguistic theorizing (e.g., Chien and Wexler, 1990). In Chomsky's original conception of binding theory, the use and interpretation of pronouns and reflexives is governed by two related principles of the grammar: Roughly speaking, Principle A requires reflexives in simple transitive sentences to refer to the same referent as the subject (resulting in a so-called coreferential interpretation), and Principle B requires pronouns not to corefer with the subject. It is thus expected that children would show mastery of pronouns and reflexives at more or less the same moment in their language development. However, language acquisition research revealed that children's interpretation of object pronouns in English is delayed in comparison to their interpretation of reflexives (e.g., Chien and Wexler, 1990; Grimshaw and Rosen, 1990). For example, in a situation in which two referents are present in the discourse, children until the age of 6 incorrectly allow the object pronoun in sentences like (1) to corefer with the subject. At the same time, they correctly interpret reflexives such as in (2) as coreferring with the subject from age 4.

(1) The elephant is hitting him.
(2) The elephant is hitting himself.

This phenomenon is known as the Delay of Principle B Effect, or Pronoun Interpretation Problem. Only around the age of 10 or 11 years old, children's performance on object pronoun interpretation is adult-like (Philip and Coopmans, 1996; Başkent et al., 2013).

In English, the pronoun *him* and the reflexive *himself* are quite similar in form. In contrast, the Dutch pronoun *hem* ('him') and the Dutch reflexive *zichzelf* ('himself/herself') are clearly distinct forms. Nevertheless, the Pronoun Interpretation Problem is also observed in Dutch (e.g., Philip and Coopmans, 1996; Spenader et al., 2009; van Rij et al., 2010). This indicates that the Pronoun Interpretation Problem is not caused by children's confusion of the two forms. Whereas the Pronoun Interpretation Problem occurs in children's typical acquisition of English, Dutch and several other languages, it does not occur in all languages and for example is absent in Romance languages. Thus, the Pronoun Interpretation Problem is not a universal phenomenon in language acquisition but rather appears to depend on certain grammatical properties of the language. As yet, no satisfactory explanation has been given for this cross-linguistic variation, since it is not clear what properties the languages have in common that show or do not show a Pronoun Interpretation Problem. For example, whereas English and Dutch show the Pronoun Interpretation Problem, the closely related language German does not (Ruigendijk, 2008; see also Ruigendijk et al., 2010); as such, German patterns with the Romance languages, which differ from German in that they have clitic pronouns. Although this cross-linguistic variation in the Pronoun Interpretation Problem is relevant for generalizing the findings of the present study, the present study focuses on Dutch with the aim to shed more light on the interaction between grammar and cognitive processes in pronoun interpretation in languages that show a Pronoun Interpretation Problem.

Explaining the Pronoun Interpretation Problem

In the linguistic literature, various explanations have been put forward for the Pronoun Interpretation Problem. The three explanations most relevant for the current study are discussed below, namely the pragmatic explanation, the working memory explanation, and the perspective taking explanation. These explanations all assume that the interpretation of reflexives is fully determined by the grammar, but that the interpretation of pronouns requires some additional process: pragmatics, reference-set-computation, or bidirectional optimization.

Chien and Wexler (1990) argue that children possess the relevant grammatical knowledge of the binding principles required for a mature interpretation of object pronouns and reflexives (cf. Chomsky, 1981), but still lack the pragmatic skills for their mature usage in context (cf. Thornton and Wexler, 1999). Chien and Wexler's (1990) pragmatic explanation is based on a distinction between syntactic binding (e.g., the relation between the reflexive *himself* and the quantified subject *every elephant* in the sentence "Every elephant is hitting himself") and pragmatic coreference (e.g., the relation between the pronoun *he* and its non-local referential antecedent *an elephant* in the sentence pair "There is an elephant. He is large"). According to their explanation, children have knowledge of the restrictions on syntactic binding but have difficulty with the restrictions on pragmatic coreference. In particular, Chien and Wexler refer to so-called 'accidental coreference' as a source of confusion for children. Accidental coreference occurs when the object pronoun and the referential subject of the sentence accidentally refer to the same individual (as *he* and *him* do in "That must be John. At least he looks like him"), despite the fact that this is disallowed by the binding principles. Accidental coreference is only possible in certain (rare) contexts. To explain why English-speaking children show a Pronoun Interpretation Problem, Chien and Wexler (1990) argue that children have pragmatic difficulty with distinguishing between contexts in which accidental coreference is permitted and contexts in which it is not. Crucially, accidental coreference is not allowed in sentences like (1), but children may not yet have knowledge of this pragmatic restriction.

Under the view that children's errors with object pronouns are due to their confusion about accidental coreference, children should also show problems in pronoun production and use object

pronouns to express a coreferential meaning in all contexts, so also in contexts in which accidental coreference is not allowed. However, English-speaking children between 2;3 and 3;1 years old already produce object pronouns correctly in their spontaneous speech (Bloom et al., 1994) and English- and Dutch-speaking children's production of object pronouns in an experimental setting was found to be adult-like from age 4;6 (for English: De Villiers et al., 2006; Matthews et al., 2009; for Dutch: Spenader et al., 2009). This makes an explanation in terms of lack of pragmatic skills unlikely. Additionally, the distinction Chien and Wexler (1990) found between children's pronoun interpretation in syntactic binding environments and pragmatic coreference environments has been questioned by later studies as an artifact of their experimental materials (e.g., Elbourne, 2005; Conroy et al., 2009).

More recent explanations of the Pronoun Interpretation Problem attribute this problem to children's limited processing resources (e.g., Reinhart, 2006, 2011; Ruigendijk et al., 2011). For example, Reinhart (2011) argues that the Pronoun Interpretation Problem results from children's insufficient working memory capacity (see also Grodzinsky and Reinhart, 1993; Montgomery and Evans, 2009). According to Reinhart (2011), there are two means by which object pronouns can be interpreted: by syntactic binding and by pragmatic coreference. If the grammar allows two interpretational possibilities, the process of reference-set computation is required (Reinhart, 2006, 2011). Reference-set computation compares the different structures and their interpretations, and discards an interpretation if there is a more economical way to obtain that interpretation. Adults use reference-set computation to block pragmatic coreference between an object pronoun and the local subject, as pragmatic coreference is assumed to be a less economical way to express a coreferential interpretation than syntactic binding. Reinhart claims that children have insufficient working memory to perform this costly computation and therefore resort to guessing in their interpretation of object pronouns (Reinhart, 2011). Reference-set computation does not apply in production, since speakers already know which meaning they want to express. Therefore, children's production of object pronouns is predicted to be adult-like (Reinhart, 2006).

Another explanation of the Pronoun Interpretation Problem linking this problem to children's cognitive limitations is proposed by Hendriks and Spenader (2006). They argue that the Pronoun Interpretation Problem is caused by core properties of the grammar itself. Instead of formulating their account in terms of universally valid syntactic principles, as Chien and Wexler (1990) and Reinhart (2006, 2011) do, they formulate their account in terms of violable constraints that differ in strength, as in Optimality Theory (Prince and Smolensky, 2004). The optimal form or meaning is the form or meaning that satisfies the constraints of the grammar best. The constraints determine, for a given input, what is the optimal output for that input in production (when the input is a meaning and the output is a form) or comprehension (when the input is a form and the output is a meaning). As the constraints of the grammar are sensitive to whether they evaluate forms or meanings, they may yield a different form-meaning mapping in comprehension than in production (Smolensky, 1996). To achieve communicative success in spite of these potentially different outcomes in production and comprehension, it has been argued that production and comprehension must be taken into account simultaneously in determining the mature pattern of forms and meanings, through a procedure known as bidirectional optimization (e.g., de Hoop and Krämer, 2006; Legendre et al., 2016). This procedure of bidirectional optimization can be seen as the formalization, within the grammar, of the process of perspective taking (Hendriks, 2014).

In Hendriks and Spenader's (2006) constraint-based account, the constraints of the grammar select both a coreferential and a non-coreferential interpretation as the optimal meaning for an object pronoun, resulting in ambiguity for this pronoun. When encountering an object pronoun, adult listeners are able to block the coreferential interpretation for the pronoun by taking into account the perspective of the speaker: if the speaker would have wanted to express a coreferential interpretation, the speaker would have used a reflexive instead of a pronoun. Since the speaker did not use a reflexive, the speaker must have intended to express a non-coreferential interpretation. Young children are argued to not yet be able to take into account the perspective of the speaker in their interpretation of object pronouns in a consistent way. Hence, they consider pronouns to be ambiguous, thus showing the Pronoun Interpretation Problem. Such perspective taking is expected to require theory of mind abilities (Hendriks, 2014). Indeed, first-order theory of mind is generally acquired well before children show adult-like performance on pronoun interpretation (De Villiers et al., 2006). Furthermore, perspective taking may also require inhibition skills, since the listener must suppress the coreferential meaning in order to select the correct non-coreferential meaning for the pronoun. In Hendriks and Spenader's constraint-based account, the same constraints giving rise to ambiguity of object pronouns in comprehension result in the correct interpretation of reflexives in comprehension and the correct selection of a pronoun or reflexive in production. Thus, children's production of object pronouns is predicted to be adult-like.

The role of inhibition is not only compatible with the perspective-taking explanation, but in principle follows from all accounts of pronoun processing that assume several potential antecedents for the pronoun to be activated during initial stages of processing and assume the grammatical antecedent to compete with binding theory-incompatible antecedents (e.g., Badecker and Straub, 2002; Clackson et al., 2011). Inhibition is needed to subsequently suppress the antecedent that is incompatible with the binding principles. This contrasts with so-called initial-filter models of pronoun processing, that assume that the principles of binding theory are applied early during sentence processing and act as an initial filter, immediately ruling out antecedents that are not compatible with the binding principles (Nicol and Swinney, 1989).

In sum, while the pragmatic explanation attributes children's pronoun interpretation problems to their lack of pragmatic knowledge and predicts that children also make errors with pronouns in production, Reinhart's explanation based on reference-set computation predicts that errors in pronoun

interpretation are caused by insufficient working memory, and Hendriks and Spenader's explanation based on bidirectional optimization predicts that these errors result from a failure to take into account the speaker's perspective, which requires theory of mind abilities and inhibition skills.

Language in Children With ASD and Children With ADHD

The present study aims to clarify how children acquire object pronoun interpretation and production by investigating the role of three possible underlying cognitive mechanisms in pronoun interpretation and production, namely working memory, theory of mind, and inhibition. We designed our study in such a way that we maximized the variation in cognitive mechanisms as well as outcome measures by including children with autism spectrum disorder (ASD), children with attention-deficit/hyperactivity disorder (ADHD) and a group of typically developing (TD) children in our sample. Children with ASD are known to have difficulties in social interaction and communication and show restricted, repetitive behaviors and interests (DSM-5, American Psychiatric Association, 2013). Children with ADHD show a persistent pattern of inattention and/or hyperactivity–impulsivity (DSM-5, American Psychiatric Association, 2013).

Problems in theory of mind have been frequently reported in children with ASD (e.g., Baron-Cohen et al., 1993) and sometimes in children with ADHD (Buitelaar et al., 1999; Kuijper et al., 2015, 2017; Mary et al., 2016). Furthermore, working memory problems and problems in inhibition have been reported in children with ADHD and children with ASD (e.g., Pennington and Ozonoff, 1996; Nydén et al., 2001; Geurts et al., 2004b; Hill, 2004; Martinussen et al., 2005; Happé et al., 2006; Schmitt et al., 2018; Habib et al., 2019). Thus, the deficits in social and cognitive functioning found in children with ASD partly overlap with those in children with ADHD (Bishop and Baird, 2001; Nijmeijer et al., 2010; Rommelse et al., 2011; Demopoulos et al., 2013; Johnson et al., 2015).

Besides difficulties with theory of mind, working memory, and inhibition, both children with ASD and children with ADHD exhibit problems with language and communication. Pragmatic problems are among the core deficits of ASD (DSM-5, American Psychiatric Association, 2013). While the pragmatic deficits in ASD are well documented, less is known about problems in ASD with the structural, or morphosyntactic, properties of language. Some studies did not find morphosyntactic impairments in children with ASD (Bartolucci et al., 1976; Tager-Flusberg, 1981). In contrast, other studies found evidence for morphosyntactic impairments or delays in (subgroups of) children with ASD (Kjelgaard and Tager-Flusberg, 2001; Eigsti et al., 2007; Durrleman et al., 2017). These results indicate that there is considerable heterogeneity in language impairments in ASD (for an overview, see Boucher, 2012).

In ADHD, language deficits are not part of the diagnosis. However, recent studies using parental and teacher questionnaires suggest that in children with ADHD pragmatic use of language is often impaired (for an overview, see Green et al., 2014). Most studies investigating language impairments in ADHD did not find morphosyntactic impairments in children with ADHD (e.g., Kim and Kaiser, 2000; Geurts et al., 2004a; Geurts and Embrechts, 2008; Helland et al., 2012), but some did (Oram et al., 1999; Papaeliou et al., 2015). The language and communication problems of children with ADHD may therefore partly overlap with those observed in children with ASD (e.g., Geurts and Embrechts, 2008).

Although the findings on morphosyntactic impairments of children with ASD and ADHD are equivocal, it may well be that children with ASD or ADHD experience a greater delay in object pronoun interpretation than typically developing children, due to cognitive deficits. Perovic et al. (2013), however, found that high-functioning children with ASD and TD children demonstrated similar difficulties in their comprehension of object pronouns in English. To our knowledge, object pronoun interpretation has not been investigated yet in children with ADHD. The production of object pronouns has been studied in ASD, but mainly in languages such as French and Greek that have clitic pronouns occurring in a special position to the immediate left of the verb (e.g., Terzi et al., 2014; Tuller et al., 2017; Prévost et al., 2018). This contributes to the complexity of the construction and may explain the difficulty these children have with the production of clitic object pronouns. Thus, in addition to our main aim of investigating possible cognitive mechanisms underlying the Pronoun Interpretation Problem, our study will also yield further insight into the relation between pronoun comprehension and pronoun production in children with ASD and children with ADHD.

In our study we focus on children in the age range of 6–12 years, as in this age range in TD children the Pronoun Interpretation Problem gradually decreases (Başkent et al., 2013). Therefore, we expect most variation in object pronoun interpretation performance in this age range. To investigate possible cognitive mechanisms underlying the interpretation of object pronouns, we administer a theory of mind task, a working memory task, and an inhibition task. Following Hendriks and Spenader (2006), object pronoun interpretation is expected to be associated with theory of mind and inhibition. Alternatively, following Reinhart's (2011) account, object pronoun interpretation is hypothesized to be associated with working memory.

MATERIALS AND METHODS

Participants

In total 126 Dutch-speaking children were tested (51 with ASD, 37 with ADHD, and 38 TD children), ranging in age from 6;1 to 12;10 (M = 9;1, SD = 1;9).

Children With ASD

Children in the ASD group were diagnosed with Autistic Disorder (n = 10), PDD-NOS (n = 34) or Asperger's Disorder (n = 7) by independent clinicians on the basis of the DSM-IV-TR criteria (American Psychiatric Association, 2000). Additional inclusion criteria were that the children had a Full Scale Intelligence Quotient (FSIQ) above 75 and verbal communication skills. Furthermore, both the Autism Diagnostic

Interview Revised (ADI-R: Rutter et al., 2003) and the Autism Diagnostic Observation Schema (ADOS, Lord et al., 1999) were administered by certified psychologists. Children in this study were included in the ASD group if they met the ADOS criteria for autism or ASD and/or the ADI-R criteria for autism or ASD (cf. Risi et al.'s ASD2 criteria, Risi et al., 2006). Three children from the ASD group were excluded from further analysis because they did not meet these criteria. One more child was excluded later because he finished neither the pronoun and reflexive comprehension task nor the production task (see section "Procedure"), leaving 47 children in the ASD group. To document the extent to which ADHD symptoms were present, the Parent Interview for Child Symptoms (PICS: Ickowicz et al., 2006) was administered. Seven children in the ASD group scored above the ADHD cut-offs on the PICS (see **Table 1**). In line with their clinical ASD diagnosis, we included these children in the ASD group.

Children With ADHD

Children in the ADHD group were diagnosed with Combined type (n = 19), Predominantly Hyperactive-Impulsive type (n = 12) or Predominantly Inattentive type (n = 6) by independent clinicians on the basis of the DSM-IV-TR criteria (American Psychiatric Association, 2000). Furthermore, both the Parent Interview for Child Symptoms (PICS: Ickowicz et al., 2006) and the Teacher Telephone Interview-IV (TTI-IV: Tannock et al., 2002) were administered by trained clinicians. Six children with ADHD lacked TTI information. Four of them already scored above the cut-off for ADHD based on parent information alone. The remaining two children scored 1 point below the cut-off for ADHD. Since these children scored comparable on the PICS to the other children in the ADHD group (for whom TTI scores combined with their PICS scores reached the cut-off), we included them in the analyses. Seven children in the ADHD group scored within ASD criteria on the ADOS

TABLE 1 | Mean scores (standard deviations) of age, clinical interviews, WISC-III, PPVT, False Belief task, n-back task, and stop task.

	ASD (n = 47)		ADHD (n = 36)		TD (n = 38)		Group differences (Bonferroni corrected *post hoc* analyses)
	M	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	
% Male	87		83		66		
Age	9.3	(1.10)	8.9	(1.7)	9.0	(1.9)	n.s.
ADI-R[1]							
Social Interaction	16.40	(6.06)	4.58	(4.10)	1.82	(3.09)	ASD*** > ADHD > TD*
Communication	12.62	(4.38)	4.03	(2.68)	1.34	(1.55)	ASD*** > ADHD > TD**
Stereotyped Behavior	4.40	(2.59)	1.42	(1.56)	0.32	(0.66)	ASD*** > ADHD > TD*
Behavior < 3 years	3.00	(0.98)	1.47	(1.54)	0.13	(0.41)	ASD*** > ADHD > TD***
ADOS module 3[2]							
Communication	2.67	(1.43)	1.09	(0.92)	0.53	(0.76)	ASD*** > ADHD, TD
Social interaction	7.26	(3.12)	2.57	(1.96)	1.50	(1.72)	ASD*** > ADHD, TD
Com + Soc	9.93	(4.17)	3.66	(2.57)	2.03	(1.99)	ASD*** > ADHD, TD
RRB	1.13	(1.24)	0.29	(0.57)	0.16	(0.44)	ASD*** > ADHD, TD
Social Affect	8.89	(4.19)	2.83	(2.36)	1.74	(2.02)	ASD*** > ADHD, TD
SA + RRB	10.02	(4.68)	3.11	(2.37)	1.89	(2.15)	ASD*** > ADHD, TD
PICS[3]							
Inattention	2.26	(2.07)	3.61	(2.18)	0.11	(0.39)	ADHD** > ASD > TD***
Hyperactivity/impulsivity	1.98	(1.97)	5.22	(2.45)	0.29	(0.57)	ADHD*** > ASD > TD***
WISC-III							
Block Design	9.87	(3.57)	8.33	(3.02)	11.16	(3.23)	ADHD < TD**
Vocabulary	8.81	(3.18)	9.44	(2.10)	11.82	(2.51)	ASD***, ADHD** < TD
Estimated Full scale IQ	96.19	(17.47)	93.26	(12.80)	109.02	(13.64)	ASD, ADHD < TD***
PPVT							
WBQ	104.85	(14.33)	99.97	(12.57)	108.84	(10.72)	ADHD < TD*
False Belief Task							
Proportion correct FB1	0.89	(0.19)	0.88	(0.14)	0.94	(0.11)	n.s.
Proportion correct FB2	0.56	(0.40)	0.55	(0.34)	0.78	(0.29)	ASD, ADHD < TD*
N-Back Task							
Number correct 2back	39.02	(7.95)	38.19	(7.45)	41.77	(5.28)	n.s.
Stop Task							
SSRT	257.39	(96.51)	254.84	(94.25)	256.74	(77.59)	n.s.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. n.s., non-significant.
[1] Five children in the ADHD group scored on the ADI-R above the cut-off for ASD (on the basis of Risi et al.'s criteria, Risi et al., 2006).
[2] Two children in the ADHD group scored above the ADOS criteria for ASD.
[3] Seven children in the ASD group scored within our criteria for ADHD on the PICS (above or one point below the cut-off on the PICS).

or ADI-R (see **Table 1**). In line with their clinical diagnosis, we included these children in the ADHD group. One child was excluded later for task-related reasons (see section "Procedure"), leaving 36 children in the ADHD group.

TD Children

Children in the TD group had not been diagnosed with ASD or ADHD. The ADOS, ADI-R and PICS were administered by trained clinicians in this group as well. None of the children scored above the cut-offs for ASD or ADHD described above.

Materials

Background Variables

IQ of the children was assessed by two subtests (Vocabulary and Block Design) of the Dutch Wechsler Intelligence Scale for Children (WISC-III NL: Kort et al., 2002). Verbal ability was assessed by the Dutch version of the Peabody Picture Vocabulary Test-III (PPVT: Dunn and Dunn, 1997; Schlichting, 2005). Group means and standard deviations for age, IQ, PPVT, and clinical interviews can be found in **Table 1**.

Pronoun and Reflexive Comprehension Task

To test the comprehension of object pronouns and reflexives, we carried out a Picture Verification Task. Children saw one picture at a time. The picture showed two animals engaged in an other-oriented action (**Figure 1**) or a self-oriented action (**Figure 2**).

At the same time, the child heard an introductory sentence, followed by a test sentence with either an object pronoun or a reflexive [see example (3) and (4)].

(3) Introductory sentence:
Een krokodil en een olifant zijn op de stoep.
'An alligator and an elephant are on the sidewalk.'

(4) Test sentence:
De olifant slaat hem/zichzelf.
'The elephant is hitting him/himself.'

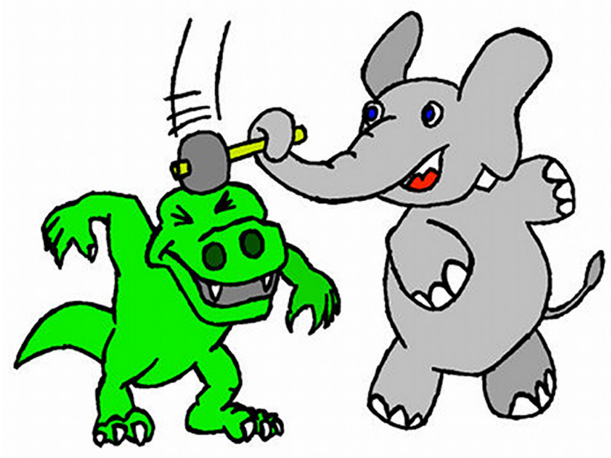


FIGURE 1 | Example of picture showing an other-oriented action.

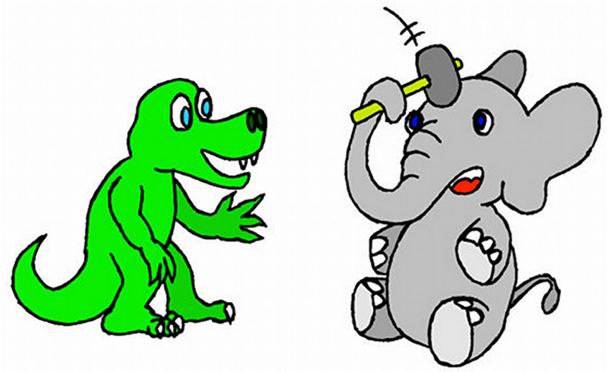


FIGURE 2 | Example of picture showing a self-oriented action.

The materials were based on the materials of Spenader et al. (2009) and van Rij et al. (2010). The transitive verbs that were used in the test sentences were the Dutch translations of *to tickle*, *to hit*, *to bite*, *to point to*, *to draw*, *to paint*, *to tie*, *to make up*, and *to dress*. The child was asked whether or not the recorded sentence matched the picture. Children had to respond by pressing the yes-key when the sentence matched the picture, and by pressing the no-key when the sentence did not match the picture. On trials for which the children decided that the sentence did not match the picture, they were asked to explain why. A second tester noted these justifications.

The task started with two practice items to determine whether the children understood the task. The comprehension task consisted of 34 items: 2 practice items, 16 test items (eight items in the reflexive condition and eight items in the pronoun condition), and 16 control items without an object pronoun or reflexive. The control items were included to measure children's general understanding of the task. In half of the items the sentence matched the picture (match condition). In the other half of the items the sentence and the picture did not match (mismatch condition). Mismatch items contained either a picture of an other-oriented action in combination with a sentence with a reflexive, or a picture of a self-oriented action in combination with a sentence with a pronoun.

We expect children exhibiting the Pronoun Interpretation Problem to make more errors in the pronoun mismatch condition than in the pronoun match condition (cf. Chien and Wexler, 1990; van Rij et al., 2010). Because these children are expected to allow both interpretations of the object pronoun, they will correctly accept the non-coreferential interpretation in the match condition, but also incorrectly accept the coreferential interpretation in the mismatch condition, leading to lower performance on the mismatch condition than on the match condition. Furthermore, we expect these children to not make errors in the reflexive condition if they are not impaired in their syntactic abilities. The reflexive conditions (match and mismatch) thus serve as control conditions to measure children's mastery of the syntactic knowledge required for interpreting the test sentences.

Pronoun and Reflexive Production Task

To check whether children's production of object pronouns and reflexives is adult-like in the same sentence context that is used in the comprehension task, we carried out a sentence elicitation task. This production task was designed to be similar to the comprehension task, as it is well-known from the literature on linguistic reference that pronoun interpretation and use is highly dependent on contextual features such as the structure of the linguistic discourse and visual information. This also holds for object pronouns in simple transitive sentences, which were used in the comprehension task. For example, Dutch-speaking children's as well as adults' online processing of object pronouns in simple transitive sentences is influenced by the linear order in which the potential antecedents of the pronoun are mentioned in the preceding sentence (van Rij et al., 2016). To rule out the possibility that observed differences between production and comprehension outcomes are caused by subtle differences in verbal or visual materials or task instructions, we kept the two tasks as similar as possible. Thus, the production task allows us to test whether the children obey the binding principles in production.

The visual materials of the production task were based on the materials of Spenader et al. (2009). Pictures that were used in the production task were similar to those in the comprehension task. When a picture with an other-oriented action was used in the comprehension task, the corresponding picture with the self-directed action was used in the production task and vice versa. In this way, no picture was shown in both comprehension and production, to avoid possible priming effects. The production task consisted of 16 items in total: two practice items and 14 test items. No filler items were used. Half of the items displayed an other-oriented action, the other half a self-oriented action.

Children saw one picture at a time. They were instructed to first introduce both animals and then to describe the action, leading to sentences like "I see an elephant and a crocodile. The elephant is hitting himself." The production task started with two practice items to determine whether the children understood the task, before they were presented with the test items.

Theory of Mind

To test theory of mind, we used a second-order False Belief task adopted from Hollebrandse et al. (2014). False Belief tasks require one to understand that another person has his or her own beliefs and that these can be different from one's own beliefs (e.g., Baron-Cohen et al., 1985). The task measured both first-order False Belief (FB) (involving the belief of another person) and second-order FB (involving the belief of another person about someone else's belief). We used Hollebrandse et al.'s (2014) verbal rather than their low-verbal second-order FB task for this study, as their low-verbal task turned out to be much more demanding for children in the age range tested than their verbal task, for reasons unrelated to theory of mind abilities (see Hollebrandse et al., 2014, for discussion). As most typically developing children pass first-order FB tasks around age 4 (see the meta-analysis of Wellman et al., 2001), and our participant group is between 6 and 12 years old, we expect ceiling performance on the first-order FB questions. Therefore, of specific interest to our study is children's performance on the second-order FB questions.

Each story in the FB task starts with an initial belief that is shared by the two main characters in the story (e.g., Sam and Maria both believe that they are selling cookies at the bake sale). This belief changes in the middle of the story for the first character without the second character knowing about this (e.g., while Maria has gone out to buy cookies, Sam hears that they are selling apple pie instead), and next changes for the second character without the first character knowing about this (e.g., Maria finds out at the bake sale that they are only selling waffles, without Sam knowing about this). As a result, the story involves three distinct beliefs: the second character's true belief about the actual situation and two false beliefs. The first-order FB question asks about the first character's false belief about the situation (e.g., what does Sam think they are selling at the bake sale?). The second-order FB question asks about the second character's false belief about the first character's belief, and is broken down into two separate questions to avoid asking syntactically too complex questions (e.g., Maria is asked what Sam thinks they are selling at the bake sale, and then the child is asked what Maria will answer, thus effectively asking what Maria thinks that Sam thinks they are selling at the bake sale). See Hollebrandse et al. (2014, Appendix 1) for a sample item.

The task consisted of eight stories read to the child by the experimenter. Each story was accompanied by four pictures that were presented one by one on a computer screen. The task was divided in two blocks with a short break in between. The order of stories was counterbalanced across participants. Each story contained one second-order FB question and two first-order FB questions. The first first-order FB question was asked in the middle of the story, when the first false belief was introduced. At the end of the story, the second-order FB question was asked, followed by the first-order FB question. The first-order FB question was asked again at the end of the story in order to check whether children had difficulties with the length and complexity of the story.

One item was removed from further analysis since item analysis showed that the response on this item differed from the other seven items: on the second first-order FB question, mean accuracy on this item was only 0.48, while mean accuracy on other items varied between 0.79 and 0.92. Additionally, on this item, mean accuracy on the second-order FB question was higher (0.80) than on the easier first-order FB question (0.48). Inspection of this item revealed that its content differed from the other items in that an extra belief had inadvertently been introduced, which made the correct first-order FB answer less plausible. Two dependent measures were calculated: mean accuracy on the first first-order FB question (FB1) and mean accuracy on the second-order FB question (FB2).

Working Memory

Working memory is the ability to temporarily maintain and manipulate information (Baddeley and Hitch, 1974). It can be operationalized in different ways. Because of the known language and communication difficulties of children with ASD and children with ADHD, we wanted to reduce the verbal load

of the working memory task by using a visual task, rather than a verbal task such as a listening span task or digit span task. Specifically, we operationalized working memory by the n-back task (Owen et al., 2005). The n-back task is a continuous performance task to measure working memory capacity. The task is commonly used in psychology and cognitive neuroscience (e.g., Williams et al., 2005; Cui et al., 2010; Chatham et al., 2011; de Vries and Geurts, 2014) and requires sustained maintenance and updating of information in working memory.

The n-back task in our study included three experimental conditions: 0-back (baseline), 1-back, and 2-back. In each condition, pictures were presented on a computer screen with a stimulus duration of 1000 milliseconds, followed by an interstimulus interval of 1500 milliseconds. In the 0-back condition, participants were instructed to press the yes-button when they saw a picture of a car, and to press the no-button when another picture appeared. In the 1-back condition, participants had to press the yes-button when the picture matched the picture immediately preceding it, and otherwise press the no-button. In the 2-back condition, participants had to press the yes-button when the picture matched the picture that appeared two pictures back, and otherwise press the no-button. Studies have shown that 2-back tasks seem suitable for children in our age range (e.g., Schleepen and Jonkman, 2010). The task was divided in different blocks, which were presented in random order. Each block started with 0-back, followed by 1-back and then 2-back. In this way, children got used to the task and were able to understand the more difficult 2-back condition. Participants started with a practice session of 15 trials per condition (0-, 1-, and 2-back), followed by the test session consisting of four blocks of 15 trials per condition (resulting in a total of 60 trials per condition). The total number correct on the 2-back condition was calculated as a measure of working memory.

Response Inhibition

The study also included a task to measure response inhibition. Response inhibition is the capacity to suppress an ongoing motor response that is no longer relevant. To capture response inhibition, we used a stop task, which is considered a relatively pure, reliable and valid measure of prepotent response inhibition (Tannock et al., 1989; Kindlon et al., 1995; de Vries and Geurts, 2014). Like the n-back task, the stop task is often used in psychology and cognitive neuroscience, and measures individual, clinical and developmental differences in the inhibition of responses. In this study we adopted the stop task from van den Wildenberg and Christoffels (2010). This is a non-verbal response inhibition task, which we preferred over a verbal task for the same reason as mentioned for the n-back task.

In this stop task, simple drawings of a tree and a door were presented on the computer screen. During go-trials, participants were asked to press the button corresponding with the picture on a two-button box. In 30% of the trials, a visual stop-signal was presented: a red square frame surrounding the picture border. When confronted with the stop-signal, participants had to inhibit the go-response by not pressing the button. The interval between the onset of the go-picture and the onset of the stop-signal (stop-signal delay) was set at 200 ms on the first stop-trial. An online tracking algorithm adjusted stop-signal delay as a function of individual stopping performance (Levitt, 1971). If the participant was able to stop, the stop-signal delay increased by 50 ms, thereby decreasing the chances of successful inhibition on the next stop-trial. After a failed-inhibition trial, the stop-signal delay decreased by 50 ms. This adaptive algorithm ensured successful inhibition on about 50% of the stop-trials, a procedure that yields reliable estimates of the Stop Signal Reaction Time (SSRT: Band et al., 2003). SSRT was calculated as a measure of response inhibition.

Procedure

Children and their parents were recruited by brochures at schools and in outpatient clinics for child and adolescent psychiatry in Groningen. They took part in a larger study on language and communication in ASD and ADHD (Kuijper, 2016). The study was reviewed and approved by the research ethics committee CETO of the University of Groningen. Parents of all child participants gave written informed consent prior to participation in the study. Children and parents came to the lab together. Children were tested individually on a single day in a quiet testing room with two experimenters present. After every task children had a short break.

Two participants were excluded from further analysis: one (ASD) because he finished neither the comprehension task nor the production task, leaving 47 children in the ASD group, and the other (ADHD) because he scored below 0.75 on the control items in the comprehension task, leaving 36 children in the ADHD group. Furthermore, one child (ASD) conducted only half of the False Belief task and was removed from analyses involving this task. One child (ASD) did not finish the n-back task and was removed from analyses involving the n-back task. Another child (ADHD) did not complete the stop task and consequently was excluded from analyses involving this task. Finally, one child (ADHD) finished neither the n-back nor the stop task and was excluded from analyses including these tasks.

Coding of Production Data

Children's answers on the production task were voice-recorded. Only active transitive sentences containing a subject and an object that referred to one of the two animals in the picture were included in analyses (93.1% of all items). In the production task, more answers are acceptable than only object pronouns or reflexives. For pictures showing an other-oriented action, the use of a full noun phrase (e.g., "the elephant is hitting the crocodile") to describe such actions is compatible with the binding principles. In fact, such a choice is pragmatically felicitous as well, as adults produce mainly full noun phases in this sentence context (see Spenader et al., 2009). Both the use of object pronouns (e.g., "the elephant is hitting him") and the use of full noun phrases were therefore coded as correct responses in this condition. For pictures showing a self-oriented action, only the use of a reflexive (e.g., "the elephant is hitting himself") was treated as accurate. All items were scored independently by two coders, who were blind to the participant's diagnosis. The coders scored the grammatical form of the object (pronoun, reflexive, or full noun phrase). Inter-scorer agreement was high (Cohen's $\kappa = 0.95$).

Data Analysis

The data were analyzed using Generalized Linear Mixed Models (GLMM). A logit link was used to accommodate the repeatedly measured binary outcome variable (i.e., accuracy of pronoun interpretation, denoted below as Accuracy) (Jaeger, 2008; Heck et al., 2012). Compound symmetry was used as covariance matrix. First we tested for differences between groups in pronoun comprehension. Contrasts between diagnostic groups and controls (ASD vs. TD and ADHD vs. TD) were dummy-coded and included as fixed factors in the analysis. Whether the sentence matched the picture (coded as 0) or not (coded as 1) was additionally included as a fixed factor. This last factor was included because previous studies showed clear differences between match and mismatch conditions, likely caused by a yes-bias (see also Chien and Wexler, 1990; van Rij et al., 2010). In addition to these three main effects (denoted as ASD, ADHD, and Match) we included two two-way interactions (ASD*Match, ADHD*Match) in the model. A two-way interaction or main effect that had no effect on Accuracy ($p > 0.05$) was removed from the model.

Next, we examined possible cognitive mechanisms underlying object pronoun interpretation by including the relevant parameters derived from the False Belief task (FB1 and FB2), the n-back task (working memory, or WM), and the stop task (SSRT), respectively. All four were mean-centered around a value of zero and were included, in four separate analyses, as fixed factors in the aforementioned model. Interactions that had no effect on Accuracy ($p > 0.05$) were removed from the model. Finally, we tested whether found associations held up when all main and interaction effects with a significance value of $p \leq 0.05$ were examined simultaneously in a multiple GLMM analysis.

RESULTS

Pronoun and Reflexive Comprehension Task

In line with our expectations, neither the reflexive match condition nor the reflexive mismatch condition yielded a substantial number of errors (see **Table 2**). Therefore we did not statistically test for differences in reflexive interpretation between the groups. Below, our focus is on the two object pronoun conditions. Despite the rather small differences in performance in the object pronoun conditions, there was enough variance to build a meaningful GLMM.

Clinical Groups

As expected (Chien and Wexler, 1990; van Rij et al., 2010), a significant effect of Match was found (see **Table 3**), indicating that more errors were made in the object pronoun mismatch condition than in the object pronoun match condition. Interactions of ASD or ADHD with Match did not contribute significantly to participants' scores on the comprehension task (all p-values > 0.05), showing that this effect held for all groups. In addition, the main effects of ASD and ADHD did not significantly contribute to Accuracy.

With no differences among the groups, we conclude that errors in object pronoun interpretation are not explained by the presence of ASD or ADHD. In subsequent analyses, main and interaction effects related to diagnostic group were removed, leaving a model that included two main effects (Mechanism and Match) and one interaction effect (Mechanism*Match).

Because the TD group differs from the ASD and ADHD group in mean IQ-score and the TD group differs from the ADHD group in mean PPVT-score (see **Table 1**), we checked post-hoc if group differences in pronoun interpretation between ASD, ADHD and TD emerge, by (i) selecting part of our TD group ($n = 27$) to match the IQs of both other groups, and by (ii) selecting part of our TD group ($n = 34$) to match the PPVT of the ADHD group. No group differences in pronoun interpretation emerge when we use the subgroups matched on IQ or verbal ability in the two *post hoc* analyses (see **Table 4**).

Mechanisms

No interaction effect of Match with any of the cognitive mechanisms was found (all p-values > 0.05). Therefore, in the final model only the main effects of each of the cognitive mechanisms and Match were included, first separately, and next in the multiple GLMM. We found a main effect of FB2 (see **Table 5**). Lower scores on second-order False Belief questions were associated with lower Accuracy scores in both the Object pronoun match and the Object pronoun mismatch condition. We also found a significant main effect of SSRT. Higher SSRT scores

TABLE 3 | Estimated effects for Clinical group and Match on accuracy in object pronoun interpretation.

Predictor	Estimate	*SE*	*p*
Match	−1.94***	0.28	<0.001
ASD vs. TD	−0.26	0.50	0.60
ADHD vs. TD	−0.79	0.50	0.11

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE 2 | Mean proportions correct responses and standard deviations per group and per condition in the comprehension task.

	Object pronoun match		Object pronoun mismatch		Reflexive match		Reflexive mismatch	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
TD	0.98	0.07	0.87	0.29	0.99	0.04	1	0.00
ASD	0.98	0.07	0.84	0.28	0.99	0.05	0.99	0.05
ADHD	0.94	0.12	0.75	0.34	0.97	0.08	0.99	0.04

TABLE 4 | Estimated effects for Clinical group and Match on accuracy in object pronoun interpretation for the IQ-matched subgroup and for the PPVT-matched subgroup.

	IQ-matched subgroup (n = 110)			PPVT-matched subgroup (n = 117)		
Predictor	**Estimate**	***SE***	***p***	**Estimate**	***SE***	***p***
Match	−1.97***	0.30	<0.001	1.94***	0.28	<0.001
ASD vs. TD	−0.13	0.51	0.81	−0.14	0.50	0.79
ADHD vs. TD	−0.67	0.51	0.19	−0.67	0.50	0.18

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE 5 | Estimated effects of Mechanism and Match on accuracy in object pronoun interpretation.

	FB1			FB2			SSRT			WM		
Predictor	**Estimate**	***SE***	***p***	**Estimate**	***SE***	***p***	**Estimate**	***SE***	***p***	**Estimate**	***SE***	***p***
Match	−2.00***	0.30	<0.001	−2.02***	0.30	<0.001	−2.05***	0.30	<0.001	−1.99***	0.29	<.001
Mechanism	1.71	1.00	0.085	1.23*	0.51	0.017	−0.006**	0.002	0.008	0.031	0.02	0.13

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

(indicating lower inhibition) were associated with more errors in the object pronoun conditions. No significant effects of FB1 or working memory were found. In all four analyses, the main effect of Match remained significant: more errors were made in the object pronoun mismatch condition than in the object pronoun match condition.

FB2, SSRT and Match were included in a multiple GLMM (**Table 6**). All aforementioned associations remained significant. Thus, when adjusted for the effect of SSRT, lower scores on FB2 questions were still associated with lower Accuracy scores in the object pronoun conditions. Vice versa, when adjusted for the effect of FB2, higher SSRT scores were still associated with lower Accuracy scores in the object pronoun conditions. Furthermore, a main effect for Match remained: adjusted for the effects of FB2 and SSRT, children still performed worse in the object pronoun mismatch condition than in the object pronoun match condition.

In a *post hoc* analysis we added age to our model. In our study we focused on children in the age range of 6–12 years, during which the Pronoun Interpretation Problem gradually disappears. With age being associated with FB2 and SSRT, age was added to our model to study the extent to which age would subsume the effects of FB2 and SSRT.

Table 7 shows that the effects of FB2 and SSRT were attenuated when age was included, confirming that children's pronoun interpretation errors decrease with age and indicating that age is more strongly linked to object pronoun interpretation than theory of mind and inhibition. The main effect of Match remained significant: children made more pronoun interpretation errors in the mismatch condition than in the match condition.

TABLE 6 | Estimated multiple mechanisms model of accuracy in object pronoun interpretation.

	Estimate	***SE***	***p***
Intercept	3.72***	0.31	<0.001
Match	−2.07***	0.31	<0.001
FB2	1.13*	0.50	0.025
SSRT	−0.005*	0.002	0.018

*Only main effects and interactions $p < 0.05$ in univariate analyses were included in multiple analyses. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.*

Pronoun and Reflexive Production Task

In production, consistent with our expectations, children hardly made any mistakes (see **Table 8**). With all three groups performing at ceiling, we did not test for group differences in production accuracy.

Recall that, for the other-oriented action, both the use of an object pronoun and the use of a full noun phrase were scored as correct responses. Only in 5% of the cases an object pronoun was used. In the remaining 95% of the cases a full noun phrase was used. This corresponds with the pattern of production displayed by Dutch adults, who also mainly used full noun phrases to describe an other-oriented action in a similar experiment (Spenader et al., 2009). Importantly, children hardly ever incorrectly use an object pronoun [4 out of 769 scorable sentences, produced by three children (two ADHD and one ASD)] or a reflexive [4 out of 794 scorable sentences, produced by only one child (ASD)].

We tested, *post hoc*, if children with ASD or ADHD differed from TD children in their use of full noun phrases and object pronouns. A GLMM was performed on all items in the other-oriented condition, with full noun phrase (yes or no) as binary dependent variable and two dummy-coded contrasts between diagnostic groups and controls (ASD vs. TD and ADHD vs. TD) as fixed factors. No significant differences between the groups were found (all p-values > 0.05): children with ASD used a full noun phrase in 96% of the cases, children with ADHD in 95% of the cases and TD children in 94% of the cases. This indicates that children with ASD and children with ADHD use the same linguistic forms as TD children to express other-oriented and self-oriented actions.

TABLE 7 | Estimated multiple mechanisms model of accuracy in object pronoun interpretation, including Age.

	Estimate	*SE*	*p*
Intercept	1.41	0.93	0.13
Match	−2.10***	0.31	<0.001
FB2	0.82	0.53	0.12
SSRT	−0.004	0.002	0.07
Age	0.02**	0.008	0.008

*Only main effects and interactions $p < 0.05$ in univariate analyses were included in multiple analyses. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.*

TABLE 8 | Mean proportions correct responses in the production task.

	Other-oriented action		Self-oriented action	
	Mean	***SD***	**Mean**	***SD***
TD	1.0	0.00	1.0	0.06
ASD	0.99	0.11	0.99	0.08
ADHD	1.0	0.00	0.99	0.12

DISCUSSION

The aim of this study was to clarify how children acquire object pronoun interpretation and production by investigating the possible cognitive mechanisms underlying the Pronoun Interpretation Problem, as different theoretical accounts see a role for different cognitive mechanisms. We found that both second-order False Belief performance and Stop Signal Reaction Time were associated with performance on the object pronoun interpretation task. These results suggest that theory of mind and inhibition are necessary for object pronoun interpretation. This finding is compatible with the perspective taking account of the Pronoun Interpretation Problem by Hendriks and Spenader (2006). According to Hendriks and Spenader's account, object pronouns are potentially ambiguous and listeners must consider the perspective of the speaker to block the incorrect coreferential interpretation for the object pronoun. The results of this study suggest that the Pronoun Interpretation Problem arises if children fail to consider the perspective of the speaker because of insufficient theory of mind abilities, or fail to suppress the incorrect interpretation of the pronoun because of poor inhibition skills.

We did not find a relation between working memory and performance on object pronoun interpretation and thus found no support for Reinhart's (2006, 2011) claim that sufficient working memory is necessary for the costly operation of reference-set computation that is needed for object pronoun interpretation. The absence of a relation with working memory corroborates the results of Perovic and Wexler (2018). They found that children with Williams Syndrome, who are generally reported not to have memory deficits, nevertheless showed difficulties with pronoun interpretation in simple transitive sentences in English. However, these children did not receive a working memory task to confirm that they did not have memory deficits. Contrasting with these findings, in children with Developmental Language Disorders (DLD) Montgomery and Evans (2009) found a relation between working memory, as measured by a listening span task, and performance on the interpretation of complex sentences, including embedded pronominal sentences such as "Bugs Bunny says Daffy Duck is hugging him." However, performance on different sentence types was combined in this study and also included performance on embedded reflexive sentences and passive sentences. Additionally, the embedded pronominal sentences in this study were more complex than the simple transitive pronominal sentences in the current study (see also Ladányi et al., 2017, who found a relation in children with DLD between performance on the n-back task and the interpretation of pronouns and reflexives in embedded sentences in Hungarian). Because of these differences with the current study, it is possible that the relation with working memory reported in previous studies with children with DLD is due to other features of the linguistic materials than the presence of object pronouns, for example the syntactic complexity of the test sentences used. This explanation is supported by the close link found between working memory capacities and complex syntax in children's comprehension of language, as measured with different working memory tasks and different syntactic constructions (Delage and Frauenfelder, 2019).

Regarding children's production of pronouns and reflexives, we did not find support for Chien and Wexler's (1990) pragmatic explanation of the Pronoun Interpretation Problem, as the children hardly made any binding errors in their production of object pronouns or reflexives. That is, they rarely produced a reflexive when a pronoun or full noun phrase was the correct form to use (which would constitute a violation of Principle A), and they rarely produced a pronoun when a reflexive was the correct form to use (which would constitute a violation of Principle B, in generative syntactic terms). Thus, the children observed the constraints of the grammar in their production of these forms. This finding is in line with previous experiments with typically developing children, showing that children produce object pronouns in an adult-like way from a young age (De Villiers et al., 2006; Matthews et al., 2009; Spenader et al., 2009).

Because of their known difficulties with theory of mind, working memory, and inhibition, we had expected children with ASD and children with ADHD to have more problems with object pronoun interpretation than TD children. However, we did not find any differences in object pronoun interpretation between children with ASD, children with ADHD, and TD children: all three groups made errors in object pronoun interpretation. As expected, we also found that all three groups performed at ceiling on the reflexive conditions and on the production task. That is, the TD children as well as the children with ASD or ADHD in our study only had problems with the interpretation of object pronouns (particularly emerging in the mismatch condition), and did not have difficulty with the interpretation of reflexives (either in the match condition or in the mismatch condition) or with the production of pronouns and reflexives.

That children with ASD and TD children show a similar Pronoun Interpretation Problem corroborates the findings by Perovic et al. (2013). Perovic et al. (2013) consider the

Pronoun Interpretation Problem to be pragmatic in nature (cf. Chien and Wexler, 1990). At first glance, this leaves unexplained why they did not find differences between children with ASD and TD children in object pronoun interpretation. After all, if the Pronoun Interpretation Problem is a pragmatic problem, why would children with ASD, who are known for their pragmatic deficits, not make more errors in object pronoun interpretation than TD children?

Perovic et al. (2013) argue that there may be different kinds of pragmatics: a kind of pragmatics related to social rules and a kind of pragmatics more directly related to language (cf. Schaeffer, 2003). This latter so-called "linguistic pragmatics" may not be affected in ASD, according to Perovic et al. (2013). We propose an alternative explanation of these findings. Rather than positing two types of pragmatics, one of which is unaffected in ASD, we propose that perspective taking need not be a pragmatic process but can also be part of the grammar. According to Hendriks and Spenader's (2006) account of the Pronoun Interpretation Problem, the interpretation of an object pronoun requires listeners to take into account the perspective of a hypothetical speaker in order to determine the interpretation of the pronoun (see also Hendriks, 2014). That is, listeners must apply the relevant constraints of the grammar to determine the optimal meaning of the pronoun, and must additionally place themselves in the perspective of a hypothetical speaker and apply the same constraints to determine the optimal form for this optimal meaning. In a final step, the listener must check whether the input form in comprehension and the output form in production match, or in other words: whether a speaker would have used a pronoun to express the selected interpretation. If so, the selected interpretation is considered to be correct, but if not, the selected interpretation must be suppressed and another interpretation must be checked. This process of "grammaticalized perspective taking," which requires listeners to take the perspective of a hypothetical speaker and express a particular meaning as if they were the speaker, may be different from taking the perspective of an actual speaker, who may or may not be sitting in front of the listener. The latter form of perspective taking is much more challenging for listeners, since it differs per speaker and per situation. In contrast, grammaticalized perspective taking may be less demanding, as it does not vary per situation and therefore could be gradually automatized (as is shown in computational cognitive simulations to be psychologically plausible, see van Rij et al., 2010; Vogelzang et al., 2021). Such an automatized process can be understood as being part of the grammar of a mature native speaker.

This view of object pronoun interpretation as a process of grammaticalized perspective taking is supported by the finding of similar difficulties with pronoun interpretation in non-advanced second-language learners as in children acquiring their native language. The finding of a Pronoun Interpretation Problem in second-language learners has been put forward as evidence in favor of Reinhart's costly operation of reference-set computation and against an explanation in terms of lack of linguistic knowledge (Slabakova et al., 2017), but is also consistent with the proposed computationally complex process of perspective taking. These results may thus provide support for the claim that this grammaticalized perspective taking is unaffected in children with ASD and ADHD. This corroborates previous findings of similar linguistic performance in ASD children, ADHD children, and TD children (Kim and Kaiser, 2000; Geurts et al., 2004a; Geurts and Embrechts, 2008; Helland et al., 2012). In contrast, taking the perspective of an actual speaker may be involved in pragmatic skills such as turn-taking and conversational rapport, both of which are found to be impaired in ASD and ADHD (e.g., Geurts et al., 2004a; Green et al., 2014).

Most of the ASD children in our study could be classified as "language normal" (based on their PPVT scores and the vocabulary subtest of the WISC-III, cf. Kjelgaard and Tager-Flusberg, 2001). Perovic et al. (2013) found that the linguistic performance of ASD children with language impairment differed from the linguistic performance of ASD children without language impairment. However, they only found differences in the interpretation of reflexives, while both groups of ASD children performed similarly on the interpretation of object pronouns. A crucial difference between the study of Perovic et al. (2013) and the present study is the type of task that is used. Perovic et al. (2013) used a Picture Selection Task, which tests for preference of interpretation, whereas our study used a Picture Verification Task, which tests for acceptability of interpretation. On the basis of the study of Perovic et al. (2013) it can be concluded that ASD children with language impairment have a preference for a non-coreferential interpretation for object pronouns and reflexives. However, it is not clear whether these children would incorrectly accept a coreferential interpretation for pronouns, which is what the Pronoun Interpretation Problem entails. To further unravel differences between ASD children with and without language impairment, such children could be tested on their interpretation of object pronouns and reflexives using a Picture Verification Task or some other task testing acceptability rather than preference for one of the two relevant interpretations.

A finding of our study that may at first sight be surprising is the fact that we found an effect of second-order False Belief understanding, but no effect of first-order False Belief understanding. The absence of an association between first-order False Belief understanding and object pronoun interpretation is probably due to ceiling performance in first-order False Belief understanding (see **Table 1**). First-order False Belief understanding is generally mastered around age 4, so at least 2 years before object pronouns are understood correctly (De Villiers et al., 2006). Because we expected a ceiling effect for first-order False Belief understanding in our 6- to 12-year-old children, we included a task that also measured second-order False Belief understanding. Since accurate second-order False Belief understanding is dependent on accurate first-order False Belief understanding, a slower development of first-order theory of mind is expected to result in a slower development of second-order theory of mind as well, thus allowing us to investigate the relation between pronoun interpretation and theory of mind

by looking at second-order False Belief understanding. Second-order False Belief understanding was found to relate to object pronoun interpretation, which indicates that perspective taking is important in interpreting object pronouns.

The False Belief task used in this study is a highly verbal task, which also depends on general language skills. Therefore, it could be argued that the observed relation between False Belief understanding and performance on pronoun comprehension merely reflects children's general language abilities. However, if true, we would expect this to also be reflected in children's performance on reflexive comprehension. The reflexive condition can be considered a control condition, assessing children's general language comprehension abilities and, more specifically, their syntactic abilities. Since the children in our study did not have any problems in the reflexive condition, their general language comprehension abilities appear to be intact. Previous studies found significant relations between various aspects of language and False Belief understanding (for an overview, see Milligan et al., 2007). Our study adds to this the observation of a relation between object pronoun interpretation and False Belief understanding. Yet it would be worthwhile to examine the relation between object pronoun interpretation and theory of mind using other theory of mind tasks, for example low-verbal theory of mind tasks or the (more natural) strange stories task (Happé, 1994).

Although we did not find an association between working memory and performance in object pronoun interpretation, it should be kept in mind that working memory is a broad concept and many different tasks for its measurement have been developed. In our study, an n-back task with non-verbal stimuli (pictures) was used. It is possible that working memory tasks with verbal stimuli are associated with object pronoun interpretation. However, meta-analyses show that both working memory tasks with verbal stimuli and with non-verbal stimuli relate to general language comprehension (Daneman and Merikle, 1996) and that both give rise to similar activation patterns in neuroimaging studies (Owen et al., 2005). Additionally, in a related study with the same children (Kuijper et al., 2015), a relation was found between performance on the n-back task with non-verbal stimuli and performance on another linguistic task than the one reported on here. This other linguistic task tested speakers' referential choice between using a pronominal subject and using a full noun phrase subject in production, which is dependent on how well the speaker can keep track of the different referents mentioned in the preceding linguistic discourse. This indicates that the n-back task used in this study relates to at least some aspect of linguistic performance that requires working memory. Since no association was found between the n-back task and performance on object pronoun interpretation in the present study, this strongly suggests that object pronoun interpretation and working memory are unrelated.

In contrast to working memory, inhibition was found to be associated with object pronoun interpretation in our study. In our study, we used a stop task to measure prepotent response inhibition. Yet, it may be worthwhile to also investigate the relation between pronoun interpretation and other types of inhibition, in particular interference control (i.e., cognitive inhibition). A final consideration with regard to the cognitive processes that were studied here pertains to the role of age. In a *post hoc* analysis we added age to our final model, leading to attenuation of the effects of inhibition and theory of mind. Age, as the umbrella variable, was more strongly linked to object pronoun interpretation than the specific effects of theory of mind and inhibition. The effect of age shows that, in addition to theory of mind and inhibition, other cognitive factors are likely involved in pronoun interpretation which also develop with age and which we have not included in this study. These cognitive factors (i.e., the included as well as non-included ones) are all subsumed by the overarching factor of age. Although the theoretical literature on object pronoun interpretation is not explicit about this, possibly cognitive flexibility (to switch from the incorrect interpretation to the alternative correct interpretation, cf. Kissine, 2012) or focused attention (to process speech in real-time, see, e.g., Wolfgramm et al., 2016) play a role too.

In summary, the current study provides insight into the Pronoun Interpretation Problem and the cognitive mechanisms underlying this comprehension delay in children's language development. We found that both theory of mind and inhibition skills were associated with performance on object pronoun interpretation. This provides support for Hendriks and Spenader's (2006) perspective taking account of object pronoun interpretation, which holds that listeners must take into account the perspective of a hypothetical speaker and thus block the incorrect interpretation for the pronoun. Furthermore, our study showed that the performance of children with ASD or ADHD was comparable to that of TD children: the three groups demonstrated similar difficulties in their interpretation of object pronouns and neither of the groups showed difficulties in the production of object pronouns and reflexives. This suggests that children with ASD and children with ADHD do not have more problems than TD children in taking into account the grammatical perspective of a hypothetical speaker, despite their possible difficulties in perspective taking with actual conversational partners.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

This study, involving human participants, was reviewed and approved by the Commissie Ethische Toetsing Onderzoek (CETO) of the University of Groningen. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SK, CH, and PH designed the study. SK carried out the experiment. SK, CH, and PH analyzed the results and wrote the

manuscript. All the authors read and approved the final version of the manuscript.

FUNDING

This investigation was supported by a grant from the Netherlands Organisation for Scientific Research (NWO, Grant No. 277-70-005), awarded to PH.

ACKNOWLEDGMENTS

We are grateful to the children and their parents who participated in this study. We thank Accare University Center for Child and Adolescent Psychiatry Groningen, Autisme Team Noord-Nederland and all research assistants for their help in the data collection. Special thanks go to Jessica Overweg and Gisi Cannizzaro for their assistance and useful suggestions. The content of the manuscript previously appeared as part of the Ph.D. thesis of Sanne Kuijper (2016).

REFERENCES

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edn. Washington, DC: American Psychiatric Association.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Washington, DC: American Psychiatric Association.

Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89.

Badecker, W., and Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *J. Exp. Psychol. Learn.* 28, 748–769. doi: 10.1037/0278-7393.28.4.748

Band, G. P. H., van der Molen, M. W., and Logan, G. D. (2003). Horse-race model simulations of the stop-signal procedure. *Acta Psychol. (Amst).* 112, 105–142. doi: 10.1016/s0001-6918(02)00079-3

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Tager-Flusberg, H., and Cohen, D. (1993). *Understanding Other Minds: Perspectives from Autism*. Oxford: Oxford University Press.

Bartolucci, G., Pierce, S., Streiner, D., and Eppel, P. T. (1976). Phonological investigation of verbal autistic and mentally retarded subjects. *J. Autism Child Schizophr.* 6, 303–316. doi: 10.1007/bf01537908

Başkent, D., van Rij, J., Ng, Z. Y., Free, R., and Hendriks, P. (2013). Perception of spectrally degraded reflexives and pronouns by children. *J. Acoust. Soc. Am.* 134, 3844–3852. doi: 10.1121/1.4824341

Bishop, D. V. M., and Baird, G. (2001). Parent and teacher report of pragmatic aspects of communication: use of the Children's Communication Checklist in a clinical setting. *Dev. Med. Child Neurol.* 43, 809–818. doi: 10.1017/s0012162201001475

Bloom, P., Barss, A., Nicol, J., and Conway, L. (1994). Children's knowledge of binding and coreference: evidence from spontaneous speech. *Language* 70, 53–71. doi: 10.2307/416740

Boucher, J. (2012). Research review: structural language in autistic spectrum disorder - characteristics and causes. *J. Child Psychol. Psychiatry.* 53, 219–233. doi: 10.1111/j.1469-7610.2011.02508.x

Buitelaar, J. K., van der Wees, M., Swaab-Barneveld, H., and van der Gaag, R. J. (1999). Theory of mind and emotion-recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. *Dev. Psychopathol.* 11, 39–58. doi: 10.1017/s0954579499001947

Chatham, C. H., Herd, S. A., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R., et al. (2011). From an executive network to executive control: a computational model of the n-back task. *J. Cogn. Neurosci.* 23, 3598–3619. doi: 10.1162/jocn_a_00047

Chien, Y.-C., and Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Lang. Acquis.* 1, 225–295. doi: 10.1207/s15327817la0103_2

Chomsky, N. (1981). *Lectures on Government and Binding: the Pisa Lectures*. Dordrecht: Foris Publications.

Clackson, K., Felser, C., and Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: evidence from eye-movements during listening. *J. Mem. Lang.* 65, 128–144. doi: 10.1016/j.jml.2011.04.007

Conroy, A., Takahashi, E., Lidz, J., and Phillips, C. (2009). Equal treatment for all antecedents: how children succeed with Principle B. *Linguist. Inq.* 40, 446–486. doi: 10.1162/ling.2009.40.3.446

Cui, J., Gao, D., Chen, Y., Zou, X., and Wang, Y. (2010). Working memory in early-school-age children with Asperger's syndrome. *J. Autism Dev. Disord.* 40, 958–967. doi: 10.1007/s10803-010-0943-9

Daneman, M., and Merikle, P. M. (1996). Working memory and language comprehension: a meta-analysis. *Psychon. Bull. Rev.* 3, 422–433. doi: 10.3758/bf03214546

de Hoop, H., and Krämer, I. (2006). Children's optimal interpretations of indefinite subjects and objects. *Lang. Acquis.* 13, 103–123. doi: 10.1207/s15327817la1302_4

Demopoulos, C., Hopkins, J., and Davis, A. (2013). A comparison of social cognitive profiles in children with autism spectrum disorders and attention-deficit/hyperactivity disorder: a matter of quantitative but not qualitative difference? *J. Autism Dev. Disord.* 43, 1157–1170. doi: 10.1007/s10803-012-1657-y

De Villiers, J. G., Cahillane, J., and Altreuter, E. (2006). "What can production reveal about Principle B," in *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition-North America*, ed. K. U. Deen, J. Nomura, B. Schulz, and B. D. Schwartz (University of Connecticut Occasional Papers in Linguistics: Storrs, CT), 89–100.

de Vries, M., and Geurts, H. M. (2014). Beyond individual differences: are working memory and inhibition informative specifiers within ASD? *J. Neural Transm.* 121, 1183–1198. doi: 10.1007/s00702-014-1225-z

Delage, H., and Frauenfelder, U. H. (2019). Syntax and working memory in typically-developing children: focus on syntactic complexity. *Lang. Interact. Acquis.* 10, 141-176.

Dunn, L. M., and Dunn, L. M. (1997). *Examiner's manual for the PPVT-III: Peabody Picture Vocabulary Test*. 3rd Edn, Circle Pines, MN: American Guidance Service.

Durrleman, S., Delage, H., Prévost, P., and Tuller, L. (2017). The comprehension of passives in autism spectrum disorder. *Glossa* 2, 1–30. doi: 10.5334/gjgl.205

Eigsti, I-M., Bennetto, L., and Dadlani, M. B. (2007). Beyond pragmatics: morphosyntactic development in autism. *J. Autism. Dev. Disord.* 37, 1007–1023. doi: 10.1007/s10803-006-0239-2

Elbourne, P. (2005). On the acquisition of Principle B. *Linguist. Inq.* 36, 333–365.

Geurts, H. M., and Embrechts, M. (2008). Language profiles in ASD, SLI, and ADHD. *J. Autism. Dev. Disord.* 38, 1931–1943. doi: 10.1007/s10803-008-0587-1

Geurts, H. M., Verte, S., Oosterlaan, J., Roeyers, H., and Sergeant, J. A. (2004b). How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *J. Child. Psychol. Psychiatry* 45, 836–854. doi: 10.1111/j.1469-7610.2004.00276.x

Geurts, H. M., Verte, S., Oosterlaan, J., Roeyers, H., Hartman, C. A., Mulder, E. J., et al. (2004a). Can the Children's Communication Checklist differentiate between children with autism, children with ADHD, and normal controls? *J. Child Psychol. Psychiatry* 45, 1437–1453. doi: 10.1111/j.1469-7610.2004.00326.x

Green, B. C., Johnson, K. A., and Bretherton, L. (2014). Pragmatic language difficulties in children with hyperactivity and attention problems: an integrated review. *Int. J. Lang. Commun. Disord.* 49, 15–29. doi: 10.1111/1460-6984.12056

Grimshaw, J., and Rosen, S. T. (1990). Knowledge and obedience: the developmental status of the Binding Theory. *Linguist Inq.* 21, 187–222.

Grodzinsky, Y., and Reinhart, T. (1993). The innateness of binding and coreference. *Linguist Inq.* 24, 69–101.

Habib, A., Harris, L., Pollick, F., and Melville, C. (2019). A meta-analysis of working memory in individuals with autism spectrum disorders. *PLoS One* 14:e0216198. doi: 10.1371/journal.pone.0216198

Happé, F. G. E. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord.* 24, 129–154. doi: 10.1007/bf02172093

Happé, F., Booth, R., Charlton, R., and Hughes, C. (2006). Executive function deficits in autism spectrum disorders and attention–deficit/hyperactivity disorder: examining profiles across domains and ages. *Brain Cogn.* 61, 25–39. doi: 10.1016/j.bandc.2006.03.004

Heck, R. H., Thomas, S. L., and Tabata, L. N. (2012). *Multilevel Modeling of Categorical Outcomes using IBM SPSS*. New York, NY: Routledge.

Helland, W. A., Biringer, E., Helland, T., and Heimann, M. (2012). Exploring language profiles for children with ADHD and children with asperger syndrome. *J. Atten. Disord.* 16, 34–43. doi: 10.1177/1087054710378233

Hendriks, P. (2014). *Asymmetries Between Language Production and Comprehension*. Studies in Theoretical Psycholinguistics, Vol. 42. Dordrecht: Springer.

Hendriks, P., and Spenader, J. (2006). When production precedes comprehension: an optimization approach to the acquisition of pronouns. *Lang. Acquis.* 13, 319–348. doi: 10.1207/s15327817la1304_3

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Dev. Rev.* 24, 189–233. doi: 10.1016/j.dr.2004.01.001

Hollebrandse, B., van Hout, A., and Hendriks, P. (2014). Children's first and second-order false-belief reasoning in a verbal and a low-verbal task. *Synthese* 191, 321–333. doi: 10.1007/s11229-012-0169-9

Ickowicz, A., Schachar, R. J., Sugarman, R., Chen, S. X., Millette, C., and Cook, L. (2006). The parent interview for child symptoms: a situation-specific clinical research interview for attention-deficit hyperactivity and related disorders. *Can. J. Psychiatry* 51, 325–328. doi: 10.1177/070674370605100508

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007

Johnson, M. H., Gliga, T., Jones, E., and Charman, T. (2015). Annual research review: infant development, autism, and ADHD – early pathways to emerging disorders. *J. Child. Psychol. Psychiatry* 56, 228–247. doi: 10.1111/jcpp.12328

Kim, O. H., and Kaiser, A. P. (2000). Language characteristics of children with ADHD. *Commun. Disord. Q.* 21, 154–165. doi: 10.1177/152574010002100304

Kindlon, D., Mezzacappa, E., and Earls, F. (1995). Psychometric properties of impulsivity measures: temporal stability, validity and factor structure. *J. Child Psychol. Psyc.* 36, 645–661. doi: 10.1111/j.1469-7610.1995.tb02319.x

Kissine, M. (2012). Pragmatics, cognitive flexibility and autism spectrum disorders. *Mind. Lang.* 27, 1–28. doi: 10.1111/j.1468-0017.2011.01433.x

Kjelgaard, M. M., and Tager-Flusberg, H. (2001). An investigation of language impairment in autism: implications for genetic subgroups. *Lang. Cogn. Process.* 16, 287–308. doi: 10.1080/01690960042000058

Kort, W., Compaan, E. L., Bleichrodt, N., Resing, W. C. M., Schittekatte, M., Bosmans, M., et al. (2002). *WISC-III NL Handleiding*. London: The Psychological Corporation.

Kuijper, S. J. M. (2016). *Communication Abilities of Children with ASD and ADHD: Production, Comprehension, and Cognitive Mechanisms*. Ph.D. Dissertation. Groningen: University of Groningen.

Kuijper, S. J. M., Hartman, C. A., and Hendriks, P. (2015). Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns. *PLoS One* 10:e0132408. doi: 10.1371/journal.pone.0132408

Kuijper, S. J. M., Hartman, C. A., Bogaerds-Hazenberg, S. T. M., and Hendriks, P. (2017). Narrative production in children with autism spectrum disorder (ASD) and children with attention-deficit/hyperactivity disorder (ADHD): similarities and differences. *J. Abnorm. Psychol.* 126, 63–75. doi: 10.1037/abn0000231

Ladányi, E., Kas, B., and Lukács, Á (2017). The role of cognitive control in anaphor resolution in children with specific language impairment. *Appl. Psycholinguist* 38, 1173–1199. doi: 10.1017/s0142716417000091

Legendre, G., Putnam, M. T., de Swart, H., and Zaroukian, E. (eds) (2016). *Optimality-Theoretic Syntax, Semantics and Pragmatics: From Uni- to Bidirectional Optimization*. Oxford: Oxford University Press.

Levinson, S. (1991). Pragmatic reduction of the binding conditions revisited. *J. Linguist.* 27, 107–161. doi: 10.1017/s0022226700012433

Levinson, S. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT press.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

Lord, C., Rutter, M., DiLavore, P. C., and Risi, S. (1999). *Autism Diagnostic Observation Schedule (ADOS): Manual*. Los Angeles, CA: Western Psychological Services.

Martinussen, R., Hayden, J., Hogg-Johnson, S., and Tannock, R. (2005). A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* 44, 377–384.

Mary, A., Slama, H., Mousty, P., Massat, I., Capiau, T., Drabs, V., et al. (2016). Executive and attentional contributions to Theory of Mind deficit in attention deficit/hyperactivity disorder (ADHD). *Child. Neuropsychol.* 22, 345–365. doi: 10.1080/09297049.2015.1012491

Matthews, D., Lieven, E., Theakston, A., and Tomasello, M. (2009). Pronoun co-referencing errors: challenges for generativist and usage-based accounts. *Cogn. Linguist.* 20, 599–626.

Milligan, K., Astington, J. W., and Dack, L. A. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Dev.* 78, 622–646. doi: 10.1111/j.1467-8624.2007.01018.x

Montgomery, J. W., and Evans, J. L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *J. Speech Lang. Hear R.* 52, 269–288. doi: 10.1044/1092-4388(2008/07-0116)

Nicol, J., and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *J. Psycholinguist. Res.* 18, 5–19. doi: 10.1007/bf01069043

Nijmeijer, J. S., Arias-Vásquez, A., Rommelse, N. N. J., Altink, M. E., Anney, R. J. L., Asherson, P., et al. (2010). Identifying loci for the overlap between attention-deficit/hyperactivity disorder and autism spectrum disorder using a genome-wide QTL linkage approach. *J. Am. Acad. Child Adolesc. Psychiatry* 49, 675–685. doi: 10.1016/j.jaac.2010.03.015

Nydén, A., Billstedt, E., Hjelmquist, E., and Gillberg, C. (2001). Neurocognitive stability in Asperger syndrome, ADHD, and reading and writing disorder: a pilot study. *Dev. Med. Child. Neurol.* 43, 165–171. doi: 10.1111/j.1469-8749.2001.tb00182.x

Oram, J., Fine, J., Okamoto, C., and Tannock, R. (1999). Assessing the language of children with attention deficit hyperactivity disorder. *Am. J. Speech Lang. Pathol.* 8, 72–80.

Owen, A. M., Mcmillan, K. M., Laird, A. R., and Bullmore, E. (2005). N-Back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Map.* 25, 46–59. doi: 10.1002/hbm.20131

Papaeliou, C. F., Maniadaki, K., and Kakouros, E. (2015). Association between story recall and other language abilities in schoolchildren with ADHD. *J. Atten. Disord.* 19, 53–62. doi: 10.1177/1087054712446812

Pennington, B. F., and Ozonoff, S. (1996). Executive functions and developmental psychopathology. *J. Child Psychol. Psyc.* 37, 51–87. doi: 10.1111/j.1469-7610.1996.tb01380.x

Perovic, A., and Wexler, K. (2018). Teasing apart explanations of a developmental delay in binding: experimental evidence from the comparison of SLI and Williams Syndrome. *Lang. Acquis.* 25, 24–38. doi: 10.1080/10489223.2016.1187618

Perovic, A., Modyanova, N., and Wexler, K. (2013). Comparison of grammar in neurodevelopmental disorders: the case of binding in williams syndrome and autism with and without language impairment. *Lang. Acquis.* 20, 133–154. doi: 10.1080/10489223.2013.766742

Philip, W., and Coopmans, P. (1996). "The double dutch delay of principle b effect," in *Proceedings of the 20th Boston University Conference on Language Development*, eds A. Stringfellow, D. Cahana-Amitay, E. Hughes, and A. Zukowski (Somerville, MA: Cascadilla Press), 576–587.

Prévost, P., Tuller, L., Zebib, R., Barthez, M. A., Malvy, J., and Bonnet-Brilhault, F. (2018). Pragmatic versus structural difficulties in the production of pronominal clitics in French-speaking children with autism spectrum disorder. *Autism Dev. Lang. Impair.* 3:2396941518799643.

Prince, A., and Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell.

Reinhart, T. (2006). *Interface Strategies: Optimal and Costly Computations*. Cambridge, MA: MIT Press.

Reinhart, T. (2011). "Processing or pragmatics? explaining the coreference delay," in *The Processing and Acquisition of Reference*, eds E. Gibson and N. J. Pearlmutter (Cambridge, MA: MIT Press), 157–194. doi: 10.7551/mitpress/9780262015127.003.0007

Reinhart, T., and Reuland, E. (1993). Reflexivity. *Linguist. Inq.* 24, 657–720.

Reuland, E. (2011). *Anaphora and Language Design*. Cambridge, MA: MIT Press.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., et al. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J. Am. Acad. Child Adolesc. Psychiatry* 45, 1094–1103. doi: 10.1097/01.chi.0000227880.42780.0e

Rommelse, N. N. J., Geurts, H. M., Franke, B., Buitelaar, J. K., and Hartman, C. A. (2011). A review on cognitive and brain endophenotypes that may be common in autism spectrum disorder and attention-deficit/hyperactivity disorder and facilitate the search for pleiotropic genes. *Neurosci. Biobehav. Rev.* 35, 1363–1396. doi: 10.1016/j.neubiorev.2011.02.015

Ruigendijk, E. (2008). "Reference assignment in German preschool children," in *Language Acquisition and Development: Proceedings of Gala 2007*, eds A. Gavarró and M. J. Freitas (Cambridge: Cambridge University Press), 370–380.

Ruigendijk, E., Baauw, S., Zuckerman, S., Vasic, N., de Lange, J., and Avrutin, S. (2011). in *A cross-Linguistic Study on the Interpretation of Pronouns by Children and Agrammatic Speakers: Evidence from Dutch, Spanish and Italian*, eds E. Gibson and N. J. Pearlmutter (Cambridge MA: MIT Press), 133–155.

Ruigendijk, E., Friedmann, N., Novogrodsky, R., and Balaban, N. (2010). Symmetry in comprehension and production of pronouns: a comparison of German and Hebrew. *Lingua* 120, 1991–2005. doi: 10.1016/j.lingua.2010.02.009

Rutter, M., Le Couteur, A., and Lord, C. (2003). *Autism Diagnostic Interview Revised (ADIR): Manual*. Los Angeles, CA: Western Psychological Services.

Schaeffer, J. (2003). "Pragmatics and SLI," in *Language Competence Across Populations: Towards a Definition of Specific Language Impairment*, eds Y. Levy and J. Schaeffer (Hillsdale NJ: Lawrence Erlbaum Associates), 1–29.

Schleepen, T. M., and Jonkman, L. M. (2010). The development of non-spatial working memory capacity during childhood and adolescence and the role of interference control: an N-Back task study. *Dev. Neuropsychol.* 35, 37–56. doi: 10.1080/87565640903325733

Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Amsterdam: Harcourt test Publishers.

Schmitt, L. M., White, S. P., Cook, E. H., Sweeney, J. A., and Mosconi, M. W. (2018). Cognitive mechanisms of inhibitory control deficits in autism spectrum disorder. *J. Child Psychol. Psyc.* 59, 586–595. doi: 10.1111/jcpp.12837

Slabakova, R., White, L., and Guzzo, N. B. (2017). Pronoun interpretation in the second language: effects of computational complexity. *Front. Psychol.* 8:1236. doi: 10.3389/fpsyg.2017.01236

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguist. Inq.* 27, 720–731.

Spenader, J., Smits, E.-J., and Hendriks, P. (2009). Coherent discourse solves the pronoun interpretation problem. *J. Child Lang.* 36, 23–52. doi: 10.1017/s0305000908008854

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *J. Autism Dev. Disord.* 11, 45–56. doi: 10.1007/bf01531340

Tannock, R., Hum, M., Masellis, M., Humphries, T., and Schachar, R. (2002). *Teacher Telephone Interview for Attention-Deficit/Hyperativity Disorder and Related Disorders: DSM-IV version*. Toronto, ON: The hospital for sick children.

Tannock, R., Schachar, R. J., Carr, R. P., Chajczyk, D., and Logan, G. D. (1989). Effects of methylphenidate on inhibitory control in hyperactive children. *J. Abnorm. Child Psych.* 17, 473–491. doi: 10.1007/bf00916508

Terzi, A., Marinis, T., Kotsopoulou, A., and Francis, K. (2014). Grammatical abilities of Greek-speaking children with autism. *Lang. Acquis.* 21, 4–44. doi: 10.1080/10489223.2013.855216

Thornton, R., and Wexler, K. (1999). *Principle B, VP Ellipsis and Interpretation in Child Grammar*. Cambridge, MA: MIT Press.

Tuller, L., Ferré, S., Prévost, P., Barthez, M.-A., Malvy, J., and Bonnet-Brilhault, F. (2017). "The effect of computational complexity on the acquisition of French by children with ASD," in *Innovative Investigations of Language in Autism Spectrum Disorder*, ed. L. Naigle (Berlin: de Gruyter), 115–140. doi: 10.1037/15964-007

van den Wildenberg, W. P. M., and Christoffels, I. K. (2010). STOP TALKING! Inhibition of speech is affected by word frequency and dysfunctional impulsivity. *Front. Psychol.* 1:145. doi: 10.3389/fpsyg.2010.00145

van Rij, J., Hollebrandse, B., and Hendriks, P. (2016). "Children's eye gaze reveals their use of discourse context in object pronoun resolution," in *Empirical Perspectives on Anaphora Resolution*, eds A. Holler and K. Suckow (Berlin: De Gruyter), 267–293.

van Rij, J., van Rijn, H., and Hendriks, P. (2010). Cognitive architectures and language acquisition: a case study in pronoun comprehension. *J. Child Lang.* 37, 731–766. doi: 10.1017/s0305000909990560

Vogelzang, M., Guasti, M. T., van Rijn, H., and Hendriks, P. (2021). How children process reduced forms: a computational cognitive modeling approach to pronoun processing in discourse. *Cogn. Sci.* 45:e12951.

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child. Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304

Williams, D. L., Goldstein, G., Carpenter, P. A., and Minshew, N. J. (2005). Verbal and spatial working memory in autism. *J. Autism. Dev. Disord.* 35, 747–756. doi: 10.1007/s10803-005-0021-x

Wolfgramm, C., Suter, N., and Göksel, E. (2016). Examining the role of concentration, vocabulary and self-concept in listening and reading comprehension. *Int. J. List.* 30, 25–46. doi: 10.1080/10904018.2015.1065746

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

frontiers in Communication

ORIGINAL RESEARCH
published: 15 June 2021
doi: 10.3389/fcomm.2021.625757

On the Interaction of Gestural and Linguistic Perspective Taking

Stefan Hinterwimmer[1]*, *Umesh Patil*[2] *and Cornelia Ebert*[3]

[1]Department of German Studies, Bergische Universität Wuppertal, Wuppertal, Germany, [2]Department of German Language and Literature I, Universität zu Köln, Cologne, Germany, [3]Department of Linguistics, Goethe-Universität Frankfurt, Frankfurt, Germany

In this paper, we investigate the question of whether and how perspective taking at the linguistic level interacts with perspective taking at the level of co-speech gestures. In an experimental rating study, we compared test items clearly expressing the perspective of an individual participating in the event described by the sentence with test items which clearly express the speaker's or narrator's perspective. Each test item was videotaped in two different versions: In one version, the speaker performed a co-speech gesture in which she enacted the event described by the sentence from a participant's point of view (i.e. with a character viewpoint gesture). In the other version, she performed a co-speech gesture depicting the event described by the sentence as if it was observed from a distance (i.e. with an observer viewpoint gesture). Both versions of each test item were shown to participants who then had to decide which of the two versions they find more natural. Based on the experimental results we argue that there is no general need for perspective taking on the linguistic level to be aligned with perspective taking on the gestural level. Rather, there is clear preference for the more informative gesture.

Keywords: perspective taking, free indirect discourse, viewpoint, co-speech gestures, observer viewpoint gestures, character viewpoint gestures, anti-logophoricity

OPEN ACCESS

Edited by:
Jorrig Vogels,
University of Groningen, Netherlands

Reviewed by:
Marisa Casillas,
Max Planck Institute for Psycholinguistics, Netherlands
Gisela Redeker,
University of Groningen, Netherlands
Natalie Dowling,
University of Chicago, United States,
in collaboration with reviewer MC

***Correspondence:**
Stefan Hinterwimmer
hinterwimmer@uni-wuppertal.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *03 November 2020*
Accepted: *24 May 2021*
Published: *15 June 2021*

Citation:
Hinterwimmer S, Patil U and Ebert C (2021) On the Interaction of Gestural and Linguistic Perspective Taking. Front. Commun. 6:625757. doi: 10.3389/fcomm.2021.625757

INTRODUCTION

Perspective taking is an integral part of the information conveyed by sentences that are contained in narrative texts. One and the same event can be described from a detached observer's perspective or as if it was perceived by a participant. Free Indirect Discourse (FID) is a particularly clear way for a sentence to express the perspective of a protagonist in a narrative text. In FID, all perspective-dependent expressions (i.e. deictic expressions such as tomorrow and here, evaluative expressions, interjections etc.) are interpreted from the perspective of some contextually salient protagonist, while pronouns and tenses are interpreted from the (possibly entirely abstract) narrator's perspective (Rauh 1978; Banfield 1982; Doron 1991; Schlenker 2004; Eckardt 2014; Maier 2015). Additionally, the proposition denoted by a sentence in FID is interpreted as a thought or utterance that the respective protagonist has or makes at the reference time of the ongoing story (Eckardt 2014). The second sentence in (1), for instance, which is a clear instance of FID, is thus interpreted as a thought that Masha has at the time of her staring at Wilfred–the same thought as the one rendered as direct discourse in (1b).

(1) a. Masha stared at Wilfred in disbelief. Had that idiot really invited her to his birthday party tomorrow evening?

b. Masha stared at Wilfred in disbelief. She thought: 'Has that idiot really invited me to his birthday party tomorrow evening?'

At the same time, it is well-known that perspective taking can also be expressed at the level of co-speech gestures, i.e. gestures that speakers produce while uttering sentences. In particular, there are two kinds of iconic gestures that are often used by speakers when they describe scenes or events to their interlocuters and that clearly reveal a perspective: character viewpoint gestures (CVG), on the one hand, and observer viewpoint gestures (OVG), on the other (McNeill, 1992; Parrill, 2010, Parrill, 2012; Stec, 2012, Stec, 2016). When performing the former, the speaker impersonates an individual participating in the event described by the sentence and enacts the event from that person's point of view by using her entire body in combination with facial expressions. When performing the latter, in contrast, the speaker depicts the event described by the sentence as if it was observed from a distance, usually by using the hands exclusively which then represent a participant, with the hand's trajectory representing that participant's path, for instance.

Based on Parrill (2010) assumption that gestural and linguistic viewpoint have a common conceptual source, we investigate the question in this paper of whether and how perspective taking at the linguistic level interacts with perspective taking at the level of co-speech gestures. In an experimental study conducted in German we compared test items clearly expressing the perspective of an individual participating in the events described with test items which clearly express the speaker's or narrator's perspective. Test items of the former kind were always construed in such a way that the opening sentence describes the feelings, intentions or thoughts of a protagonist in a particular situation which the second sentence specifies in more detail, while the final sentence renders a thought that the respective protagonist has in that situation in the form of FID. Test items of the latter kind, in contrast, always contained an opening sentence in the form of a general statement in present tense expressing an evaluation of or opinion about the individual participating in the event described by the following sentences. Additionally, that individual was referred to by a demonstrative pronoun. As shown by (Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017), German demonstrative pronouns are anti-logophoric pronouns, i.e. they cannot refer to an individual whose perspective is expressed by the sentence containing them.

Each test item was videotaped in two different versions: In one version, the speaker performed a CVG while uttering the respective final sentence. In the second version, the speaker performed an OVG. Both versions of each test item were shown to participants who then had to decide which of the two versions they find more natural. If Parrill (2010) is right in assuming that gestural and linguistic viewpoint have a common conceptual source, it is to be expected that those versions of the test items are preferred where the linguistically expressed perspective aligns with the perspective expressed on the gestural level. This expectation is based on the assumption that combinations of speech and gesture convey a complex multimodal message including a perspective that is planned by a central cognitive process and then dispatched into disparate channels (see De Ruiter, 1998, De Ruiter, 2000, De Ruiter, 2007; Kendon, 2004). Crucially, while the information conveyed via the two channels may be either redundant (thus avoiding misunderstanding) or complementary, the default should be for it to be coherent (see also Cassell et al., 1999, but see Goldin-Meadow, 1999 for arguments that mismatches between speech and gestures are sometimes productive and useful). Concerning perspective, this means that at least in the absence of intervening factors there should be a preference for perspective alignment. Consequently, the prediction is that participants choose the CVG-version of test items instantiating FID and thus expressing the participant's perspective more often than the OVG-version. For test items expressing the speaker's or narrator's perspective, in contrast, it should be the other way around, i.e. participants should choose the OVG-version more often than the CVG-version.

An exception to this general reasoning would be the case of multiple perspectives. A speaker might plan to express more than one perspective at the same time. In this vein, Parrill (2009) investigates dual viewpoint gestures (first noted by McNeill, 1992, see also Cassell et al., 1999 for discussion), i.e. gestures that simultaneously express more than one viewpoint. If gestures can express more than one viewpoint at the same time, it should also be possible to express more than one viewpoint in gesture and speech. For example, one could express a certain viewpoint A in speech and a certain viewpoint B in gesture, or, a viewpoint A in speech and a dual viewpoint expressing A and B in gesture. This possibility might have confounded our results and we will come back to it below.

As we will see, the predictions that FID couples with CVGs and linguistic narrator's perspective with OVGs are only partially confirmed by the experimental results: While CVGs were generally preferred in combination with both kinds of test items, this tendency was stronger in test items expressing the respective participant's perspective than in test items expressing the speaker's or narrator's perspective. We take this as evidence that there is no general need for perspective taking on the linguistic level to be aligned with perspective taking on the gestural level, contrary to our expectations. Nevertheless, there seems to be a general preference for more informative gestures, with CVGs being more informative than OVGs insofar as they are less schematic and more detailed. Since a speaker might wish to express dual viewpoints, she might choose to express one viewpoint in speech and another one, possibly one that is more informative, with gesture.

The paper is structured as follows. In *Free Indirect Discourse* and *Demonstrative pronouns as anti-logophoric pronouns*. we provide some theoretical background on FID and the anti-logophoricity of German demonstrative pronouns, respectively, and in *Observer viewpoint and character viewpoint gestures* on CVGs and OVGs. The experimental study is presented and discussed in *The experimental study*. *Conclusion* concludes the paper.

THEORETICAL BACKGROUND

Free Indirect Discourse

As already said in the introduction, FID is a particularly clear form of linguistically encoded perspective taking that is largely confined to narrative texts (but see Redeker, 1996 for evidence that FID is often found in journalistic texts as well and Stokke, 2020 for discussion of the use of FID in non-fictional texts such as biographies of historical figures; see Brinton, 1980, Banfield, 1982, Stokke, 2013, Hinterwimmer, 2017 and Abrusán, 2020 for another form of linguistically encoded perspective taking dubbed free perception report, viewpoint shifting, or protagonist projection). A sentence in FID conveys a character's thoughts or utterances without there being any overt indication of a shift from the (potentially entirely abstract) narrator's external perspective to the internal perspective of the respective character. An overt indication of such a shift would be the presence of a propositional attitude verb, where the external argument slot is occupied by a phrase referring to that character in combination with quotation marks, as in direct discourse (DD). What FID shares with DD is the interpretation of local and temporal deictic expressions with respect to the context of the individual whose thoughts or utterances are rendered. At the same time, pronouns and tenses are interpreted with respect to the narrator's context, as in indirect discourse. Consider again (1a) from above, repeated here as (2a): The second sentence is interpreted as a question that Masha asks herself while she is staring at Wilfred in disbelief. Crucially, the deictic temporal adverb tomorrow, which is usually interpreted with respect to the context of utterance (Kaplan 1989), is interpreted in the same way in (2a) as it is interpreted in (2b), namely as referring to the day following the day on which she has the thought reported by the respective sentence. In contrast to (2b), however, where that thought is rendered as DD, Masha is referred to by the third person pronoun *her* instead of the first person pronoun *me*, and the auxiliary verb *have* is marked for past instead of present tense.

(2) a. Masha stared at Wilfred in disbelief. Had that idiot really invited her to his birthday party tomorrow evening?
b. Masha stared at Wilfred in disbelief. She thought: "Has that idiot really invited me to his birthday party tomorrow evening?"

Deictic expressions thus do not behave uniformly in FID: While the vast majority of them is interpreted with respect to the context set up by the preceding sentence, pronouns and tenses are always interpreted with respect to the narrator's context. Concerning evaluative expressions such as *that idiot* in (2a), interjections such as *wow*, *ouch* and *oops* or exclamatives, they are always interpreted with respect to the respective character's perspective–in (2a), for instance, it is Masha who considers Wilfred an idiot, not the narrator. Likewise, the exclamative in (3) expresses Tom's delight and surprise in virtue of the extent to which he is (or rather believes himself to be) smart, not the narrator's.

(3) Tom leaned back in his chair, smiling at the man he had just cheated out of 5,000 dollars. Wow, how smart he was!

Concerning the question of how the distinctive properties of FID just outlined are to be captured, there are two lines of analysis that have been proposed in the formal semantics literature: double context analyses (Schlenker 2004, Sharvit 2008, Eckardt 2014; see Rauh 1978, Banfield 1982 and Doron 1991 for earlier implementations of similar ideas in different frameworks) and the mixed quotation approach. The basic idea behind double context analyses is that sentences in FID are interpreted not just with respect to a context of utterance (Kaplan 1989), but with respect to two different contexts. Concerning the technical details, the following exposition is based on Eckardt (2014). Eckardt assumes that while ordinary utterances in everyday communication are just interpreted with respect to a context of utterance *C*, narrative texts allow the addition of a second context *c*, with *C* being the context of the (potentially entirely abstract) narrator and *c* being the context of some character that is prominent at that point in the discourse (see Hinterwimmer 2019 and Hinterwimmer and Meuser 2019 for detailed discussion of the conditions under which characters are prominent enough to serve as potential anchors for FID). This second context *c* is implicitly introduced by the preceding discourse and it consists of the character functioning as the author (i.e. the speaker or thinker) of *c* and the spatial and temporal location of that character at the reference time of the ongoing story.

Whenever only *C* is present, all context-sensitive expressions are interpreted with respect to *C*. Crucially, however, all context-sensitive expressions with the exception of pronouns and tenses have to be interpreted with respect to *c* whenever *c* is available, while pronouns and tenses always have to be interpreted with respect to *C*. Additionally, whenever a sentence is interpreted both with respect to *C* and *c*, the proposition *p* (or set of propositions *Q* in cases such as (2a), where the sentence in FID is a polar question) it denotes is not directly added to the set of propositions that characterize the fictional story worlds. Rather, the proposition that the author of *c* believes *p* (or asks herself *Q*) is added, thus ensuring that sentences in FID are interpreted as the respective character's thoughts rather than assertions by the narrator.

Concerning the second sentence in (2a), for instance, the deictic temporal adverb tomorrow is interpreted with respect to the temporal parameter of *c*, i.e. as referring to the day following the day on which Masha stared at Wilfred in disbelief, and the negative evaluation of Wilfred as an idiot is attributed to the author of *c*, i.e. Masha. The past tense marking of the auxiliary verb *have* and the third person features on the pronouns her and him, in contrast, are interpreted with respect to the narrator's context *C*, thus requiring temporal location before the time of *C* rather than *c* and distinctness from the author (and addressee) of *C* rather than *c* (which is why *her* can be interpreted as referring to Masha), respectively. Finally, the sentence is interpreted as a question that Mary is asking herself at the time of *c*, i.e. at the time of her staring at Wilfred in disbelief.

A fundamentally different analysis of FID is proposed by Maier (2015, 2017: see also Dirscherl and Pafel, 2015). On this approach, FID is a special, highly conventionalized form of mixed quotation: Sentences in FID are quotes of thoughts or utterances, with tenses and pronouns being unquoted. In contrast to more familiar forms of mixed quotation, as they are found in newspaper articles, the quoted parts are not typographically marked as such (via quotation marks or italics, for instance) and there is no introductory clause such as *x said/thought* signaling that the following clause is to be interpreted as the partial quote of a thought or utterance of *x*. On the mixed quotation approach (2a) corresponds to the (simplified) schematic representation in (4a) and is interpreted as paraphrased (in simplified form) in (4b), the idea being that speaking and thinking events have both a form and a content and can be decomposed into subevents corresponding to parts of the respective thought or utterance.

(4) a. Mary stared at Wilfred in disbelief. Had 'that idiot really invited' her 'to his birthday party tomorrow evening'? b. There is an event *e* of Mary staring in disbelief that is located before the time of *C* and there is an event e_1 of Masha thinking that is located at the time of *e* and there are subevents e_2 and e_3 of e_1, and the form of e_1 is the form of e_2 concatenated with *that idiot really invited* concatenated with the form of e_3 concatenated with *to his birthday party tomorrow evening*, and the content of e_2 is the denotation of had and the content of e_3 is the denotation of her.

Concerning the question of why pronouns and tenses are systematically unquoted in FID (Maier, 2015, Maier, 2017), assumes this to be the result of a pragmatic tendency that can be observed in other forms of mixed quotation as well and has become fully conventionalized in the case of FID. Note that in spite of the profound differences in technical implementation, the resulting interpretations are rather similar to those assumed by the double context analyses: A speaking or thinking event has to be accommodated by the reader and the content of this event is the semantic object denoted by the respective sentence in FID. The quoted context-sensitive expressions are ultimately interpreted with respect to the context of the character whose thought or utterance is being quoted, while the unquoted ones receive their standard interpretation with respect to the narrator's context. Nevertheless, there is an argument in favor of the mixed quotation approach: As pointed out by Maier (2015), there are cases of FID where a character's thoughts or utterances are rendered in the non-standard dialect spoken by that character, while the surrounding text is written in standard language. While such cases are easily accounted for in the mixed quotation approach, it is hard to see how they could be captured by double context analyses.

Having discussed FID as a phenomenon where the perspective of some character becomes highly prominent (without the narrator's perspective disappearing completely, as evidenced by the interpretation of pronouns and tenses), we will discuss the use of German demonstrative pronouns in the following section and argue that these pronouns indicate that the speaker's or narrator's perspective is highly prominent.

Demonstrative Pronouns as Anti-logophoric Pronouns

Similarly to languages such as Dutch, Finnish and Catalan (see, e.g. Kaiser and Trueswell, 2008; Kaiser, 2010, Kaiser, 2011a, Kaiser, 2011b, Kaiser, 2013; Mayol and Clark, 2010), German does not only have personal pronouns (henceforth: PPros), but also demonstrative pronouns. The latter come in two varieties: the *der/die/das* series and the *dieser/diese/dieses* series. Since *diese* pronouns are largely confined to the formal register (see Patil et al., 2020 for recent discussion), we will set them aside for the purposes of this paper and concentrate on the contrast between PPros and demonstrative pronouns of the *der/die/das* variety, which we will henceforth refer to as DPros.

In the past, research on the contrast between DPros and PPros has mostly focused on cases like (5) (adapted from Bosch et al., 2007), where two potential antecedents with congruent gender features have been introduced in the preceding linguistic context and where there is genuine ambiguity in the resolution options of the pronouns. In such cases, there is a strong tendency for DPros to pick up the less prominent and for PPros to pick up the more prominent antecedent. Prominence has been defined in terms of (grammatical) subjecthood (Bosch et al., 2007; Hinterwimmer and Brocher 2018), topicality (Bosch and Umbach 2006; Hinterwimmer 2015), and (proto-)agentivity (Schumacher et al., 2016, Schumacher et al., 2017).

(5) Paul_i wollte mit Peter_j laufen gehen. Aber $\{\text{er}_{i,j}/\text{der}_j\}$ war leider erkältet.

Paul_i wanted to go running with Peter_j. But he $\{\text{PPro}_{i,j}/\text{DPro}_j\}$ had a cold unfortunately *(adapted from Bosch et al., 2007).*

In (5), Paul is more prominent than Peter since the proper name referring to him is the grammatical subject of the preceding sentence and the agent of the verb contained in that sentence, while the proper name referring to Paul is (contained in) the prepositional object and the theme. Additionally, Peter is also the aboutness topic (Reinhart, 1981) of that sentence by default, because the proper name referring to him is the subject of the sentence and because it occupies the leftmost position of the sentence. Consequently, there is a very strong tendency for the DPro *der* in (5) to pick up Peter, while the PPro *er* can be understood as picking up either Paul or Peter, with a (comparatively weak) preference for Paul.

As observed by Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017; see Hinterwimmer et al., 2020 and Hinterwimmer to appear for additional empirical evidence), however, there are cases like (6a) where referents that are maximally prominent in terms of subjecthood, topicality and (proto-)agentivity can nevertheless easily be picked up by DPros. At the same time, the contrast with (6b), where this is impossible or at least leads to rather strong markedness, shows that it is not the case that DPros are only prohibited from picking up maximally prominent referents in cases of potential ambiguity, i.e. whenever there is more than one potential antecedent with matching gender features available.

(6) Peter$_i$ seufzte, als er die Tür öffnete, und sah, dass die Wohnung mal wieder in einem fürchterlichen Zustand war.
Peter$_i$ sighed when he opened the door and saw that the flat was in a terrible state again.
a. Der $_i$ kann sich einfach nicht gegen seinen Mitbewohner durchsetzen.
He(DPro$_i$) is simply unable to stand his ground against his flatmate.
b. Verdammt, der$_{*i}$/er$_i$ hatte doch gestern erst aufgeräumt.
Damn, he(DPro$_{*i}$)/he$_i$ had only tidied up yesterday, after all. (Hinterwimmer et al., 2020: 114, ex. (8))

Peter is not only the subject and the agent of the matrix as well as the temporal adjunct clause in the opening sentence in (6), but its referent is presumably also the aboutness topic (Reinhart 1981) of the sentence, i.e. the proposition denoted by that sentence is understood as being about Peter rather than the door or the flat. If DPros avoid maximally prominent referents, Peter should thus be unavailable as an antecedent not only for the DPro in (6b), but also for the one in (6a). At the same time, if DPros were only prohibited from picking up maximally prominent referents in cases of potential ambiguity, Peter should not only be available as an antecedent for the DPro in (6a), but also in (6b). So, what distinguishes the continuation of (6) in (6a) from the one in (6b)?.

According to Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017, it is the fact that (6a) can only be interpreted as expressing the narrator's (or speaker's, if it is uttered in oral conversation) perspective, while (6b) is most plausibly interpreted as expressing Peter's perspective. In other words, while the narrator is the perspectival center with respect to the proposition denoted by (6a), Peter is the perspectival center with respect to the proposition denoted by (6b). In the case of (6a), it is the switch from past tense in the opening sentence, which is an instance of neutral narration, to present tense in the continuation, which breaks narrative continuity and in combination with the content establishes the narrator (or speaker) as the perspectival center. Consequently, (6a) is understood as a general statement about Peter's character by the narrator.

The continuation in (6b), in contrast, is most likely understood as a thought of Peter rendered in FID, i.e. the most plausible reading is one according to which Peter thought I only tidied up yesterday, after all when he saw the chaos in the kitchen. This is indicated by the content in combination with the presence of the deictic temporal adverb *gestern* ("yesterday"), the evaluative expression *verdammt* ("damn") and the modal particle *doch*. Concerning the deictic temporal adverb *gestern*, it is most plausibly interpreted as referring to the day preceding the day on which Peter came home in the evening, i.e. with respect to Peter's context, not with respect to the narrator's (or speaker's) context. Likewise, the evaluative expression *verdammt* ("damn") is more plausibly interpreted as expressing Peter's rather than the narrator's (or speaker's) frustration. Finally, the modal particle *doch*, which (very roughly) indicates that the proposition denoted by the clause containing it violates a previously held assumption, is more plausibly interpreted as violating Peter's rather than the narrator's (or speaker's) expectations.

From contrasts like the one between (6a) and (6b) in the context of (6), Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017 draw the conclusion that DPros avoid discourse referents functioning as perspectival centers as antecedents, where the term perspectival center is defined as in (7).

(7) A discourse referent α is the perspectival center with respect to a proposition *p* if *p* is the content of a mental state of the semantic value of α (i.e. g(*α*), where g is the assignment function).

Since the proposition denoted by (6b) (on its most plausible interpretation) is the content of a thought of Peter, Peter is the perspectival center with respect to that proposition and can accordingly not be picked up by a DPro, but only by a PPro (PPros being neutral in this respect, i.e. they can, but do not have to pick up discourse referents functioning as perspectival centers). Concerning the proposition denoted by (6a), in contrast, there is a strong tendency for it to be interpreted as the content of a thought of the narrator (or speaker). Consequently, the narrator (or speaker) is the perspectival center with respect to that proposition, and the DPro can accordingly be interpreted as picking up Peter, in spite of him being maximally prominent in terms of subjecthood (proto-)agentivity, and topicality. Because of their avoidance of perspectival centers, DPros are dubbed anti-logophoric pronouns in Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017 (see Charnavel and Mateu, 2015 and Yashima, 2015 for arguments that anti-logophoric pronouns exist in French, Spanish and Japanese, as well), i.e. the counterparts of pronouns existing in many West African and Asian languages that can only be used to pick up discourse referents functioning as perspectival centers and have been dubbed logophoric pronouns (Clements, 1975; Sells, 1987; Sundaresan, 2012; Nishigauchi, 2014; Pearson, 2015).

But what about cases like (5), where anti-logophoricity does not obviously play a role regarding the resolution options of DPros? Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017 argue that in the absence of the speaker or narrator functioning as perspectival center (i.e. in instances of neutral narration) the respective aboutness topic is the perspectival center by default, i.e. the proposition denoted by the respective sentence is interpreted as the content of a mental state of the topical referent, where that mental state need not be a conscious thought but can also be a state of perceiving. Evidence for this assumption is provided by the following observation (see Hinterwimmer et al., 2020 for additional empirical evidence): In the variant of (5) (repeated here as (8a)) given in (8b), where the second sentence is construed in such a way that it clearly expresses an evaluative comment by the speaker or narrator, the DPro can easily be understood as picking up Paul.

(8) a. Paul$_i$ wollte mit Peter$_j$ laufen gehen. Aber {er$_{i,j}$/der$_j$} war leider erkältet.
Paul$_i$ wanted to go running with Peter$_j$. But he {PPro$_{i,j}$/DPro$_j$} had a cold unfortunately.
b. Paul$_i$ wollte mit Peter$_j$ laufen gehen. {Er$_{i,j}$/Der$_{i,j}$} sucht sich immer Leute als Trainingspartner aus, die nicht richtig fit sind.

Paul$_i$ wanted to go running with Peter$_j$. He {PPro$_{i,j}$/DPro$_{i,j}$} always picks people as training partners who are not really fit.

In the experiment to be discussed in *The experimental study*, we make use of the anti-logophoricity of DPros as indicators that the sentences containing them have to be interpreted as expressing the speaker's or narrator's perspective (see Zeman, 2019, to appear for discussion of other indicators of the narrator's presence as a perspective taker in narrative texts) rather than the perspective of the discourse referent picked up by the respective DPro. Accordingly, the items which are meant to express the speaker's or narrator's perspective are construed in such a way that they license the use of DPros to pick up the respective topical referent, with the presence of the DPro enforcing such an interpretation. The items which are meant to express the respective topical referent's perspective, in contrast, always contain an instance of FID that can only be interpreted as a thought of that protagonist.

Having discussed perspective taking on the linguistic level, we will discuss perspective taking on the level of co-speech gestures in the following section.



FIGURE 1 | Original scene

Observer Viewpoint and Character Viewpoint Gestures

As already mentioned in the introduction, it is well-known that perspective taking can also be expressed at the level of co-speech gestures, i.e. via gestures that speakers produce while uttering sentences. While there are gestures that are produced without speech and either replace certain parts of speech (pro-speech), precede (pre-speech) or follow it (post-speech), most gestures are produced during speech (co-speech) (see McNeill, 1992 for a descriptive approach of the different types of gestures and Schlenker, 2018 for recent discussion in the formal semantic realm). Empirical studies have shown that a co-speech gesture and the corresponding spoken language segment are not only semantically, but also temporally aligned in systematic ways. Usually the apex (or more generally: the "stroke") of a gesture coincides with or directly precedes an intonational peak, the main accent, of the semantically associated phrase (Pittenger et al., 1960; Kita and Özyürek, 2003; Loehr, 2004). Typically, the content of a speech-accompanying gesture semantically interacts closely with the corresponding phrase and triggers a complex meaning ensemble that is dependent on the utterance context (Kopp et al., 2004). In the following, we will only be concerned with a certain type of gestures among many very different gesture types, namely iconic gestures. Iconic gestures depict some aspect of what they are meant to represent. For example, a "round"-gesture indicating the shape of some round item depicts roundness and thus always bears a certain similarity to roundness or a round object. At the same time, iconic gestures are more or less idiosyncratic and dependent on the person that performs them. While one speaker chooses to illustrate the roundness of a certain object by way of a static two-handed gesture representing roundness, another might use a dynamic gesture drawing a circle in the air with the index finger. While iconic gestures are non-conventionalized and, as their name says, very iconic, these are not properties that are shared by gestures in general. Emblematic gestures, for example, like the "thumbs-up" or the "victory" sign are symbols that have to be performed according to the conventions of a certain cultural community and they do not necessarily bear similarity to what they represent. An emblem can have a certain meaning in one community and a very different one or none in another. In the following, we will only be concerned with iconic gestures.

Crucially, it has been argued that there are two kinds of iconic gestures which are often used by speakers when they describe scenes or events to their interlocuters and that clearly reveal a perspective: character viewpoint gestures (CVG), on the one hand, and observer viewpoint gestures (OVG) gestures, on the other (McNeill, 1992; Parrill, 2010, Parrill, 2012; Stec, 2012, Stec, 2016). When producing a CVG, the speaker impersonates an individual participating in the event described by the sentence and enacts the event from that person's point of view. When producing an OVG, in contrast, the speaker depicts the event described by the sentence as if it was observed from a distance. CVGs typically involve the speaker's entire body and face, while OVGs are usually performed exclusively with the hands (McNeill 1992).

As an illustration, consider the two different gesture types in **Figures 1**–**3** from Parrill (2010: 651). She discusses two gestures performed by participants in an experiment in which they had to describe to their interlocuters a scene from a cartoon they had just seen: an event of a skunk hopping across the room. In one case, the speaker performed an OVG in which the hand represents the skunk and the trajectory of the hand represents the trajectory of the skunk. Crucially, the event was depicted as if observed from a distance and in a rather schematic way. In the other case, in contrast, the speaker performed a CVG in which he enacted the hopping skunk with his entire body.



FIGURE 2 | Observer Viewpoint Gestures

The Relationship Between Linguistic and Gestural Viewpoint

We follow De Ruiter, 1998, De Ruiter, 2000, De Ruiter, 2007; see also Kendon, 2004) in assuming that combinations of speech and gesture convey a complex multimodal message that is planned by a central cognitive process and then dispatched into disparate channels. The information conveyed via the two channels may be either redundant (thus avoiding misunderstanding) or complementary. Although Goldin-Meadow. (1999) has shown that in cases of learning novel concepts that involve transition between different cognitive states, mismatches between gesture and speech can be quite productive and useful, the default should be for the information conveyed via one channel to be coherent with information conveyed via the other one. If Parrill (2010) is right in assuming that gestural and linguistic viewpoint have a common conceptual source, perspective is an integral part of the multimodal message conveyed by the combination of speech and gesture. Consequently, the default should be for such a message to be coherent with regard to perspective as well, i.e. it should express a single perspective by default. Concerning perspective, this means that at least in the absence of intervening factors there should be a preference for perspective alignment. This should concern production as well as comprehension, i.e. the speaker should in the default case plan to convey a message that is coherent with regard to perspective and the listener should in the default case expect the speaker to convey such a message.

There is, however, the complicating case of multiple perspectives. It is possible that a speaker plans to convey a thought from more than one perspective. Although there is some discussion about the possibilities of linguistic realizations of such multiple perspectives within the evidentiality literature (see Evans, 2005 and Bergqvist, 2015 for some discussion), a systematic investigation of the linguistic tools to simultaneously convey multiple perspectives is still outstanding. As for the gestural realization of perspective, McNeill (1992), Cassell et al. (1999) and Parrill (2009) discuss exactly such examples of dual viewpoints. Cassell et al. (1999) present an example where someone hands something to himself while uttering she got something. Here, the arm and hand embodies the giver and the rest of the body the receiver. McNeill (1992) and Parrill (2009) discuss similar cases as well as cases where someone performs an OVG and a CVG at the same time. Parrill (2009) reports an example where the narrator talks about a character and impersonates the reported character by performing a body lean to mimic the action of this character, hence a CVG, and at the same time indicates the trajectory of a certain path taken by the character, clearly an OVG.

So far, the relation between linguistic and gestural perspective has not been systematically investigated. If, however, there are techniques to represent dual perspectives within speech alone and within gesture alone, as we have pointed out above, it is not implausible to assume that a speaker can also choose to convey two perspectives at the same time and realize one via gesture and one via speech. We believe, however, it is equally fair to assume that such dual viewpoint realizations need specific licensing conditions and are the exception rather than the general case. Future research will have to shed light on this.

Existing research on the relationship of viewpoint gestures and speech has for the most part focused on general factors influencing the frequency with which co-speech gestures are performed or which types of gestures are preferred, depending on the accompanying speech. McNeill (1992), for example,



FIGURE 3 | Character Viewpoint Gesture.

observed that linguistic complexity influences which type of co-speech gesture is performed, with utterances containing transitive verbs having a tendency to be accompanied by CVGs and sentences with intransitive verbs having a tendency to be accompanied by OVGs. Additionally McNeill (1992), notes a tendency for causally central events to be accompanied by CVGs. The first observation was confirmed by Parrill (2010), while the second was disconfirmed: While causally central events were more often accompanied by gestures than peripheral events, there was no contrast between CVGs and OVGs. At the same time, Parrill (2010), Parrill (2012) found other factors that had an influence on which type of gesture was chosen by speakers. First, she found that speakers performed CVGs more often than OVGs when the information conveyed by the utterance was new to the hearer, while for shared information it was the other way around. Second, she found that the internal structure of the reported event was a crucial factor, too, with events involving the display of affects or a prominent use of the character's hands and torso triggering more CVGs than OVGs, while for events involving trajectories it was the other way around.

Having discussed perspective taking on the linguistic level, on the level of co-speech gestures, as well as their relationship, we will discuss an experimental study in which we systematically investigated the interaction of the two kinds of perspective taking in light of our assumptions concerning perspectival coherence in *The experimental study* (see Ebert and Hinterwimmer to appear for a study of self-pointing CVGs in reported speech vs. direct speech vs. FID and a proposal to account for such and other demonstrations in quotation on basis of Ebert et al. (2020) account for co-speech gestures).

THE EXPERIMENTAL STUDY

As stated above, we assume that

(a) Gesture and speech together convey a multimodal message that is planned by a central cognitive process and then dispatched into disparate channels (De Ruiter, 1998, De Ruiter, 2000, De Ruiter, 2007; see also; Kendon, 2004).
(b) Gestural and linguistic viewpoint have the same conceptual source (Parrill, 2010).
(c) Perspective is an integral part of the multimodal message to be conveyed and,
(d) the default is for this message to be coherent.

We thus predict a strong preference for gestural and linguistic perspective to be aligned in a single utterance, at least in the absence of intervening factors.

In order to test this prediction, we conducted a forced-choice experiment in German in which we tested whether utterances in which the linguistically expressed perspective aligns with the perspective expressed on the gestural level are preferred to test items in which this is not the case. We compared variants of test items clearly expressing the perspective of an individual participating in the events described with variants of test items which clearly express the speaker's or narrator's perspective. Variants of the former kind were always construed in such a way that the opening sentence describes the feelings, intentions or thoughts of a protagonist in a particular situation which the second sentence specifies in more detail, while the final sentence renders a thought that the respective protagonist has in that situation in the form of FID. Variants of the latter kind, in contrast, always contained an opening sentence in the form of a general statement in present tense expressing an evaluation of or opinion about the individual participating in the event described by the following sentence. Additionally, that individual was referred to by a DPro. The two variants of each test item basically described the same situation. Each variant of each test item was videotaped in two different versions: In one version, the speaker performed a CVG while uttering the respective final sentence. In the second version, the speaker performed an OVG. Two examples are provided in (9) and (10). The respective CVGs and OVGs, which are described beneath the items, were performed while uttering the portion of the respective sentence marked in boldface. All stimuli, anonymized data, and codes can be accessed via the following link: https://osf.io/4bqpx/.

(9) a. Leon ist ein begeisterter Sportler. Als der sich neulich beim Fußballspielen den Ball erkämpfte, **kickte er ihn sofort in Richtung Tor (narrator perspective = NP)**.
Leon is an enthusiastic athlete. When he (DPro) recently won the ball while playing soccer, he immediately kicked it in the direction of the goal.
b. Leon spielte am Wochenende Fußball. Nach einigem Gerangel hatte er sich den Ball erkämpft. **Toll, jetzt konnte er ihn direkt in Richtung Tor schießen (character perspective = CP)**
Leon played soccer on the weekend. After some scramble, he had finally won the ball. Great, now he could directly kick it in the direction of the goal!
CVG: The speaker performs a kicking movement with her right leg and foot, displaying an enthusiastic facial expression.
OVG: The speaker presses her index finger on her thumb, then releasing it quickly, thus imitating a kicking movement with the index finger.

(10) a. Denise ist ein richtiger Tollpatsch. Als die neulich nach Feierabend das Büro verließ, hat sie nicht richtig aufgepasst **und ist voll gegen die Tür geknallt! (NP)**
Denise is a real klutz. When she (DPro) recently left the office at the end of the work day, she did not really pay attention and slammed into the door at full tilt.
b. Denise hatte es eilig. Beim Verlassen des Büros passte sie nicht richtig auf. **Autsch, jetzt war sie voll gegen die Tür geknallt (CP).**
Denise was in a hurry. When leaving the office, she did not really pay attention. Ouch, now she had slammed into the door at full tilt!

CVG: The speaker throws back her head and imitates the astonished facial expression of someone banging her head against a door, eyes wide open.
OVG: The speaker moves the index finger of her right hand quickly into the direction of her vertically upheld left hand and lets it collide with her hand and bounce back.

The variants of the test items were evenly distributed across two lists, so that each participant saw only either the NP or the CP variant of each test item in both conditions, i.e. in the version where the speaker performs a CVG while uttering it and in the version where she performs an OVG while uttering it. The participants then had to choose the video with the gesture that they thought better fits the spoken utterance. The method we employed is thus similar to the one employed in grammaticality and/or pragmatic felicity judgement tasks, the underlying reasoning being that speakers have intuitions regarding the interaction of linguistic and gestural perspective taking that reflect underlying unconscious principles in the same way as they have intuitions reflecting unconscious syntactic, semantic or pragmatic principles.

Now, if there is a strong preference for linguistic and gestural perspective to be aligned, the CVG version should be chosen more often for the CP variants of the test items, while the OVG should be chosen more often for the NP version of the test items. In (9b), for example, the sentence in boldface renders a thought of Leon that expresses Leon's perspective on the event of him kicking the ball in the direction of the goal. When performing the CVG described above while uttering that sentence, the speaker enacts the event of Paul kicking the ball in the direction of the goal from his perspective. Consequently, the linguistically and the gesturally expressed perspectives align when (9b) is combined with the CVG. When performing the OVG described above while uttering the sentence in boldface in (9b), in contrast, she depicts the event of Leon kicking the ball in the direction of the goal from an outside perspective. There is thus a mismatch between the linguistically and the gesturally expressed perspective.

In (9a), the opening sentence in combination with the use of the DPro to refer to Leon in the temporal adjunct clause ensures that the sentence in boldface is attributed to the narrator or speaker, i.e. it reports the event of Leon kicking the ball in the direction of the goal from the narrator's or speaker's perspective. When the speaker performs the OVG described above while uttering the sentence in boldface in (9a), the linguistically expressed perspective aligns with the gesturally expressed one, since the outside perspective conveyed by the OVG can easily be construed as the narrator's or speaker's perspective. When she performs the CVG while uttering the sentence in (9a), in contrast, there is a mismatch between the linguistically expressed narrator's or speaker's perspective and the gesturally expressed perspective, which is Leon's perspective.

Consequently, if there is a requirement or a strong preference for the linguistically and the gesturally expressed perspective to be aligned (9a) should be chosen more often in combination with the OVG, and (9b) in combination with the CVG. At the same time, as we have seen in *Observer viewpoint and character viewpoint gestures* above, there is a preference for complex events (i.e. events involving at least two participants) as well as for events contained in sentences introducing new information to be accompanied by CVGs. The sentences in boldface in (9) and (10) clearly convey information that is new to the participants and they contain transitive verbs. Consequently, if there is no requirement or strong preference for the linguistically expressed perspective to align with the gesturally expressed perspective, CVGs should be preferred across both conditions according to the above-mentioned findings of Parrill (2010, 2012).

In order to make sure that our stimuli were actually interpreted as intended, i.e. as either expressing the character's or the narrator's perspective, we conducted an informal forced-choice study[1]. Participants saw muted versions of both the CVG- and the OVG-variant of each test item together with a short, neutral description of the situation reported by both variants of the respective item. For (9a–b), for example, the following description was provided: *The following video is about a soccer player who is kicking the ball into the goal.* Participants (n = 18) then had to decide for each item which of the two videos corresponds to the character's and which one to an observer's perspective, i.e., in effect, whether the gesture they saw was a CVG or an OVG[2]. Additionally, they had to indicate on a scale from 1 to 5 how sure they were of their judgement, with 1 expressing minimal and 5 expressing maximal confidence. In terms of the design, the experimental task was almost equivalent to showing both video versions (CharVideo and ObsVideo) of an item and asking participants to choose which one is more likely to be a CVG and which one is more likely to be an OVG (i.e. it was almost equivalent to a classical forced-choice design)[3].

As **Figure 4**, Panel **(A)** shows, in the overwhelming majority of cases, the gestures were interpreted as intended, i.e. videos with a CVG were interpreted as conveying the character's perspective and videos with an OVG as conveying an observer's perspective. Additionally, as Panel **(B)** shows, in the vast majority of cases, participants were confident in their choices, i.e. there were very few low confidence responses (i.e. only 14% of all responses were given with confidence lower than 3).

We now turn to a detailed description and discussion of the experimental study itself.

[1]We are grateful to two anonymous reviewers as well as the editor for urging us to conduct this study.

[2]Participants were recruited through Prolific (https://prolific.ac/) for monetary compensation (3,75 £; 7,50 £/h), just as in the main experiment. Only persons who had not participated in the main experiment could participate. Two participants had to be excluded because they self-identified as non-native speakers of German. The study also included simple questions which were included in order to check whether participants were paying attention. No one had to be excluded on the basis of answering these questions incorrectly.

[3]We chose that design in order to keep the study similar to the design of the main experiment.

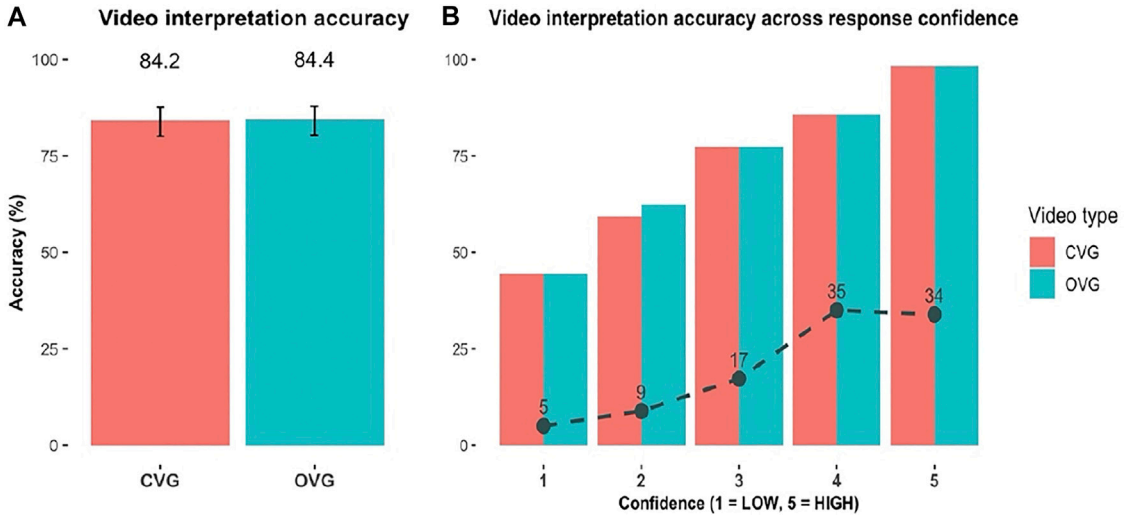


FIGURE 4 | The bar plots show mean percentages of accurately recognizing CVG and OVG videos. Panel **(A)** shows the mean accuracy for the two types of videos (the confidence intervals are calculated using prop.test() function in R). Panel **(B)** shows accuracy across five levels of response confidence with "1" denoting least and "5" most confident responses. The overplotted dots joined with a dashed line in Panel **(B)** show percentages of responses in each level of the confidence scale (the numbers above the dots are the percentages).

METHOD

Participants

Eighty-five native speakers of German (45 males, 40 females, age 18–65) were recruited through Prolific (https://prolific.ac/) for monetary compensation (3,75 £; 7,50 £/h).

MATERIALS

We constructed 20 experimental items similar to those in (9) and (10), interspersed with 24 fillers. For each item there was an NP variant (similar to (9a) and (10a)) and a CP variant (similar to (9b) and (10b)). Both variants always described the same situation. Each variant of each test item was videotaped in two different versions: In one version, the speaker performed a CVG while uttering the respective final sentence. In the second version, the speaker performed an OVG. Consequently, each item came in four different conditions: NP-CVG, NP-OVG, CP-CVG and CP-OVG.

The fillers consisted of two sentences and involved pointing to a location in the gesture space. In the first sentence, two discourse referents were introduced by referential expressions accompanied by pointing gestures that anchored the referents in the gesture space. In 12 filler items both referents had the same gender (e.g. *Gestern auf der Party hat Peter Linus beleidigt.* Engl.: *Yesterday at the party, Peter insulted Linus.*, plus pointing to a point left in the central gesture space in front of the speaker's body when uttering Peter and to a point right when uttering Linus.), and in 12 filler items the gender was different (e.g. *Gestern hat Martin Claudia zum Abendessen eingeladen.* Engl.: *Yesterday Martin invited Claudia for dinner.*, plus pointing to a point left in the central gesture space in front of the speaker's body when uttering *Martin* and to a point right when uttering *Claudia.*). The second sentence always contained a pronoun and the speaker pointed to the location associated with the object referent while uttering the pronoun. For each filler item there were two different versions. For the 12 filler items where both referents had the same gender there was a version with a DPro and a version with a PPro (e.g. *Der hat dann sofort angefangen zu weinen.* Engl.: *He (DPro) then started crying immediately*, and *Er hat dann sofort angefangen zu weinen*. Engl.: *He then started crying immediately*). For the 12 filler items where the gender was different for the two referents there was a version with a male DPro and a version with a female DPro (e.g. *Der hat sich sehr darüber gefreut*. Engl.: *He (DPro) was very happy about it* and *Die hat sich sehr darüber gefreut*. Engl.: *She (DPro) was very happy about it*).

Procedure

The experiment, which involved a forced-choice task, was conducted online. The NP- and the CP-variants of the test items were evenly distributed across two lists and presented in pseudo-randomized order, interspersed with the fillers. Consequently, participants saw either both the NP-OVG and the NP-CVG version of the respective test item or the CP-CVG and the CP-OVG version. Concerning the filler items, they always saw both versions, i.e. the versions with the DPro and the PPro and the versions with the DPro matching the gender of the referent associated with the location pointed at and the version not matching it.

The task for the participants was to choose the version of the test items or the fillers in which the combination of language and gesture is more natural according to their intuitions. The question that appeared below the video was: *Welche Geste passt besser zur*

sprachlichen Äußerung in den Videos? Engl.: *Which gesture better fits the spoken utterance in the videos?*). Participants were told beforehand to pay good attention to sound and picture and that they had the option to replay the videos. They were also told that in some cases the utterances were identical in the two versions and the gestures accompanying them were different (as in the test items), while in other cases it was the other way around (as in the filler items). Before the experiment started, the participants were shown two trial items to familiarize them with the form of the experiment. They also had to do four simple matching tasks, the purpose of which was to check whether they were paying attention: They were shown two pictures of animals, fruits etc. and had to decide which of the two pictures matched a word such as *cat*, *apple* etc. Participants who would choose the wrong picture in one of the four tasks were to be excluded from the final analysis (however, there was no such case).

Data Analysis

We excluded seven participants from the analysis because they either did not complete the task or completed it in less than 10 min (the approximate duration was 30 min). All data processing and analyses were carried out in R (Core Team, 2020). Since the responses were binomial (CVG or OVG), we analyzed the proportions of CVG responses using mixed-effects logistic regression through the R package brms (Bürkner, 2017). We used condition as the predictor variable with CP as the reference level. To avoid the extreme probability values (0 and 1), we used weakly informative priors, N(0, 2.5), instead of the default priors of brms for logistic regression (Student-*t* with df = 3, mu = 0 and sigma = 2.5). The model was run with four sampling chains each of which ran for 5,000 iterations with a warm-up period of 2000 iterations. We also fit another model with the same form but without any predictors, an intercept only model, to statistically test if the CVG option was chosen more often than the OVG option.

For each effect we report its mean and 95% CrI under the posterior distribution. We use CrI to make inferences about the presence of an effect. If the 95% CrI for an effect does not include zero we consider that there is compelling evidence for that effect. We also report the posterior probability of an effect being greater than zero or less than zero depending on the sign of the estimated parameter mean. The posterior probability is calculated using the posterior sample for a parameter generated by the statistical model and it is the proportion of the sample less than or greater than zero.

Results

The response proportions are plotted in **Figure 5**, and the results of the data analysis are listed in **Table 1** for the model with condition as the predictor and **Table 2** for the intercept only model. We found that overall the CVG option was chosen clearly more often than the OVG option (**Table 2**). Moreover, the proportions of CVG responses were influenced by the condition (CP vs. NP) such that in the NP condition the CVG option was chosen less often than in the CP condition; although this effect was small and the support for it was weaker (**Table 1**).

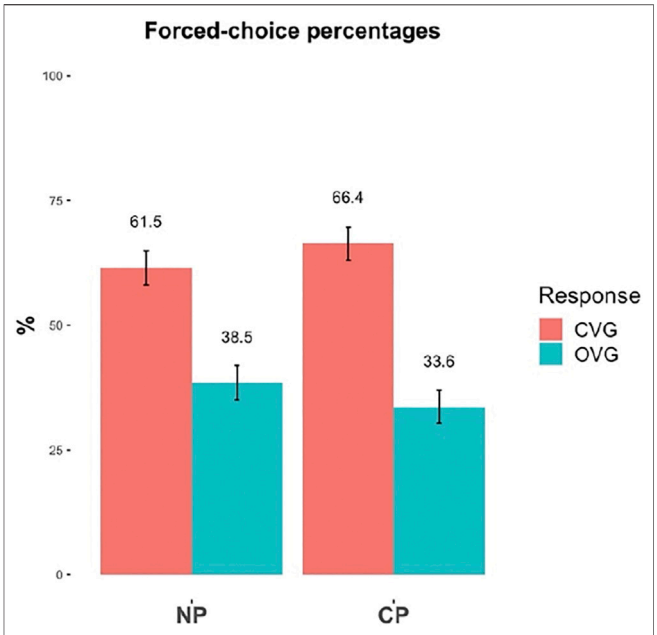


FIGURE 5 | Each bar shows the percentage of choosing an option (CVG or OVG) across the two conditions, NP and CP (the confidence intervals are calculated using prop.test() function in R).

TABLE 1 | Results from statistical analysis—estimates of the model, corresponding 95% credible intervals (95% CrI) and the posterior probabilities (Post. Prob.). Effects for which the CrI excludes zero are shown in bold. 'Intercept' denotes the effect of selecting CVG more often than OVG in the CP condition, and Narrator denotes the effect of selecting CVG in the Narrator condition compared to the CP.

Effect	Estimate	95% CrI	Post. prob
Intercept	1.10	[0.35, 1.89]	0.997
Narrator	−0.32	[−0.75, 0.12]	0.926

TABLE 2 | Results from statistical analysis—estimates of the model, corresponding 95% credible intervals (95% CrI) and the posterior probabilities (Post. Prob.). Effects for which the CrI excludes zero are shown in bold. In this intercept-only model 'Intercept' denotes the effect of selecting CVG more often than OVG.

Effect	Estimate	95% CrI	Post. prob
Intercept	0.90	[0.20, 1.63]	0.993

DISCUSSION

At first sight, the experimental results seem to be incompatible with the assumption of a strong preference for the linguistically expressed perspective to align with the perspective expressed on the level of co-speech gestures. If such a strong preference existed, participants should have chosen the combination of CP and CVG more often than the combination of CP and OVG, on the one hand, and the combination of NP and OVG more often than the

combination of NP and CVG, on the other. Rather, they showed a clear preference for CVGs across both the CP and the NP condition. This is in line with the findings of McNeill (1992) and Parrill (2010), Parrill (2012) that sentences with transitive verbs and sentences introducing new information have a tendency to be accompanied by CVG since the relevant portions of our test items all introduced information that was new to the participants and the vast majority of them contained transitive sentences.

Although the preference for CVGs was slightly stronger in the CP condition than in the NP condition, this preference did not turn out to be significant, so no conclusions can be derived from it at this point. Nevertheless, there is still the possibility that linguistic and gestural perspective are preferably aligned, but that this is a constraint that ranks below the requirement that sentences introducing new information and describing rather complex events should be accompanied by CVGs. Concerning the question of why the two latter preferences exist, we would like to tentatively suggest that this is due to CVGs always being more informative than OVGs, which only depict events in a rather generic and schematic way. CVGs, in contrast, by making use of the speaker's entire body in combination with her facial expression, convey much more fine-grained and detailed information which is particularly useful when the sentence they accompany introduces new information and when the event described by that sentence is rather complex.

Recall that we derived our hypothesis of a strong preference for gestural and linguistic perspective to be aligned in a single utterance from the following assumptions:

(a) Gesture and speech together convey a multimodal message that is planned by a central cognitive process and then dispatched into disparate channels (De Ruiter, 1998, De Ruiter, 2000, De Ruiter, 2007; see also Kendon, 2004).
(b) Gestural and linguistic viewpoint have the same conceptual source (Parrill, 2010).
(c) Perspective is an integral part of the multimodal message to be conveyed.
(d) The default is for this message to be coherent.

Since we still consider these assumptions to be very plausible, the most straightforward way to reconcile them with the findings of our study would be to assume the preference for informative gestures to be strong enough to overwrite the default. At the same time, the differences between the OVGs and the CVGs in our study can be interpreted as revealing a potential limitation, which might have affected the results[4]. After all, the co-speech gestures did not only differ with respect to perspective and informativity, but also with respect to size, since the CVGs involved the speaker's entire body and facial expression, while OVGs only involved the hands. This difference in size quite plausibly made the CVGs more salient than the OVGs, which might have played a role in the general preference for CVGs. Additionally, CVGs might have been judged as more natural in virtue of the speaker being more fully engaged. One might consider to conducting a follow-up study in which the gestures are more comparable in size and speaker's engagement and therefore in saliency and naturalness. To give a concrete example, the OVG in (9a–b), repeated here as (11a–b), could be replaced by a full-body gesture where the speaker steps back and follows the path of an imaginary ball with her index finger and gaze. This would, however, mean departing from the standard view that OVGs are performed with only the hands. It might hence be more feasible to replace the CVG by a CVG where the gesturer does only a small kick of the foot but does not incorporate her upper body or facial expression.

(11) a. Leon ist ein begeisterter Sportler. Als der sich neulich beim Fußballspielen den Ball erkämpfte, **kickte er ihn sofort in Richtung Tor. (narrator perspective = NP).**
Leon is an enthusiastic athlete. When he (DPro) recently won the ball while playing soccer, he immediately kicked it in the direction of the goal.
b. Leon spielte am Wochenende Fußball. Nach einigem Gerangel hatte er sich den Ball erkämpft. **Toll, jetzt konnte er ihn direkt in Richtung Tor schießen. (character perspective = CP).**
Leon played soccer on the weekend. After some scramble, he had finally won the ball. Great, now he could directly kick it in the direction of the goal!
CVG: The speaker performs a kicking movement with her right leg and foot, displaying an enthusiastic facial expression.
OVG: The speaker presses her index finger on her thumb, then releasing it quickly, thus imitating a kicking movement with the index finger.

We are planning to conduct a follow-up study with gestures that are more comparable in saliency and naturalness in order to test whether the difference in size between the two kinds of gestures in our original study had an influence on the results. Additionally, in virtue of the preference for CVGs being potentially linked to the speaker's introducing new information, we are planning to conduct a follow-up study which does not only contain stimuli with new events, as the study reported in this paper does, but which contains both stimuli with new and stimuli with given events. Our prediction is that the preference for CVGs should at least be weaker in the stimuli with given events and potentially be overwritten by the preference for perspective alignment.

Let us finally add a grain of salt: Since the division of labour between gesture and speech in general and viewpoint issues in particular have not been systematically investigated and are not settled yet, we could not control for potential intervening factors that might elicit multiple viewpoint representations and a potential split of viewpoints on gesture and speech. As it stands, our results are equally compatible with the possibility that the gestural channel preferably transports the character

[4]We thank an anonymous reviewer for pointing this out to us.

perspective, independent of the perspective that the speech channel transports.

CONCLUSION

The topic of this paper was an investigation of the interaction of perspective taking expressed on the linguistic level with perspective taking expressed on the level of co-speech gestures. We investigated via a forced-choice experimental study whether there is a preference for the linguistic perspective to be aligned with the gestural perspective. The experimental results provided no evidence that there is such a preference for perspective alignment. There was a general preference for CVGs, which was slightly, but not significantly, stronger when the respective CVG accompanied a sentence expressing a character's perspective, however, than when it accompanied a sentence expressing the narrator's or speaker's perspective. After all, the results of our study did not support our initial hypothesis that there is a preference for perspective alignment. It might, however, still be the case that there is such a preference and it can be overwritten by the preference for more informative gestures or by a preference to transport the character's perspective in the gestural channel. In future research we are planning to conduct studies which are aimed at testing for this possibility. First, we are planning a follow-up study with CVGs and OVGS that are more comparable in saliency and naturalness than those in our original study in order to test whether the differences in saliency and naturalness were a confounding factor. Secondly, we are planning to conduct a follow-up study which contains both stimuli with new and stimuli with given events in order to test whether the preference for CVGs in our original study was at least partly due to CVGs being preferred in sentences conveying new information. Finally, we are planning a study in which the following two cases will be compared:

a. Utterances where the speaker describes an event she participated in from her first person perspective while performing a CVG that depicts the actions of another participant.
b. Utterances where the speaker describes an event she participated in from her first person perspective while performing a CVG that depicts her own actions.

If our assumptions are correct, utterances instantiating the constellation in a. should be clearly dispreferred compared to utterances instantiating the constellation in b. since the former in contrast to the latter involve conflicting perspectives. At the same time, they do not differ in informativity, since the gestures in both cases are CVGs. If there are no clear differences between the two cases, this would be a strong indication that there is not even a weak default preference for linguistic and gestural perspective to be aligned. Since only CVGs and no OVGs are involved, a gestural preference for CVGs should not influence the results, either.

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/4bqpx/.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

SH and CE jointly developed the experimental design and the analysis and were jointly responsible for the execution of the experiment, UP was responsible for the statistical analysis. The introduction and the parts on Free Indirect Discourse and DPros were written by SH alone, the section on gestures by CE, and the sections discussing the experiment, the discussion and the conclusion were jointly written by SH, CE and UP.

FUNDING

The research reported in this paper was funded by the DFG 1459 within the SFB 1252 Prominence in Language (Project-ID 281511265) (SH) and as part of the project PSIMS of the Priority Program 1727 XPRAG. de (EB 523/1-1) (CE).

ACKNOWLEDGMENTS

We thankfully acknowledge the support of the German Research Foundation (DFG). We would like to thank Theresa Stender, Sebastian Walter, Janne Schmandt, Felix Jüstel and Magdalena Schmitz for help with item generation and execution of the forced-choice study reported in *The Experimental Study*. We also thank Gregor Brinkmeier from studiumdigitale of the University of Frankfurt for supplying us with a studio and all technical equipment for the recordings as well as help with the postprocessing of the recordings.

REFERENCES

Abrusán, M. (2020). The Spectrum of Perspective Shift: Protagonist Projection vs. Free Indirect Discourse. *Linguistics And Philosophy*. *Adv. access published July* 22. doi:10.1007/s10988-020-09300-z

Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Boston: Routledge & Kegan Paul.

Bergqvist, H. (2015). Epistemic Marking and Multiple Perspective: An Introduction. *Lang. Typology Universals* 68 (2), 123–141. doi:10.1515/stuf-2015-0007

Bosch, P., Katz, G., and Umbach, C. (2007). "The Non-subject Bias of German Demonstrative Pronouns,". Editors M. Schwarz-Friesel, M. Consten, and M. Knees, 145–164. doi:10.1075/slcs.86.13bos Anaphors in textAmsterdam & Philadelphia: Benjamins.

Bosch, P., and Umbach, C. (2006). Reference Determination for Demonstrative Pronouns. Proceedings of the conference on intersentential pronominal reference in child and adult language. Editors D. Bittner and N. Gagarina (Berlin: Zentrum für Allgemeine Sprachwissenschaft, Sprachtypologie und Universalforschung), 48, 39–51.

Brinton, L. (1980). 'Represented Perception': A Study in Narrative Style. *Poetics* 9, 363–381. doi:10.1016/0304-422x(80)90028-5

Bürkner, P.-C (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw. Articles* 80 (1), 1–28. doi:10.18637/jss.v080.i01

Cassell, J., McNeill, D., and McCullough, K.-E. (1999). Speech-gesture Mismatches: Evidence for One Underlying Representation of Linguistic and Nonlinguistic Information. *P&C* 7 (1), 1–34. doi:10.1075/pc.7.1.03cas

Charnavel, I., and Mateu, V. (2015). The Clitic Binding Restriction Revisited: Evidence for Antilogophoricity. *Linguistic Rev.* 32 (4), 671–701. doi:10.1515/tlr-2015-0007

Clements, G. N. (1975). The Logophoric Pronoun in Ewe: Its Role in Discourse. *The J. West African Lang.* 10, 141–177. doi:10.17487/rfc0683

De Ruiter, J. P. (1998). Gesture and Speech Production, PhD Thesis. Nijmegen, Netherlands: University of Nijmegen.

De Ruiter, J. P. (2007). Postcards from the Mind. *Gest* 7, 21–38. doi:10.1075/gest.7.1.03rui

De Ruiter, J. P. (2000). "The Production of Gesture and Speech," in *Language and Gesture*. Editor D. McNeill (Cambridge, UK: Cambridge University Press), 284–311. doi:10.1017/cbo9780511620850.018

Dirscherl, F., and Pafel, J. (2015). Die vier Arten der Rede- und Gedankendarstellung. Zwischen Zitieren und Referieren. *Linguistische Berichte* 241, 3–47.

Doron, E. (1991). "Point of View as a Factor of Content," in *Proceedings of Semantics and Linguistic Theory*. Editors S. K. Moore and A. Z. Wyner (Ithaca, NY: Cornell University), 1, 51–64. doi:10.3765/salt.v1i0.2997

Ebert, C., and Hinterwimmer, S. (2020). "Free Indirect Discourse Meets Character Viewpoint Gestures: A Reconstruction of Davidson's Demonstration Account with Gesture Semantics," in To Appear appear Linguistic Evidence 2020 Proceedings.

Ebert, Ch., Ebert, C., and Hörnig, R. (2020). *Demonstratives as dimension shifters. Proceedings of Sinn und Bedeutung 24*. Osnabrück: University of Osnabrück, 161–178.

Eckardt, R. (2014). *The Semantics of Free Indirect Discourse. How Texts Allow to Mind-Read and Eavesdrop*. Leiden: Brill.

Evans, N. (2005). View with a View: Towards a Typology of Multiple Perspective Constructions. *Bls* 31, 93–120. doi:10.3765/bls.v31i1.3429

Goldin-Meadow, S. (1999). The Role of Gesture in Communication and Thinking. *Trends Cogn. Sci.* 3, 419–429. doi:10.1016/s1364-6613(99)01397-2

Hinterwimmer, S. (2015). "A Unified Account of the Properties of German Demonstrative Pronouns," in The proceedings of the workshop on pronominal semantics at NELS. Editors P. Grosz, P. Patel-Grosz, and I. Yanovich (Amherst, MA: GLSA Publications, University of Massachusetts), 61–107.

Hinterwimmer, S. (2020). Zum Zusammenspiel von Erzähler- und Protagonistenperspektive in den Brenner-Romanen von Wolf Haas," in *Sprachlichen Strukturen der Narration, special issue of Zeitschrift für Germanistische Linguistik*. Editor S. Zeman ((ZGL)). doi:10.1515/zgl-2020-2013

Hinterwimmer, S., and Bosch, P. (2016). "Demonstrative Pronouns and Perspective," in *The Impact of Pronominal Form on Interpretation*. Editors P. Patel and P. Patel-Grosz (Berlin/New York: De Gruyter (Studies in Generative Grammar)).

Hinterwimmer, S., and Bosch, P. (2017). "Demonstrative Pronouns and Propositional Attitudes," in *Pronouns in Embedded Contexts (Studies in Linguistics and Philosophy)*. Editors P. Patel-Grosz, P. G. Grosz, and S. Zobel (Washington, DC: Springer), 282–301. doi:10.1007/978-3-319-56706-8_4

Hinterwimmer, S., and Brocher, A. (2018). An Experimental Investigation of the Binding Options of Demonstrative Pronouns in German. *Glossa: A J. Gen. Linguistics* 3 (1), 77. doi:10.5334/gjgl.150

Hinterwimmer, S., Brocher, A., and Patil, U. (2020). Demonstrative Pronouns as Anti-logophoric Pronouns: An Experimental Investigation. *Dialogue Discourse* 11 (2), 110–127. doi:10.5087/dad.2020.204

Hinterwimmer, S., and Meuser, S. (2019). , "Erlebte Rede und Protagonistenprominenz,". *Rede- und Gedankenwiedergabe in narrativen Strukturen – Ambiguitäten und Varianz, Linguistische Berichte, Sonderheft*. Editors S. Engelberg, C. Fortmann, and I. Rapp (Hamburg: Helmut Buske Verlag), 27, 177–200.

Hinterwimmer, S. (2019). Prominent Protagonists. *J. Pragmatics* 154, 79–91. doi:10.1016/j.pragma.2017.12.003

Hinterwimmer, S. (2017). Two Kinds of Perspective Taking in Narrative Texts, *Salt* In D. Burgdorf, J. Collard, S. Maspong, and B. Stefánsdóttir (eds.), Proceedings of Semantics and Linguistic Theory(SALT) 27, 282–301. doi:10.3765/salt.v27i0.4153

Kaiser, E. (2013). Looking beyond Personal Pronouns and beyond English: Typological and Computational Complexity in Reference Resolution. *Theor. Linguistics* 39, 109–122. doi:10.1515/tl-2013-0007

Kaiser, E. (2011a). "On the Relation between Coherence Relations and Anaphoric Demonstratives in German," in *Proceedings of Sinn und Bedeutung*. Editors I. Reich, E. Horch, and D. Pauly (saarbrücken, Germany: Saarland University Press), 15, 337–351.

Kaiser, E. (2010). Effects of Contrast on Referential Form: Investigating the Distinction between strong and Weak Pronouns. *Discourse Process.* 47, 480–509. doi:10.1080/01638530903347643

Kaiser, E. (2011b). Salience and Contrast Effects in Reference Resolution: The Interpretation of Dutch Pronouns and Demonstratives. *Lang. Cogn. Process.* 26 (10), 1587–1624. doi:10.1080/01690965.2010.522915

Kaiser, E., and Trueswell, J. C. (2008). Interpreting Pronouns and Demonstratives in Finnish: Evidence for a Form-specific Approach to Reference Resolution. *Lang. Cogn. Process.* 23 (5), 709–748. doi:10.1080/01690960701771220

Kaplan, D. (1989). Demonstratives. in *Themes from Kaplan*. Editors J. Almog, J. Perry, and H. Wettstein (Oxford: Oxford University Press), 565–614.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University. doi:10.1017/cbo9780511807572

Kita, S., and Özyürek, A. (2003). What Does Cross-Linguistic Variation in Semantic Coordination of Speech and Gesture Reveal?: Evidence for an Interface Representation of Spatial Thinking and Speaking. *J. Mem. Lang.* 48 (1), 16–32. doi:10.1016/s0749-596x(02)00505-3

Kopp, S., Tepper, P., and Cassell, J. (2004). Towards Integrated Microplanning of Language and Iconic Gesture for Multimodal Output. In Proceedings of the 6th International Conference on Multimodal Interfaces, 97–104. doi:10.1145/1027933.1027952

Loehr, D. (2004). Gesture and Intonation. PhD thesis. Washington, DC: Georgetown University.

Maier, E. (2015). Quotation and Unquotation in Free Indirect Discourse. *Mind Lang.* 30 (3), 345–373. doi:10.1111/mila.12083

Maier, E. (2017). "The Pragmatics of Attraction," in *The Semantics and Pragmatics of Quotation*. Editors P. Saka and M. Johnson (Dordrecht: Springer), 259–280. doi:10.1007/978-3-319-68747-6_9

Mayol, L., and Clark, R. (2010). Pronouns in Catalan: Games of Partial Information and the Use of Linguistic Resources. *J. Pragmatics* 42, 781–799. doi:10.1016/j.pragma.2009.07.004

McNeill, D. (1992). *Hand and Mind. What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

Nishigauchi, T. (2014). Reflexive Binding: Awareness and Empathy from a Syntactic point of View. *J. East. Asian Linguist* 23, 157–206. doi:10.1007/s10831-013-9110-6

Parrill, F. (2012). Interactions between Discourse Status and Viewpoint in Co-speech Gesture. in *Viewpoint in Language: A Multimodal Perspective*. Editors B. Dancygier and E. Sweetser (Cambridge: CUP), 97–112.

Parrill, F. (2009). Dual Viewpoint Gestures. *Gest* 9 (3), 271–289. doi:10.1075/gest.9.3.01par

Parrill, F. (2010). Viewpoint in Speech-Gesture Integration: Linguistic Structure, Discourse Structure, and Event Structure. *Lang. Cogn. Process.* 25 (5), 650–668. doi:10.1080/01690960903424248

Patil, U., Bosch, P., and Hinterwimmer, S. (2019). Constraints on German Diese Demonstratives: Language Formality and Subject-Avoidance. *Glossa: A J. Gen. Linguistics* 5 (1), 1–22. doi:10.5334/gjgl.962

Pearson, H. (2015). The Interpretation of the Logophoric Pronoun in Ewe. *Nat. Lang. Semantics* 23, 77–118. doi:10.1007/s11050-015-9112-1

Pittenger, R. E., Hockett, C. F., and Danehey, J. J. (1960). *The First Five Minutes: A Sample of Microscopic Interview Analysis*. New York: Martineau.

R Core Team (2020). R: A Language and Environment for Statistical Computing [Computer Software Manual]. Vienna Austria: Retrieved from https://www.R-project.org.

Rauh, G. (1978). *Linguistische Beschreibung Deiktischer Komplexität in Narrativen Texten*. Tübingen: Narr Verlag.

Redeker, G. (1996). "Free Indirect Discourse in Newspaper Reports," in *Linguistics in the Netherlands 1996*. Editors C. Cremers and M. den Dikken (Amsterdam: Benjamins), 13, 221–232. doi:10.1075/avt.13.21red

Reinhart, T. (1981). Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica* 27, 53–94.

Schlenker, P. (2004). Context of Thought and Context of Utterance: A Note on Free Indirect Discourse and the Historical Present. *Mind Lang.* 19 (3), 279–304. doi:10.1111/j.1468-0017.2004.00259.x

Schlenker, P. (2018). Gesture Projection and Cosuppositions. *Linguist Philos.* 41 (3), 295–365. doi:10.1007/s10988-017-9225-8

Schumacher, P. B., Dangl, M., and Uzun, E. (2016). *Thematic Role as Prominence Cue during Pronoun Resolution in German*. in *Empirical Perspectives on Anaphora Resolution*. Editors A. Holler and K. Suckow (Berlin: de Gruyter), 213–240.

Schumacher, P. B. L., Roberts, L., and Järvikivi, J. (2017). Agentivity drives real-time pronoun resolution: Evidence from German er and der. *Lingua* 185, 25–41. doi:10.1016/j.lingua.2016.07.004

Sells, P. (1987). Aspects of Logophoricity. *Linguistic Inq.* 18, 445–479.

Sharvit, Y. (2008). The Puzzle of Free Indirect Discourse. *Linguist Philos.* 31 (3), 353–395. doi:10.1007/s10988-008-9039-9

Stec, K. (2016). Visible Quotation. PhD thesis. University of Groningen.

Stec, K. (2012). Meaningful Shifts. *Gest* 12 (3), 327–360. doi:10.1075/gest.12.3.03ste

Stokke, A. (2020). Free Indirect Discourse in Non-fiction. *Frontier Commun.* doi:10.3389/fcomm.2020.606616

Stokke, A. (2013). Protagonist Projection. *Mind Lang.* 28 (2), 204–232. doi:10.1111/mila.12016

Sundaresan, S. (2012). Context and (Co)reference in the Syntax and its Interfaces. PhD thesis. University of Stuttgart. doi:10.1109/infcom.2012.6195464

Yashima, Y. (2015). Antilogophoricity: In Conspiracy with the Binding Theory. PhD thesis. University of California at Los Angeles (UCLA).

Zeman, S. (2019). "Wer Spricht? Disambiguierungsfaktoren bei der Perspektivensetzung im Narrativen Diskurs," in *Rede- und Gedankenwiedergabe in narrativen Strukturen—Ambiguitäten und Varianz, Linguistische Berichte, Sonderheft*. Editors S. Engelberg, C. Fortmann, and I. Rapp (Hamburg: Helmut Buske Verlag), 27, 221–251.

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

frontiers in Communication

HYPOTHESIS AND THEORY
published: 30 June 2021
doi: 10.3389/fcomm.2021.624334

Fictive Deixis, Direct Discourse, and Viewpoint Networks

*Barbara Dancygier**

University of British Columbia, Department of English Language and Literatures, Vancouver, BC, Canada

This paper proposes a renewed and more textured understanding of the relation between deixis and direct discourse, grounded in a broader range of genres and reflecting contemporary multimodal usage. I re-consider the phenomena covered by the concept of deixis in connection to the speech situation, and, by extension, to the category of Direct Discourse, in its various functions. I propose an understanding of Direct Discourse as a construction which is a correlate of Deictic Ground. Relying on Mental Spaces Theory and the apparatus it makes available for a close analysis of viewpoint networks, I analyze examples from a range of discourse genres - textual, visual and multimodal, such as literature, political campaigns, internet memes and storefront signs. These discourse contexts use Direct Discourse Constructions but usually lack a fully profiled Deictic Ground. I propose that in such cases the Deictic Ground is not a pre-existing conceptual structure, but rather is set up ad hoc to construe non-standard uses of Direct Discourse–I refer to such construals as Fictive Deictic Grounds. In that context, I propose a re-consideration of the concept of Direct Discourse, to explain its tight correlation with the concept of deixis. I also argue for a treatment of Deictic Ground as a composite structure, which may not be fully profiled in each case, while participating in the construction of viewpoint configurations.

Keywords: direct discourse1, deixis2, constructions3, viewpoint4, viewpoint networks5, literature6, internet memes7, multimodal artifacts8

OPEN ACCESS

Edited by:
Emar Maier,
University of Groningen, Netherlands

Reviewed by:
Kobie van Krieken,
Radboud University Nijmegen, Netherlands
Gisela Redeker,
University of Groningen, Netherlands

***Correspondence:**
Barbara Dancygier
barbara.dancygier@ubc.ca

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *31 October 2020*
Accepted: *09 June 2021*
Published: *30 June 2021*

Citation:
Dancygier B (2021) Fictive Deixis, Direct Discourse, and Viewpoint Networks. Front. Commun. 6:624334. doi: 10.3389/fcomm.2021.624334

INTRODUCTION

This paper argues for the need to recognize the concept of Fictive Deictic Ground, to account for the uses of discourse in communicative contexts other than natural spoken conversation. In a standard situation, the deictic center is the contextually determined pre-condition for spoken communication–the speaker and the hearer need to share deictic space and time in order to engage in a conversation. The original formulation (Bühler, 1990 [1934/1984]) further uses the shared deictic context as an explanation of the meaning of expressions such as *I, here*, or *now*–the referents of these are determined deictically, in contrast to anaphoric usage. Importantly, Bühler's approach does not automatically represent the Hearer *you*, or the resulting view of Deictic Ground as the site of conversational discourse. Research on deictic expressions has added a number of theoretical and cross-linguistic observations (e.g., Levinson, 2008; Fillmore, 1997; Fillmore, 1982) and deixis has remained one of the core concepts in pragmatics. Recent work, however, has expanded the scope of the enquiry–adding the discussion of joint attention (Tomasello, 1995; see also Turner et al., 2019 for a discussion of Blended Classic Joint Attention–in the context of TV, film, etc.), and numerous studies of gesture and eye gaze (e.g., Stukenbrock, 2014; Stukenbrock, 2020), as ways in which the speaker and the hearer use their bodies to make joint attention

and communication possible[1]. Also, deixis is often talked about in terms of 'grounding', or Deictic Ground (Langacker, 1987; Langacker, 1991; Hanks, 1990; Brisard, 2012), to account for the numerous ways in which the use of deixis goes beyond just 'being there' for communication to happen, but rather being actively used to construe situations. I will follow Hanks in this respect and use the term Deictic Ground as a flexible construct relied on by communicators, and, as Cornish (2011) observes, also used to set up the subjective viewpoint or perspective which allows the discourse to be construed in a specific way. In what follows I introduce the concept of Fictive Deixis, to account for the cases where basic elements of deixis are missing or re-construed. To explain the phenomenon, I connect the general understanding of Deictic Ground to the use of constructional forms of spoken discourse.

The relation between Deictic Ground and spoken communication receives varied amounts of attention in various approaches. Scholars closer to Bühler follow him in the focus on 'pointing', which includes more work on gesture, eye gaze, and other embodied means of achieving joint attention. The focus on deixis-as-'pointing' is also clear in the rich literature on demonstratives (which includes a very recent Research Topic in *Frontiers in Psychology* [2020, Vol 11 https://www.frontiersin.org/research-topics/10557/demonstratives-deictic-pointing-and-the-conceptualization-of-space]). Much less attention (except the work by Fillmore (1997) and Fillmore (1982)) is given to the fact that deixis is at least partly defined by the inclusion of the speaker *I* and the hearer *you* and relies on discourse in complex ways. In this paper, I focus on the understanding of deixis as a 'speech situation', and I propose that it should further be considered in connection to the cluster of constructions known as Direct Discourse, which share the grammar of spoken communication, but appear in contexts other than natural discourse–such as narrative discourse. The role of deixis in the choices of grammatical forms of Speech and Thought Representation in narratives has been given much attention–especially in work by Duchan et al. (1995), Sanders and Redeker (1996) Vandelanotte (2004) and Vandelanotte (2009). In what follows, I will also look at examples from narratives, but only to the degree that they illustrate questions about the connection between deixis and spoken discourse.

In considering various contexts in which discourse is used against a non-typical Deictic Ground, this paper may evoke the concept of Deixis am Phantasma, as introduced by Bühler 1990 [1934/1984] and discussed from the semiotic perspective in West (2013). However, there is in fact little in common between Bühler's work and the argument presented here. Bühler foundational theory 1990 [1934/1984] is focused on pointing, such that the object pointed at may be displaced from the current place and time (being imagined, recalled from memory or a dream). The situations pointed at may have never materialized but may nevertheless be felt as vividly experienced mental images. In this paper, however, I focus on cases of deictic construals which do not rely on pointing, but on spoken discourse instead, and which evoke and construct a Fictive Deictic Ground to legitimize discourse rather than structure imaginary experience. In other words, I focus on the ways in which the concept of Deictic Ground participates in our understanding of Direct Discourse (outside of colloquial spoken contexts) and on the mutual dependence of Direct Discourse and Deictic Ground. The examples to be discussed below represent a number of communicative situations evoking new Deictic Grounds, rather than relying on existing ones.

In the remainder of the Introduction, I outline two of the theoretical concepts I will use: Viewpoint and Mental Spaces Theory (MST). In *Direct Discourse Construction and the Deictic Ground*, I further develop the approach to Direct Discourse; *The Use of Direct Discourse Construction in Literary Genres* discusses examples from literary texts, while *Multimodal Artifacts* focuses on internet memes and storefront signs. *Discourse Viewpoint* and *Final Comments* sections conclude the discussion.

The understanding of viewpoint in this study builds on several broad assumptions, which I summarize here. All these aspects of perspective-taking have been discussed, in application to various discourse types, in several collections of studies (Dancygier and Sweetser, 2012; Dancygier, Lu and Verhagen, 2016; Dancygier and Vandelanotte, 2017b; Vandelanotte and Dancygier 2017).

Viewpoint (or perspective) is here understood as **a mental alignment** expressed by a discourse participant, through one or (quite often) more of the following devices: the choice of a linguistic expression, a visual artifact, performance of a sound sequence (such as a tune or intonation pattern), gesture (hand gesture, shrug, eye-brow movement etc.), eye gaze, body posture, or mime. The alignment can focus on the experiential aspects of the basic scene assumed to be the locus of the exchange: location (direction of motion, or distance), what can be seen or heard from the location assumed, the relationship between the speaker and other participants, or the action currently being performed. In more complex instances a participant can align with a temporal perspective, an emotional angle, a humorous or ironic attitude, and an epistemic or evaluative stance. One of the most common themes in viewpoint research is an analysis of types of viewpoints (e.g., visual, enactive, epistemic or emotional) adopted by participants in an event which is narrated, rather than experienced firsthand. I will assume, though, that the difference between viewpoint expression in spontaneous conversation and in fictional narratives is due to the nature of the linguistic material, rather than to the nature of viewpoint as such.

It is typical of most artifacts that they rely on **multiple viewpoints**–multiplicity is the norm, not an exception (see Dancygier and Vandelanotte, 2016). But the many viewpoint construals available in any scene are not a loose collection–they form a **viewpoint configuration** (Dancygier, 2012; Dancygier, 2017; Dancygier and Vandelanotte, 2016; Dancygier and Vandelanotte, 2017a). Even the simplest of expressions imply a viewpoint configuration, rather than a single perspective of a single participant (though one aspect of the viewpoint structure

[1]I am referring here only to the work in which gesture and eye-gaze are discussed specifically in the context of deixis; there is a much larger body of work on gesture and eye-gaze focused on viewpoint management and other aspects of meaning construal (eg., Sweetser and Stec 2016; Brône et al., 2017). However, a broader discussion of these aspects of multimodality in interaction cannot be addressed in this paper.

may be more prominent). For example, a speaker describing their location in a room (as in *I am sitting at my desk*) is making available a range of embodied and visual viewpoint parameters. Deictic parameters (time, location, the first-person speaker *I*) participate in that viewpoint configuration, but they are not the only aspects of the viewpoint network. Just to give a few examples, the sentence further suggests the speaker's ability to reach objects on the desk, but not objects which the desk separates them from, their visual and enactive perspective such that they can see and interact with someone sitting across from them at the other side of the desk, that they can see the scene in front of them, but not behind them, etc.

If we change the tense (a grammaticalized deictic category), as in *I was sitting at my desk*, the sentence adds another layer of viewpoint to include the utterance about the past in the scope of the viewpoint parameters of the current speaker and their speech situation. The highest viewpoint level would further depend on the role the sentence plays in the discourse overall. If the speaker of the sentence is telling a story to an addressee, using distal forms (*So I was sitting at my desk when I heard the news*), the deictic viewpoint of the storyteller and the storylistener is higher than the experiential viewpoint of the participant depicted at the desk, hearing the news. But if the storyteller chooses proximal deixis (*this* as a discourse deictic, and proximal forms throughout) to tell the story (*So this is the story. I am sitting at my desk, and my radio is on. . .*), they bring the past scene and events up to the current deictic viewpoint of the story being told. At the most basic level, then, the **Deictic Ground**, with its participants (speaker and hearer), location (here) and time (now) provides the most rudimentary viewpoint configuration (Fillmore 1997; Fillmore, 1982). Deictic Ground is the site of a conversational exchange, in which the participants alternate taking the deictic role of speaker (I) and hearer (you), in the here and now. The interlocutors' shared understanding of what is or is not accessible is reflected in the use of proximal or distal indexical expressions such as *this/that* (Diessel, 2006; Dancygier, 2019).

In accounting for the multiplicity of viewpoint I rely on the theory of **Mental Spaces**–conceptual packets representing situations (real, remembered, imaginary, desired, etc.), established and manipulated as discourse progresses; [2] the start point of the network of spaces is the base space–which includes the actual Deictic Ground of an event or an exchange. While Mental Spaces Theory has proven to be especially useful in analyzing reference, it also provides a clear set of tools to describe viewpoint constructions (expressions of emotional and epistemic stance, construal of conditional and imagined situations, etc (cf. Fauconnier, 1994; Fauconnier, 1997; Fauconnier and Sweetser, 1996; Sanders and Redeker, 1996; Dancygier and Sweetser, 2005). To remain within the scope of my simplistic past tense example above (*I was sitting at my desk*), the Deictic Ground of the sentence requires that we assume the presence of a speaker, informing the listener about the prior-to-now event, taking place in a not-here room where the speaker's desk is located, and describing the speaker seated at the desk (then, not now). There are thus two mental spaces profiled–the base space when the sentence is uttered, and the Past space when the speaker sat at their desk for some time. At the same time, the base space (and its current Deictic Ground) provides a Viewpoint space, while the Past space is the Focus space. This is to say that the past situation and its Deictic Ground (as in the 'desk' example) is typically viewed as distal, from the perspective of the present communicative situation (and its Deictic Ground). The two spaces thus form a configuration, where the base space (now) is higher in the network than the Past space (then) discussed. As I have also shown, a speaker might shift away from the Ground correlated with the moment of speech and adopt the past situation as the current proximal Ground–that is, say something like *So this is the story. I am sitting at my desk,. . .*. Such configurations are unremarkable but exemplify a structure which makes switching the Ground to a different space possible and thus can yield new viewpoint effects. The mental space configurations of viewpoint networks in the three 'desk' cases (present, past, and present-as-past) are represented in **Figure 1**.

DIRECT DISCOURSE CONSTRUCTION AND THE DEICTIC GROUND

Typically, the category of Direct Discourse (DD) applies to extended narratives, where what characters say or think can be represented in three different ways: as Direct Discourse (a not-necessarily-genuine quotation), as Indirect Discourse (reported by the narrator or another character), or as Free Indirect Discourse (which is more faithful to the assumed discourse of the character, but still adjusts grammatical forms such as tense and personal pronouns to the higher viewpoint from which discourse is being reported). DD is understood in terms of the default Deictic Ground and uses appropriate proximal forms. There is no Direct Discourse without an assumed Deictic Ground, and the Deictic Ground is a prerequisite for Direct Discourse. This correlation is used below to argue that the use of Direct Discourse cannot be separated from the Deictic Ground forming the base of a network of mental spaces. Using the form of Direct Discourse assumes a viewpoint structure such that there is (at a minimum) a base mental space determining the discourse participants, time, and location. I will refer to such a mental space as **Direct Discourse Ground** (DDG). It combines deictic elements (*I, you, here, now*) with other communicative affordances, which include the use of spoken language first of all, but also using one's body to gesture, regulate joint attention with eye-gaze, etc. Importantly, a DDG can be embedded in higher DDGs (e.g., in a fictional narrative structure), and have lower level DDGs (other conversations or events reported during the base conversation) embedded in it. The viewpoint of each DDG contributes to the viewpoint network of a broader discourse structure (narrative, conversational, etc.). The need for such a multilevel understanding of discourse is further clarified by work on

[2]It might be worth pointing out here that some of the discussion of deixis to be found in Bühler 1990 [1934/1984] could quite naturally be reconsidered and further specified in terms of MST.

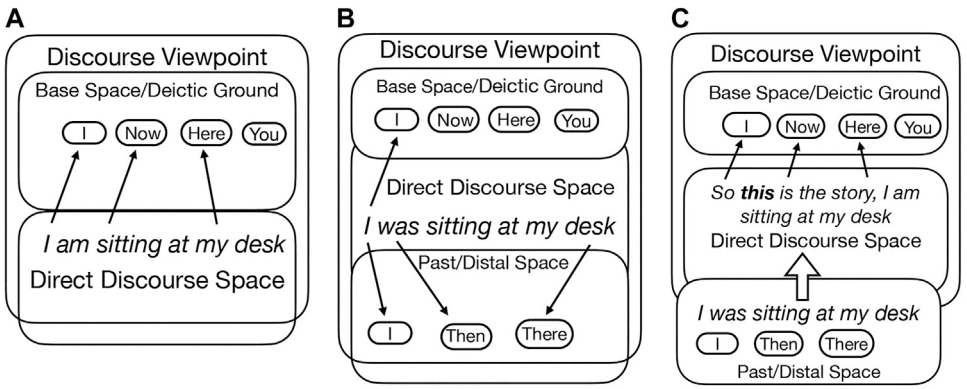


FIGURE 1 | (A) I am sitting at my desk. **(B)** I was sitting at my desk. **(C)** So this is the story. I am sitting at my desk, . . .

sign language (cf. Dudis, 2004, on body partitioning, Janzen, 2004 on ASL).

Direct Discourse Construction

In the context of Direct Discourse Ground, actual forms of DD are unique in that they are unrestricted constructionally–all that is required is the use of quotation marks, to signal the switch from the mental space currently being developed to a lower-level conversation space within it. In comparison, Indirect Discourse (ID) and Free Indirect Discourse (FID) both have formal constructional features which restrict the choice of forms (e.g., the embedded clause in Indirect Discourse undergoes a tense, pronoun and adverb shift, going, for example, from the assumed *I will finish the paper tomorrow* to indirect *She said she would finish the paper the next day*). However, in spite of its openness, Direct Discourse should also be treated as a construction. Vandelanotte (2004) shows convincingly that in spite of the use of forms correlated with the selected Deictic Ground, such choices in the narrative constitute a constructionally determined shift away from the Deictic Ground of the narrative flow as a whole. Besides, the three reporting forms (DD, ID and FID) together constitute a constructional cluster, where formal choices signal deictic concepts such as speakerhood in construction-appropriate ways (see Sanders and Redeker, 1996; Vandelanotte, 2009). I will therefore refer here to Direct Discourse Construction (DDC) and its various uses.

Every narrative sets up a number of mental spaces (more specifically, narrative spaces, cf. Dancygier, 2012). These spaces are inhabited by participants (characters) and occasionally represent conversations between these participants. When such a conversation becomes a part of what the narrative constructs, the two most likely options are Direct Discourse Construction (DDC), which shifts the viewpoint to the Deictic Ground of the scene in which the conversation occurs, and Indirect Discourse Construction (IDC), which embeds the conversational Ground in the higher narrative space. For example, if one character says to another *I have to go now*, the DDC representation would be *"I have to go now", she said*, while the IDC would be rendered as *She said she had to go right away*, where the expression *she said* is part of the narrative flow in the third person past tense narrative, and also the Viewpoint space from which the actual words of the character are represented. The constructional shift from DDC to IDC (which moves the DDC space into an embedded status with respect to the IDC *she said*-space) is marked by changing the forms appropriate to the proximal Deictic Ground (*I, Present Tense, now*) into distal forms of the embedded Ground (*she, Past Tense, right away*). Additionally, the deictic verb *go* suggests that the speaker and the listener share the current location (here), but there is no such assumption of participant proximity in the IDC version, since the speaker/narrator using the *she said* form is aligned with her/his own Ground, in the higher (Viewpoint) space in which the reported situation (Focus space) is embedded[3].

In the default set-up, DDC is used in correlation with the Deictic Ground in which the conversation happens (and thus forms a DDG). However, speakers may choose a different Deictic Ground as the backdrop to at least some parts of the conversation. Rubba (1996) describes how speakers may use proximal deictic words such as *here* to refer to a community they mentally align with, rather than to the current location. Rubba refers to such distal Ground which is talked about as if it were proximal as 'Alternate Ground'. What is important about such alternate Grounds is that deictic terms may be used to align the speaker with mentally salient spaces, and not necessarily with immediately accessible spatial and temporal spaces; spaces evoked in this manner are not imaginary and can be 'pointed at' or marked as proximal. This is made possible by embedding the spatially distal space in the current DDG, and assuming a proximal viewpoint in rendering it. Such distal-to-proximal shifts are driven strictly by viewpoint shifts and alter the overall viewpoint network being elaborated. Somewhat similar use of deictic forms has been described by Hanks (1990). As

[3]I will deliberately skip the discussion of Free Indirect Discourse, which blends the two Deictic Grounds. The workings of the distal versus proximal choices are more complex in FID, and it is worth a separate discussion, but it would not add much to my argument.

the examples throughout this paper suggest, DDGs are subject to various viewpoint shifts and embeddings and there is a wide range of such cases, in creative contexts, but also in various ordinary situations.

The point I argue for in the remainder of this paper is that DDC should not be seen solely in terms of typical sentences, spoken against the background of a Deictic Ground. Instead, we should consider how DDC emerges in various discourse contexts, how it fits into the viewpoint network of the discourse, and how its basic deictic parameters participate in the interaction. One of the assumptions guiding the analysis is that the distinction between 'real' and 'imaginary' deixis does not allow us to interpret discourse with sufficient granularity. Tying up deixis with DDC (to focus on DDGs instead of separating deixis from direct speech) gives us an opportunity to go beyond sentential and/or gestural elements and understand the mechanisms and processes involved in the construction of non-typical uses of discourse and to interpret viewpoint networks with more accuracy. Below, I consider two types of examples where DDC is used in ways deviating from standard conversational DDGs: literary discourse and multimodal artifacts. Before discussing the two communicative genres, I will briefly review the approaches to the use of DDC as an other-than-literal correlate of a fully determined Deictic Ground.

What does Direct Discourse Construction do

Many examples suggest that not every use of the Direct Discourse Construction is correlated with a fully profiled Deictic Ground. Also, while in most contexts the Deictic Ground is what makes Direct Discourse possible (as a result of the speaker and hearer roles being profiled), there are cases (which I discuss below) where Direct Discourse is used *so that a Deictic Ground can emerge*, also when crucial aspects of a default Deictic Ground are missing. Such instances confirm that the correlation between DD and DG creates various discourse affordances.

There has been some discussion of how an utterance structured as DDC can be used to signal meanings other than the default representation of speech in a Deictic Ground (Clark and Gerrig 1990; Pascual, 2006; Pascual, 2014; Pascual and Sandler 2016). The shared focus of these analyses are examples of sentences which are structured as DDC, while remaining independent of the Deictic Ground of the surrounding discourse. Examples come from written and spoken discourse.

In a written text, Direct Discourse is often represented in quotation marks (or can be seen as a quotation even if the markers of a shift to DDC are missing)–the separate Deictic Ground is thus signaled through a written convention; in spoken discourse speakers may mark the quotation with a gesture representing the scare quotes. Thus, in their genuine use, quotation marks (in writing or in gesture) represent a switch into and out of the DDG that the current speaker (or narrator) presents as not aligned with the default Ground of current discourse. This is, however, not as clear as it seems. In spoken context, if the speaker interrupts the flow of discourse to signal a switch to a different DDG, by gesture, a pause, or change of tone of voice (see Clark, 2016 for a full overview of such usage), the discourse included inside the quoted fragment signals that the speaker says something from a perspective other than their own. In fiction, something similar happens, as the narrator yields the Ground to a character or characters in conversation. Overall, quotation status consistently marks a shift to a different level of discourse, with the overarching discourse viewpoint allowing for a consistent structuring of viewpoints, depending on the genre.

Importantly, the reasons why a string of discourse is placed in quotation marks may not be restricted to a simple embedding of a piece of discourse which faithfully (verbatim) represents what was actually said in the situation described. In their now classic article, Clark and Gerrig (1990) argue that quotations do not represent authentic discourse, but that they serve as **demonstrations**. They claim that "The prototypical quotation is a demonstration of what a person did in saying something" (1990:769). Quotations are thus thought to demonstrate an act, rather than represent speech. This approach was further extended to a broader theory of depictions in Clark (2016)–the general point being that many communicative forms (such as gesture, vocal imitation, etc.) do not 'describe' anything, but rather 'demonstrate' or 'depict'. What is particularly important in the 'depiction' approach is its broad scope, but also its assumption of a special status of 'quotations'–which we can assume refers to specifically marked uses of DDC. The issue of 'faithfulness' of DDC was also taken up in Short et al. (2002). They argue that rejecting any 'faithful' value of DDC (which is Clark and Gerrig's point) is an overstatement and suggest that faithfulness should be textured in order to refer to various types of discourse.

Furthermore, Direct Discourse has been approached recently from the perspective of its possible fictive nature (Pascual 2006; Pascual, 2014; Pascual and Sandler 2016). The approach assumed in the '*fictive interaction*' work points to a broad range of uses, such as the fictive use of verbs of communication (as in *What does that tell you? Her behavior speaks for itself*) as well as textual insertion of discourse snippets ("*Any questions? Call us*", "the attitude of *yes, I can do it*"). Overall, the suggestion is that we naturally conceptualize attitudes or experience in dialogic terms. The 'fictive' aspect of such expressions is that even though they rely on verbs of communication or are represented as unattached discourse fragments, no actual conversation is implied to have taken place. The semantic mechanism whereby spoken discourse demonstrates or represents attitudes and emotional responses requires clarification–in what follows, I will refer to such cases as examples of metonymy. Importantly, the fictive utterances inserted in discourse do not lose their grammatical structure and are inserted without adjustments. Also, while expressions such as *yes, I can do it* are used metonymically to represent attitudes, they do not fit the understanding of being 'imaginary'. They are not immersed in any unreal DDG and do not require being seen in terms of DDC. The DG parameters are simply not profiled at all.

The approaches briefly mentioned here are relevant to my examples, in that they show DDC forms used beyond 'faithful' representation of discourse connected to a Deictic Ground. DDC

does not have any formal correlates of being embedded in discourse in ways that would support its interpretation–even when it is syntactically or morphologically embedded, as in many examples Pascual mentions (2006, 2014). This provides the grounds for the communicative effect of pretend-quotations being metonymic tokens of specific types of speech acts or communicative acts (so a phrase such as *Yes, I can do it* stands for a positive and determined attitude to a challenging task). Such uses do not rely on any fully profiled DDG and evoke instead any and all DDGs where such an expression of an attitude would be appropriate.

Importantly, a similar effect can be achieved in a more structured deictic context. When Barack Obama was running for President in 2008, his primary slogan was *Yes we can!*–used in campaign materials and also repeated in his speeches. The form of the slogan evokes a spoken exchange in which someone may doubt whether true change is possible (something like *Can we?* (*achieve what we want*)). Obama's use of *we* includes his followers in the attitude, while the phrase as a whole is a (non-fictive!) response to a fictive question suggesting 'doubt'. The answer comes from a specific subjectivity–the candidate himself, and so at least the speaker role in this Deictic Ground is filled–while the *Yes I can do it* generic attitude in Pascual's examples does not profile anyone in particular as the speaker.

What appears to be the case, then, is that the categories of 'demonstration' and 'fictive interaction' both rely on shorter or longer strings of DDC which tacitly *evoke* a Deictic Ground or a full DDG. However, the Ground may fill only some of the four deictic roles (speaker, hearer, time, and place). The more roles are filled, the closer the expression is to a genuine use of DDG and DDC, in a recognizable communicative context. But the fewer roles are filled, the smaller the possibility of a genuine use of DDC and the stronger the indication of metonymic evocation. What specifically is evoked depends on the type of expression, its emotional load, and the discourse context. Some aspects of these types of uses of DDC are thus worthy of note. First, the form of DDC may appear without a properly construed DG. There are various degrees of how much of DG remains unprofiled; as an extreme case, the phrase *the attitude of yes, I can do it* does not profile any of the usual deictic parameters, relying instead on the metonymic emotional value of the phrase. And yet, the emotional viewpoint expressed via the form used is easily interpretable because, as listeners, we create a set of possible DGs and contexts where the generic DDC *yes, I can do it* would signal the viewpoint intended. The DDGs evoked are not imaginary in any sense. Rather, they make it possible for the phrase to 'demonstrate' the attitude in question.

The two very similar expressions (*Yes, we can!* and *the attitude of Yes, I can do it!*) prompt viewpoint networks of different nature and complexity. Obama's slogan is licensed by a generic DDG, wherein he addresses voters to prompt the shared viewpoint of 'determination'. The second example establishes a pattern that I will elaborate on in the remainder of this paper: an instance of DDC which is not aligned with any DG and thus needs to set up a Fictive DG, where there is a speaker using a phrase that metonymically evokes an attitude.

This is still different from inserting the sentence into an IDC (*She was determined that she could do it*) where the embedding of the DDC space in the higher narrative space binds all the Deictic Ground elements to a higher narrative space which inherits deictic material from a still higher narrative space. The viewpoint of the higher space precludes reading the lower space as purely metonymic and not 'faithfully' representing the discourse. What this suggests is that the issue of 'faithfulness' of quotations may not be a matter of the type of text (as Short et al., 2002 suggest), but rather should be seen in the context of the viewpoint and deictic structure provided by a higher space. In other words, the narrator can be 'trusted' to report what 'she said', so the assumption of faithfulness is easier to accept. But in DDC, there is no such assumption of viewpoint projection from the narrative into the lower space. And it is even more clear in contexts such as fictive interaction, where the existence of a higher viewpoint space is overtly denied. We can find many more contexts in which the phrase would continue to represent conviction and determination, but the important observation is that each such instance would represent a different viewpoint network, and that the complexity of the interpretation would depend on the viewpoint spaces that would need to be set up.

In the next sections of this paper, I look at two specific (and very different) contexts, to show the crucial role viewpoint configurations play in how Deixis and Direct Discourse Construction are to be understood. I will look at literary discourse and multimodal discourse, to show the role of Direct Discourse in establishing (rather than just fitting into) its Deictic Ground.

THE USE OF DIRECT DISCOURSE CONSTRUCTION IN LITERARY GENRES

Literary discourse depends to a large degree on the use of DDC, though literary genres use it differently, with different assumptions and goals. A proper discussion of deixis in literary discourse requires a separate paper or book, but the examples below reinforce the points made so far. Literary examples of DDC are numerous, and sometimes complex, so a full discussion is beyond the limits of this paper. In earlier work (Dancygier, 2012), I discussed a number of options, but here I focus on the examples which best represent the correlation between DG and DDC.

Deictic Ground in Novelistic Prose

Dialogue (longer chains of DDCs) plays an important role in novels. In each case there are characters, well identified on the basis of the novel as a whole, communicating from the perspective of their own participation in the events of the plot. The special nature of such dialogues manifests itself on three levels. Firstly, they participate in a fictional story, and so they are embedded in narrative spaces constructed by the author and delivered by a narrator; they refer only to the fictional reality of the novel. Secondly, they are usually not represented in their (assumed)

entirety, so that just the content relevant to the story is represented, and some turns can be missed. Thirdly, the uninterrupted flow of discourse is quite different from how natural colloquial conversations would be conducted.

Importantly, novelistic dialogue is a good example of DDG. As I argued in earlier work (Dancygier, 2012), a fictional narrative sets up a Deictic Ground by virtue of relying on two important subjectivities: the narrator and the reader. Time and space are not profiled in such a communicative set-up. While it is true that contemporary novels experiment with such a frame (multiple narrators, fragmented narratives etc.), early novels typically profile a narrator addressing the reader directly–which confirms the underlying deictic set-up. Within that set-up, any dialogic part (DDC), regardless of its form, is adding to the overall higher viewpoint of the novel. The simplification of novelistic dialogue in the ways mentioned above is possible and useful because the ultimate value of DDC is adding to the story as a whole.

A brief illustration in (1) is a conversation from Dave Eggers' novel *A Heartbreaking Work of Staggering Genius*. It is a conversation between two brothers–Toph, a middle school kid, and Dave, his older brother and guardian. They are orphans, living on their own. In the episode, Toph comes back from school:

1) "What happened today?" I ask.
"Today Matthew told me that he hopes that you and Beth are in a plane and that the plane crashes and that you both die just like Mom and Dad."
"They didn't die in a plane crash."
"That's what I said."

This looks like a rather ordinary conversation: turns are taken, other conversations are reported, etc. The fragment profiles a full DDG (conversation, speaker, addressee, place and time), and includes several instances of Speech and Thought Representation (*Matthew told me . . . , he hopes . . . , I said . . .*). However, this deictically complete conversational scene continues in (2), where there is no specified deictic ground, the addressee is generic, and the lines themselves demonstrate a generic attitude.

2) Sometimes I call the parents of Toph's classmates.
"Yeah, that's what he said," I say.
"It's hard enough, you know," I say.
"No, he's okay," I continue, pouring it on this incompetent moron who raised a twisted boy. "I just don't know why Matthew would say that. I mean, why do you suppose your son wants Beth and me to die in a plane crash?" (*AHWOSG*, p. 89)

The point in (1), especially the reported conversation, is to give the reader an understanding of how other children treat Toph at school–it sets up the frame of Toph being bullied. There may have been similar conversations, and so the words do not represent one unique instance of bullying, though they are formally immersed in a specific Deictic Ground. In (2), the dialogue switches to a generic mode by relying on present tense (*I say, I continue*, in the 'repeated' or 'generic' sense of the verb form) and on the use of *Sometimes* (confirming a repeated pattern, not a unique instance). Example 2) consists exclusively of lines representing Dave manipulating various parents into feeling guilty, while the 'parent' lines are missed as irrelevant. Importantly, each of the lines sets up a Fictive DG, to give the grounding to DDC. The Fictive DG is incomplete (no addressee, specified time, or space), but it plays a role of elaboration on the topic of 'school bullies'. Together, the lines in (1) and (2) demonstrate an aspect of Toph's school experience and not any specific conversation. **Figure 2** shows how the generic discourse ground yields the meaning intended–communicating to the reader that Toph's school experience involves bullying.

The possibilities for specific roles DDC can play in a novel are almost endless. But there is a shared goal, which is focused on the DDG's contribution to the viewpoint structure of the text as a whole. And Fictive DGs play an important role in maintaining the viewpoint structure.

Direct Discourse Construction in Contemporary Poetry

We typically do not think of poetry in deictic terms, and yet contemporary poetry relies on lines from conversations very often. It is enough to consider a poem such as *The Applicant*, by Sylvia Plath, discussed in detail by Semino (1997) and Freeman (2005). Sequences of DDC, such as *. . . Stitches to show something's missing? No, no? Then/How can we give you a thing? Stop crying. Open your hand!/Empty? Empty [. . .]* use the whole constructional variety of DDC forms, but they do not build off of a deictic Ground set up earlier in the text (as narrative fiction does)–rather, they *evoke* the Fictive (and incomplete) Deictic Ground *by using DDC*. The process is thus reversed, and what ties the discourse together is its insertion in specific frames (in the case of *The Applicant*, an interview and a sales pitch). Constructing an appropriate frame

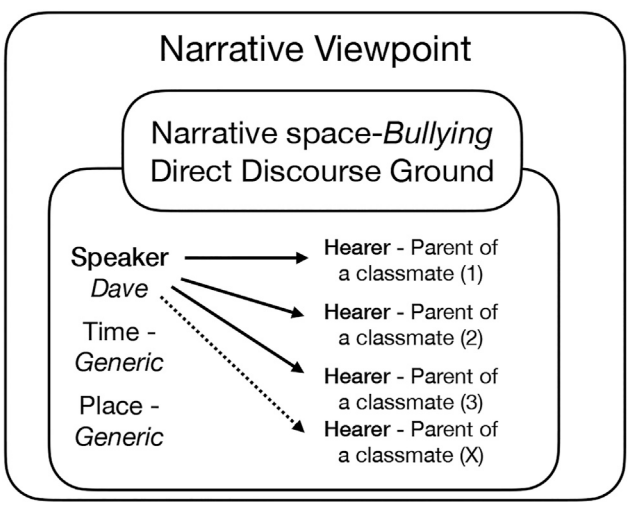


FIGURE 2 | Multiple hearers.

against which DDC lines can be understood is crucial to poetry of this kind.

It is even more visible in a poem like *Funeral*, by Wisława Szymborska, which consists entirely of lines of discourse (so that there is no voice of the 'poetic subject' or 'poem's persona' represented anywhere). The poem is quite long, all written in the style represented in (3):

3) "so suddenly, who could have seen it coming"
"stress and smoking, I kept telling him""not bad, thanks, and you" "these flowers need to be unwrapped" [. . .]
"you were smart, you brought the only umbrella"
"so what if he was more talented than they were"
"no, it`s a walk-through room, Barbara won`t take it"
"of course, he was right, but that`s no excuse"
"with body work and paint, just guess how much"

The organizing frame is determined by the title–*Funeral*. The frame evokes family members or acquaintances who do not see each other often. It also sets up a Fictive Deictic Ground, with a specific place and time, where DDC engages a number of unspecified speakers and addressees. They all participate in conversations–about the deceased but also about various everyday matters. The multiple DDC lines uttered by unidentified speakers to unidentified addressees create an ironic viewpoint–people gathered to mourn are spending a lot of time catching up on gossip instead, and then swiftly disappear into their own lives. The poem constructs an overarching viewpoint on the basis of the shared time and space of these exchanges–it is natural to understand the poem as reflecting the event of the funeral ceremony from its beginning to its end. However, the multiplicity of unspecified speakers and addresees does not match the expected deictic format. Fictive Deictic Ground is needed here so that the disjointed DDC lines can form a DDG structure which gives rise to the ironic view of the event as a whole. The structure of such discourse is represented in **Figure 3**.

In the literary cases considered above the identities of all participants or the course of the conversations evoked are not central to the viewpoint constructed. The focus is on social situations and the emotional responses of discourse participants. In reading (2), we soon notice that bullying is not really a major concern for Dave or Toph; rather, the brothers respond in a manipulative way, shaming the parents. Similarly in (3), the representation of people engaged in inconsequential chatter in the context of someone's death creates an ironic viewpoint. Importantly, the use of DDC in literary texts can evoke and set up incomplete Fictive Deictic Grounds, to construct (or 'demonstrate') viewpoints needed in the interpretation, rather than faithfully report conversations. In the next section, I consider examples from drama–the literary genre which ostensibly depends entirely on conversations.

Drama

Drama is the literary genre which uses Direct Discourse exclusively, and by definition. Anything said on the stage is ostensibly addressed to someone else (see Dancygier, 2012 for a discussion of the various 'addressees'). However, the nature of dramatic discourse is much more complex than such a description might suggest. The specificity of dramatic discourse in comparison with spontaneous conversations can be described as follows: 1. Dramatic discourse relies on two major Mental Spaces: the story space, with characters conversing on the stage, and the audience space, populated by silent spectators/listeners; 2. Actors on the stage typically address each other (and not the audience), speaking the words of the characters represented; however, the audience is still the actual addressee; 3. The DDC status of all that is said on the stage is assumed, but the discourse on the stage does not fully comply with our expectations of what spoken discourse does (it can, for example, profile aspects of the narrative which are not directly acted-out on the stage).

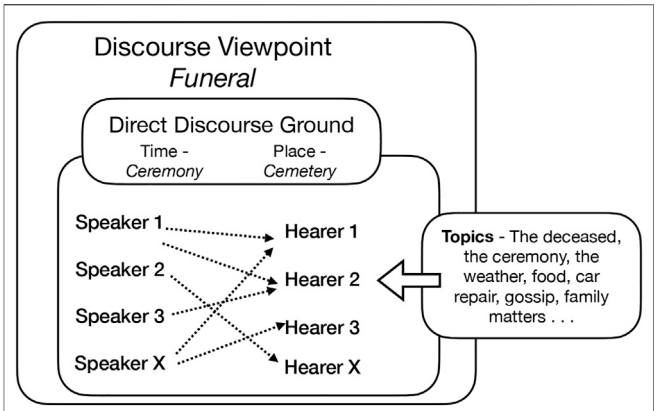


FIGURE 3 | Multiple speakers and hearers.

One of the important considerations of the discourse of drama is how it represents character's inner thoughts. In early forms (such as Shakespearean drama), characters often speak to the types of addressees that obviously cannot participate in conversations: objects, bodies, concepts or images, etc. For example, in *Romeo and Juliet*, such examples abound:

4) a. Come, gentle night, come, loving, black-brow'd night, . . . (Juliet speaks to the night)
b. Ah, dear Juliet./Why art thou yet so fair? (Romeo speaks to the body of Juliet)
c. Come bitter conduct, come unsavoury guide, . . . (to the poison he will drink)
d. Eyes, look your last, . . . (to his eyes, when he looks at Juliet for the last time).

A more thorough look at the discourse of Early Modern drama is beyond the scope of this paper, but what such examples make clear is that inner thoughts and feelings are expressed by addressing entities which are at the center of what the character feels (anticipation of the nightly arrival of the lover, surprise at the beauty of one's wife even though she is believed to be dead, expectation of relief that poison may bring, intensity of the final moment of parting, etc.). In other words, dramatic discourse sets up Fictive Deictic Grounds and DDCs with

improbable addressees (ones that cannot ever be speakers). It is worth noting that such discourse makes full use of the potential of both deixis and DDC–the deictic verb *come* is used in (4) to talk about an approaching experience (meeting the lover, dying) the speaker desires, the sentences can take imperative or interrogative forms, pronouns are consistent with the speaker or addressee status, etc. In earlier work (Dancygier, 2012), I have referred to such usage as the 'Vocative cum Imperative' Construction–a type of literary construction relying on some central features of DDC, identifying the atypical frame-relevant addressee by using a vocative form. What matters from the perspective of this paper is that lines of text structured like DDC are used in incomplete or outright impossible Fictive Deictic Grounds, to make communication of emotions possible via evocation of relevant concepts. But the evocation of Fictive DGs creates Direct Discourse Grounds (DDGs) which house viewpoints relevant to the story (secrecy, surprise, anticipation of death, etc.). Such Fictive DGs also allow for a close connection between embodied action on the stage and the words spoken (holding a poison vial, looking at Juliet, etc.). Example 4c is represented in more detail in **Figure 4**.

Literary discourse provides much material for analyzing the use of DDC, though I cannot expand the discussion here. But even these limited examples show how literary genres all rely on DDC and Fictive Deixis, profiling the deictic roles in ways appropriate to the genre. Narrative fiction builds a story by setting up complex configurations of mental/narrative spaces, with each of the spaces marking a perspective needed. It uses DDC mainly as 'demonstration', whether to represent character's speech in a specific narrative space, or generic behaviors and conversational patterns recuring in many narrative spaces. Poetry may use DDC in various ways (Dancygier and Vandelanotte, 2009) but also to evoke incomplete Fictive Deictic Grounds. Finally, drama may create specific DDC sub-constructions, to maintain the illusion of speech (against an evoked Fictive Deictic Ground) in order to communicate thoughts and feelings. Unlike the examples of fictive interaction discussed by Pascual, where there is no specific Deictic Ground referred to or evoked, literature requires the setting up of (Fictive) Deictic Grounds, so that DDC can be used, appropriately to the genre, and so that the emergent DDGs contribute to viewpoint networks.

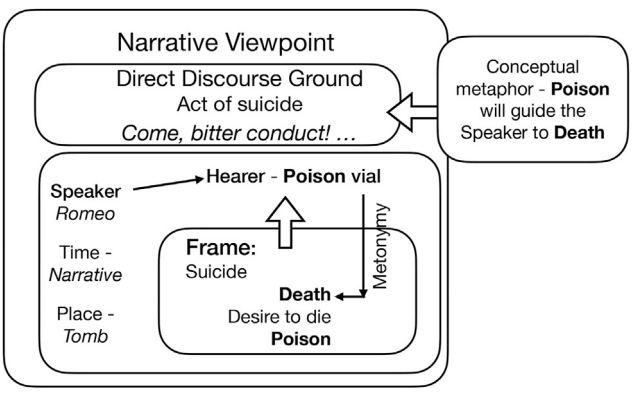


FIGURE 4 | An object as an addressee.

Given that many of the literary Deictic Grounds are incomplete, it is important to ask what aspects of deixis they profile. Judging by the examples above (and other examples I have gathered), the speaker role is the one most fully profiled. The addressee role is filled by various elements of the mental space topology, and often not given any voice at all–this is true in the Eggers example, where parents of school bullies are not profiled as participants, the addressees of gossip lines in *Funeral* are not identified at all, while addressees in *Romeo and Juliet* are material objects, body parts, or disembodied concepts (such as *night*). The speaker is given a privileged role in all these cases.

The analysis so far provides the material needed to explain the concept of Fictive Deixis. Spoken discourse (DDC) is naturally used (in various contexts) to represent attitudes. This is a natural extension of the role of DDC, because the spoken idiom is capable of expressing emotional reactions most efficiently. I follow Vandelanotte (2004) in his explanation of a clear connection between the very concept of DDC and deixis. The approach to deixis that I have built here (focusing on discourse consequences rather than 'pointing') allows for two cases: either the DG is set up by other means (as in an extended discourse of a political campaign or a novel), or it needs to be evoked to give legitimacy to DDC, and is thus Fictive–set up for the needs of specific discourse, even if there are no standard deictic dimensions available. But because such an evoked Fictive DG works with DDC to create a mental space with clear topology, the resulting DDG serves the needs of viewpoint construction without relying on all four parameters. The easiest parameters to omit are time and space, but the addressee is also often missing (see ex. 2) or replaced with non-sentient presences (objects, concepts, time of day, etc.). The deictic parameter that cannot be omitted is the speaker, to provide an entry to the viewpoint configuration of the text.

MULTIMODAL ARTIFACTS

The question of the role played by DDC and deixis in various communicative contexts becomes much more complex in artifacts using an image, or a specific design, and also language. I will look briefly at two examples of internet memes, to then consider the phenomenon that I will refer to as 'street deixis'.

Internet Memes

A genre of contemporary communication which makes a common and quite revealing use of DDC is internet memes. Memes are quite restricted formally, often using predetermined images, text slots, and phrases (Dancygier and Vandelanotte, 2017a). The way they make quick emergence of meaning possible is thus by infusing all formal aspects with maximum of meaning. One of the ways to achieve that goal is via pretend Direct Discourse, used in absence of a Deictic Ground. Examples are numerous, but I will restrict my attention to two representative cases.

Figure 5 shows one of the early memes (*Said no one ever*), which starts with a pretend quotation. The sentence quoted is not attributed to any specific speaker or situated in any Deictic Ground. In fact, it is used only to then be explicitly rejected as a possible DDC, by the *Said no one ever* phrase, communicating the meme-maker's rejection of the pretend-claim in quotes. The quotation in the meme thus evokes an attitude, somewhat similarly to the funeral conversations in (3), while constructing an ironic viewpoint (see Dancygier and Vandelanotte, 2017a for more discussion).

In **Figure 6**, one of the Scumbag Steve memes (which express complaints about various annoying misdemeanors), the quotation evokes some people's irresponsible attitude to the property of their hosts. We should also note that the grammatical structure of the meme has some unusual features: subject suppression (Ruppenhofer and Michaelis, 2010; David, 2016) and disjointed syntax, but also the placement of DDC in a context-free slot, where the viewer has to construct a Deictic Ground and the whole context for the question to make sense. There are thus two Fictive Deictic Grounds evoked: in the Top Text the meme-maker addresses the generic 'you', while the other is a fictive situation, not profiling time or space, in which a rude guest (someone like Scumbag Steve, represented in the image macro) uses DDC to refuse to make amends for damaging the host's property. Importantly, the Bottom Text question in quotation marks is addressed at the host imagined in the Top Text–the *you* addressee who had something damaged by a careless guest. The meme-viewer is thus a generic addressee of both texts, even though they are attributed to different speakers.

The roles of discourse participants (speaker and addressee) are not identified in these memes, but the presence of DDC calls for an emergence of a Fictive (generic) DG where the speaker and the addressee are profiled. The graded salience of the profiling is characteristic of Fictive DGs, and so the question in **Figure 6** can be seen as addressed to any generic person dealing with a rude guest.

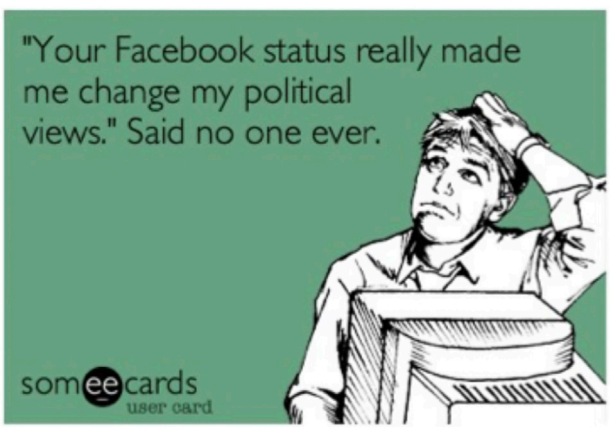


FIGURE 5 | An example of a *Said no one ever* meme.

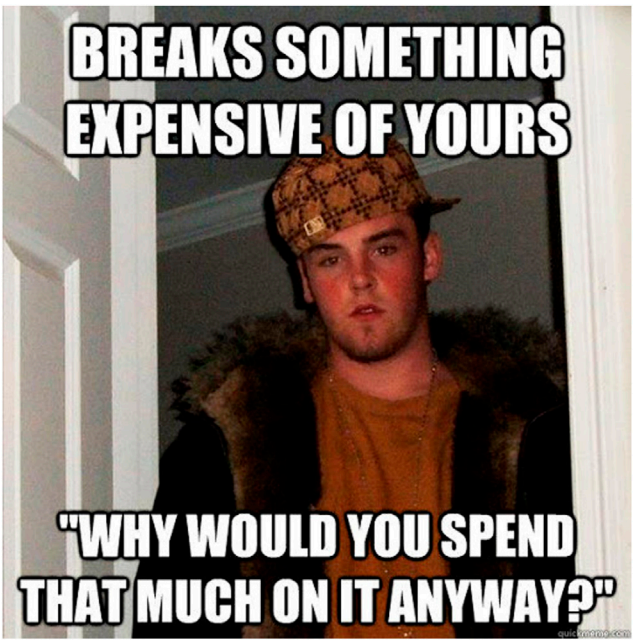


FIGURE 6 | An example of a Scumbag Steve meme.

The use of *you* and the way in which some memes involve the viewer as an addressee shows the importance of the form of memetic text in how the Deictic Ground is set up and used. The specific viewpoint networks of these (and other) memes rely on constructing a stance (such as approval/disapproval of people's beliefs and behaviors), and the viewer, even if addressed, is not a genuine participant in a DDC chain. Further confirmation of the specific nature of multimodal artifacts can be found in street signs–to be discussed in the next section.

Cityscape and 'Street Deixis'

As we walk the streets of any city, we are constantly bombarded with information–shop windows, street names, parking rules, etc. However, there seems to be a more recent emergence of complex artifacts displayed in windows and on doors. I will consider two kinds of artifacts–standard door or window signs, which can be purchased ready-made by a business owner, and some custom-made displays, related to the specific business. Overall, the signs represent another use of Fictive Deixis. Unlike the literary examples above, these signs do not rely on the surrounding text to provide a viewpoint structure, and so they construct a viewpoint on the basis of the actual location and its function.

Many small businesses rely on storefront signs to notify prospective customers about their business hours,[4] and also

[4]Business restrictions related to COVID-19 have prompted a major re-construal of storefront signs. There is on-going work on the patterns of form and meaning which have emerged as a result–the primary observation is that the storefront communication became much more personal and emotional. The work has also demonstrated the efficiency and flexibility of this form of communication (see Dancygier et al. forthcoming). (Feyaerts and Heyvaerts, 2021).



FIGURE 7 | Two examples of *Open/Closed* signs.



FIGURE 8 | A custom *Thank you* sign.

display a clear sign informing anyone approaching the shop whether it is open or closed. Such businesses typically refer to themselves as *we*. **Figure 7** shows two standard open/closed signs.

The signs use *we* to mark self-reference, the use of which Levinson (2008) has described as a weaker version of the *I* deixis. Additionally, the signs perform speech acts which imply a prior action by a prospective customer. *Yes We're Open* ostensibly answers a question, such as *Are you open?* Similarly, the invitation, issued with the use of the deictic verb *come (in)*, appears to respond to a customer who is hesitant whether the business is available. When the shop is closed, it apologizes to the willing customers who were not able to enter (*Sorry We're Closed*). Speech acts such as confirmation, invitation and apology are represented, but the actual performance only takes effect when a passer-by decides to stop and look at the sign displayed. In other words, the performance of the speech act relies on Fictive Deictic Ground: the speaker (*we*, the business) communicates something to a possible addressee, but the speech act is felicitously performed only when there is someone who passes the location of the store (*here*) at a specific time (*now*) and who is able and willing to become the recipient of the speech acts made possible by the sign. Still, even if the speech act is understood, it is not linguistically acknowledged (we do not quite imagine a passer-by reading *Sorry, We're Closed* and replying out loud *That's all right, I'll come back later*).

There are also signs performing the speech act of giving thanks. **Figure 8** represents such an example. The 'speaker' is the business, but it is now represented only by that sign, as it no longer sells flowers in the same location. The example instantiates one of the felicity conditions of the speech act of thanking–that the speaker can thank the addressee (metonymically referred to with the name of the neighborhood) for their action in the past mental space, where the act for which the speaker is expressing thanks has been performed. The surprising effect is that the relevant Fictive Deictic Ground is here used to enable the use of DDC between an absent business, a collective addressee, in a former location, and thanking the addressee now for things done in the past.

We should note that other 'thank you' acts have the power to set up mental spaces which are not reality spaces–rather, they are set up 'retroactively', to fulfill the felicity conditions of the act. This seems to be the case with most of the *Thank you for not smoking* signs–they prompt a setting-up of a non-factual mental space in the past, present or future, where the viewer of the sign refrains or has refrained from smoking. The addressee is the viewer again, regardless of their intentions and behavior, and the effect expected is to prevent them from smoking, so the thanks can become felicitous. A similar set-up applies in jocular interactions where a participant feels thanks should have been offered but weren't and says *Thank you* not to thank the hearer, but, in a sense, to put the thanks in the hearer's mouth retroactively (the tone of voice is then somewhat sarcastic). Overall, all those 'thank you' acts are made felicitous by setting up non-factual or topologically empty mental spaces (Fictive DGs) in which the gratitude-worthy acts exist.

The spatial stability of the business is what drives this kind of cityscape communication. However, there are also instances where signs displayed in cities orient the deictic center to match the location of the addressee. I am referring especially to *You are here* pointers on maps displayed in various cities (especially historic sites, where tourists might wander and feel lost). A skeletal *You are here* sign is given in **Figure 9**.

The central aspect of deixis in these examples is the location, but the construal of what counts as 'here' is quite complex. First of all, maps informing people where they are can be found in specific locations, and the role they play is to allow people exploring the city on foot to plan their itinerary to the point of interest that they want to visit. The construal thus starts with someone (tourist information office?) deciding where to locate helpful city maps, and then adding 'You are here' signs to mark the location of the map (not of a participant in a conversation!). At this stage no new information is provided to anyone, but the set-up has been created wherein the Deictic Ground is primarily structured by a location (similarly to the business location case discussed above). The people making such maps have to adopt the viewpoint of a person walking around an unfamiliar city and

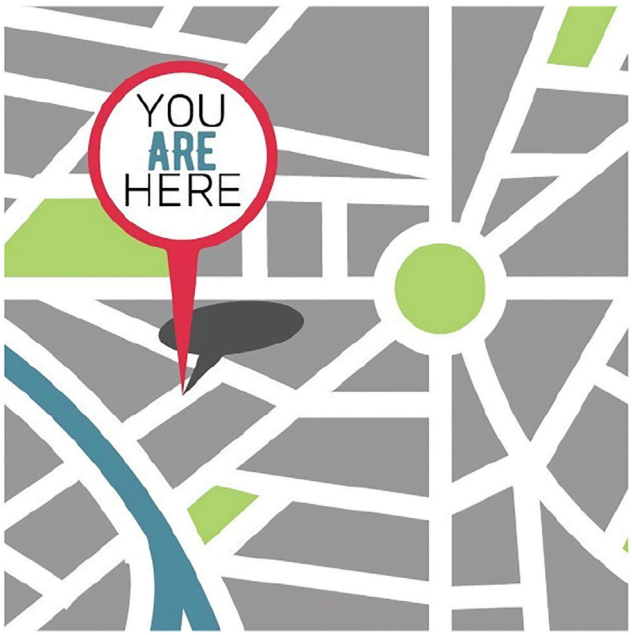


FIGURE 9 | An example of a *You are here* sign.

arriving at a map. Such a person needs to design an itinerary which starts at the location of the map. The main point, however, is that the map is a schematic visual representation of a reality of the city. When the marker points to a section of the map and says *you are here* it actually "says" something more complex. It asks the viewer to first match their real spatial location with the location of the map and the location of the *you are here* sign on it. The viewer needs to cross-map the location of the map in reality space with the spatial configuration represented by the map–in other words, to figure out the landscape surrounding the marker on the map, and then map it back onto their surroundings in the reality space. So it is not just a construction of two spaces, reality and its representation. It is a complex back and forth between the two spaces, where bits of reality have to be gradually cross-connected to the bits of the representation, in an effort to create a reliable spatial viewpoint. What is particularly interesting about such usage is the opportunistic emergence of the Fictive Deictic Ground when an unspecified passer-by stops to look at the map. The emergent DDG profiles the spatial viewpoint the passer-by needs.

Most of the signs looked at above are quite standard. My final example, however, in **Figure 10**, combines a number of important dimensions of street deixis[5].

The photo in **Figure 10** captures the window of a clinic in Vancouver which has undergone renovations. After the clinic opened for business again, they invited people with this line: *Relax, Vancouver, it's still us*. It is another example of DDC used in the context of 'street deixis', and it explicitly uses many of the features discussed above:

[5]Many thanks to Thomas McCullough for sharing this example with me.



FIGURE 10 | *Relax, Vancouver* sign.

- Imperative form (*Relax*): a bit unusually, the sign does not refer to ordinary business interaction. Instead, it speaks to the assumption that clients could be confused (because of the renewed look and appeal of the clinic). The 'confusion' frame is thus set up for the viewer to inhabit, whether they are actually confused or not.
- Addressee (*Vancouver*): similarly to *Kitsilano* above, this form refers metonymically to all inhabitants of Vancouver.
- Speaker (*us*): as usual, the clinic refers to itself in the first-person plural form.
- Location (*it*): though not strictly equivalent to *here*, this use of the pronoun seems to suggest 'the business you see here' rather than an anaphoric reference.
- Time (*still*): the time adverb *still* suggests continuation of a past state into the current situation (now)–this is, then, another case where a past mental space is being evoked. This usage refers to what a passer-by sees in the present, while reaffirming the continued identity of the clinic in spite of visual changes.

The structure of the artifact in **Figure 10** is diagrammed in **Figure 11**.

Overall, examples of street deixis rely on the interaction of several phenomena. First of all, they all rely on creating scenes of joint attention, where the passer-by is made to notice the business and the information indicated by the sign. In a sense, the sign deictically points to the relevant communication from the business, and the passer-by may respond. Secondly, the signs are all examples of Direct Discourse Construction, but without the discourse context that determines the deictic center that the discourse builds on. Broadly speaking, it depends on the assumed visual viewpoint of a person walking along a street. When a passer-by sees a written sign and decides to read it, the Fictive Deictic Ground is prompted, in which the passer-by receives the information provided, and thus becomes an addressee. At the same time, the text displayed is understood to have been provided by a 'speaker'–a business that needs to provide basic information

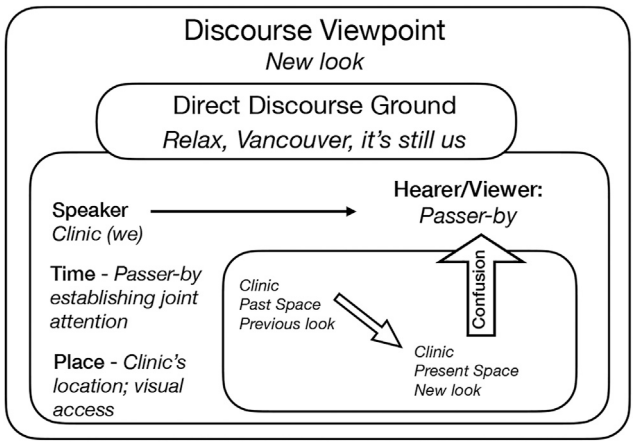


FIGURE 11 | The discourse structure of the *Relax, Vancouver* sign.

to prospective customers. Even though the text may have been displayed for a while, the Fictive Deictic Ground is triggered only when someone decides to stop by and read. The location is stable (wherever the sign is displayed), but the deictic 'now' is reset afresh for every subsequent reader/interlocutor. Importantly, as in other cases discussed, the presence of the decontextualized DDC prompts the setting-up of the Fictive DG. However, what makes these examples different is that the Fictive DG is formed in the context of real, material scenes of joint attention.

In addition to these basic deictic elements, the signs often rely on the deictic verb *come* or *come in*, asking those who have stopped to look to make the next step and enter the premises. Also, in some cases the addressee is described as a group, distinguished metonymically with reference to the location they inhabit (*Kitsilano, Vancouver*). The signs also perform several speech acts–invitation, apology, giving thanks, etc. and rely on the imperative form (which further implies the intended way to see the viewer as an addressee (*you*)). We should also note the specific use of time and space. The location of the sign becomes the deictic 'here' only when someone stops to read the sign; similarly, the time 'now' is the brief time of the interaction between a passer-by and a sign. However, past spaces can be evoked and constructed (rather than referenced) to make the speech acts felicitous.

The discourse potential of street deixis is quite limited–first of all because the speaker and the addressee do not engage in turn-taking. At the same time, the limitations of such communicative artifacts are not of the kind which would make them examples of fictive interaction. The information is communicated by someone and received by someone else, and there is an identifiable time and space of the communication–which is not included in the definition of fictive interaction. Also, the 'demonstration' status of storefront signs is not uniform, as some signs express a generalizable act of announcing that the store is open or closed, while other signs are more narrowly suited to the purpose, the time and the space. The specificity of such usage is thus due to the use of deixis. Finally, these examples would not fit well into the category of imaginary deixis, because nothing is left to the imagination here. The location, the time, and the presence of a sign are all Base-space, material elements. The information the sign provides is relevant to the nature of the business it represents. But all these parameters do not start forming a Deictic Ground until the information is received by an addressee. Consequently, street deixis examples start by missing an important parameter (an addressee) and become complete once the addressee enters the joint attention pattern and reads the sign. The DDC message remains on display without interruptions, but the Fictive DG is evoked only for the time of interaction between a passer-by and the sign. Also, it does not open an ordinary DDC format, with turn-taking patterns.

DISCOURSE VIEWPOINT

The range of examples discussed above considers the relationship between Direct Discourse and Deictic Ground in various contexts. I argued that the two phenomena are intricately connected and I have also shown how the Direct Discourse Construction needs to be seen against the foundation of a Deictic Ground–thus forming a more complex structure, the Direct Discourse Ground. To flesh out the way in which DDC and DG co-construct discourse, I have looked at three examples which represent non-standard uses of spoken discourse: literary genres, internet memes, and storefront signs. In each case I argued for an account of how DDC prompts a Fictive DG in absence of an ordinary deictic structure. One of the claims made in the analysis is that Deixis can be seen as a composite concept, such that certain uses do not fill all the parameters.

The question remains how these uses affect viewpoint networks. In my analysis of viewpoint in fictional narratives (Dancygier, 2012), I argued for the concept of Story Viewpoint Space–a narrative (mental) space which gives coherence to the narrative as a whole. Among other things, such a space determines the person and tense of the text, while also explaining how lower-level narrative spaces yield a coherent story governed by the Story Viewpoint Space. In later work, especially Dancygier and Vandelanotte (2016), where we clarified the nature of viewpoint networks in various discourse types, we expanded the concept of the top space governing the viewpoint constructed by the artifacts (including narratives), to propose the Discourse Viewpoint Space. The concept allows for a clear understanding of how discourse progresses by elaborating and texturing the viewpoint structure, and how listeners, readers, and viewers experience discourse as coherent in spite of multiple components and viewpoint shifts.

It is clear that all the artifacts and excerpts discussed above are primarily designed to create viewpointed construals of situations. The viewpoint can be aligned with a character (as in fiction or drama) or with a frame (poetry); it may involve commenting ironically on things people say or do, or engage the viewer in an unplanned interaction aimed at providing information and prompting action. The viewpoints communicated (emotional, ironic or persuasive) are constructed on the basis of setting up Fictive Deictic Grounds in contexts which do not set them up naturally (talking to the poison or to one's eyes, a 'talking' map, an

imagined confusion, a business which gives thanks, invites, or apologizes, etc.). They are Fictive Deictic Grounds and they are typically lacking some of the crucial elements of deixis, so the truncated deictic structures do not develop into natural conversations. Instead, they use various forms of DDC to construct DDGs for viewpoints–either by providing a mental space (in a novel or a play) wherein things relevant to the context can be said or by metonymically evoking emotions and attitudes. The role DDC strings play (fictional discourse or metonymic evocation) is determined by the network of viewpoints in which the DDC string is embedded.

Multimodal artifacts construct Discourse Viewpoint on the basis of visual representation spaces and discourse spaces set up. Memes further structure the viewpoint network by separating different discourse spaces (Top Text and Bottom Text) and construct their Discourse Viewpoint by resolving the contrasts set up–in the cases above, criticizing thoughtless statements or condemning rude behavior. Finally, examples of street deixis create a discourse set-up which may function as a speech act (invitation, apology, giving thanks), help structure the experience of the cityscape, etc.

The construction of Discourse Viewpoint is the ultimate goal of such discourses. The artifacts described here (as well as many other ones within the genres discussed and in other discourse contexts) start out by using DDC without its full, proper context, and in absence of a fully profiled DG. To complete at least the minimal deictic parameters required to establish the DDG for Discourse Viewpoint, a Fictive Deictic Ground is established.

FINAL COMMENTS

I want to close the discussion by pointing out some of the conditions that make the usage described possible, but also interesting. First, why can DDC play so many roles, in spite of its primary function? I argue that the metonymic evocation function of DDC is possible only when the discourse strings used clearly suggest a type of situation and the discourse appropriate to that situation. In other words, not every string of DDC can do the evocation work in the context. This is especially true in the context of a multimodal artifact, where brevity and lack of context make discourse build-up impossible. The examples above support the approach to quotations that treats them as 'demonstrations', but requires a more textured approach, depending on the nature of the Deictic Ground in various cases. There are important differences across the discourse types exemplified above.

Examples of street deixis create scenes of joint attention (quite similarly to demonstratives, see Diessel, 2006) by using DDC strings which put the person walking by in the role of a conversation participant–which other artifacts considered do not do, even when, like memes, they use *you* as a way to engage the viewer. Because the cases of street deixis are crucially dependent on the proximal deictic understanding of space and time, and on engaging the gaze of a passer-by, they are in fact the closest to a default deictic set-up. However, a Fictive DG does not lead to a conversation, to the switching of conversational roles, or a reference to distal phenomena. What the storefront sign communicates, at the moment when attention is achieved, does not lead the recipient of the message to respond to it verbally, but it may prompt the passer-by to respond by acting–e.g. entering the business, returning at a different time, or designing an itinerary suggested by the map. Such cases make it clear that what the Fictive Deictic Ground makes possible in these cases is not only communication, but also action. Importantly, the multimodal nature of the artifacts of street deixis requires an approach which goes beyond images or text, but also include aspects of multimodality in interaction–body posture and eye-gaze first of all.

The difference in deictic salience noted above also helps us define two concepts which need to be clearly distinguished: imaginary vs. fictive. The term 'fictive' was used (Talmy, 1996; Talmy, 2020; Matlock 2004) to discuss what has become known as 'fictive motion'–describing static objects in motion terms, as in *The road goes from Vancouver to Kamloops*. The term was further extended to 'fictive vision/fictive experience' (narrative instances where describing how a character conceptualizes a situation is achieved by describing what she fictively sees (Dancygier, 2012)), and to 'fictive interaction' mentioned above (using the language of spoken discourse when no speech takes place (Pascual 2006; Pascual, 2014)). Throughout this paper I have talked about 'Fictive Deixis'–examples relying on conversational DDC forms suggesting a deictic center when there is in fact no deictic center underlying the discourse, or it is only partial. What these cases share, and what distinguishes them from other discourse uses, is that the language forms evoke real world situations (motion, vision, interaction, deictic grounding) to represent mental construals. Just as the 'fictive motion' category above does not 'imagine' the road moving (but builds on mental or visual scanning instead), fictive deixis does not set up imaginary deictic grounds–rather, it structures situations in deictic terms to capitalize on the meaning potential of spoken discourse. The concept of Direct Discourse Ground represents the fact that in order to use the potential of spoken discourse, one needs to situate it in a deictic ground–pre-existing one or a fictive one.

Such an approach allows us to notice cross-discourse-genre correlations which are otherwise missed. For example, there are similarities between DDC use in Szymborska's poem and in internet memes, as examples from both genres are used for the purpose of evocation of more complex discourse appropriate to a frame (a funeral or a complaint). There are interesting correlations between the use of the deictic verb *come* in dramatic speech to non-human addressees and in storefront signs. There are also similarities and differences among fictive Deictic Grounds which profile non-prototypical hearers (the eyes, the poison, the city, any passer-by, etc.) or fail to profile any recipient at all (as in poetry). There are many aspects of Direct Discourse Constructions which can only be revealed when we consider the nature of the Fictive Deictic Ground as the primary step.

Finally, the approach taken in this paper develops concepts which can help us understand better the differences between literary language and colloquial speech, and between mono-modal and

multi-modal artifacts. Further research will uncover other ways in which Fictive Deixis is used in a range of discourse genres.

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.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This research has been funded by Insight Grant 435-2015-0158, Social Sciences and Humanities Research Council of Canada (SSHRC).

ACKNOWLEDGMENTS

I would like to thank my colleagues Lieven Vandelanotte (University of Namur, Belgium), Anna Bonifazi (University of Cologne, Germany), Arie Verhagen (Leiden University, The Netherlands), and Wei-lun Lu (Masaryk University, Brno, Czech Republic). Their insights and support have helped me greatly throughout the writing process.

REFERENCES

Brisard, F. (Editor) (2012). *Grounding: The Epistemic Footing of Deixis and Reference*. Berlin/Boston: De Gruyter, Inc.

Brône, G., Oben, B., Jehoul, A., Vranjes, J., and Feyaerts, K. (2017). Eye Gaze and Viewpoint in Multimodal Interaction Management. *Cogn. Linguistics* 28 (3), 449–483.

Bühler, K. (1990 [1934/1984]). *Theory of Language: The Representational Function of Language*, Vol. 25. Philadelphia: J. Benjamins Pub. Co, first published.1934.

Clark, H. H. (2016). Depicting as a Method of Communication. *Psychol. Rev.* 123 (3), 324–347. doi:10.1037/rev0000026

Clark, H. H., and Gerrig, R. J. (1990). Quotations as Demonstrations. *Language* 66 (4), 764–805. doi:10.2307/414729

Cornish, F. (2011). 'Strict' Anadeixis, Discourse Deixis and Text Structuring. *Lang. Sci.* 33 (5), 753–767. doi:10.1016/j.langsci.2011.01.001

Dancygier, B., and Sweetser, E. (Editors) (2012). *Viewpoint in Language: A Multimodal Perspective* (Cambridge: Cambridge University Press).

Dancygier, B. (2017). Viewpoint Phenomena in Constructions and Discourse. *Glossa* 2 (1), 37. doi:10.5334/gjgl.253

Dancygier, B. (2019). Proximal and Distal Deictics and the Construal of Narrative Time. *Cogn. Linguistics* 30 (2), 399–415. doi:10.1515/cog-2018-0044

Dancygier, B., Lu, W., and Verhagen, A. (Editors) (2016). *Viewpoint and the Fabric of Meaning: Form and Use of Viewpoint Tools across Languages and Modalities* (Berlin: De Gruyter Mouton), Vol. 55.

Dancygier, B., and Sweetser, E. (2005). *Mental Spaces in Grammar: Conditional Constructions*. Cambridge: Cambridge University Press. doi:10.1017/cbo9780511486760

Dancygier, B. (2012). *The Language of Stories: A Cognitive Approach*. Cambridge: Cambridge University Press.

Dancygier, B., and Vandelanotte, L. (2009). "Judging Distances: Mental Spaces, Distance, and Viewpoint in Literary Discourse," in Cognitive Poetics: Goals, Gains and Gaps. Editors G. Brône and J. Vandaele (Berlin: Mouton de Gruyter), 319–370.

Dancygier, B., and Vandelanotte, L. (2016). "Discourse Viewpoint as Network," in Viewpoint and the Fabric of Meaning: Form and Use of Viewpoint Tools across Languages and Modalities. Editors B. Dancygier, W. Lu, and A. Verhagen (Berlin: Mouton de Gruyter), 13–40. doi:10.1515/9783110365467-003

Dancygier, B., and Vandelanotte, L. (2017a). Internet Memes as Multimodal Constructions. *Cogn. Linguistics* 28 (3), 565–598. doi:10.1515/cog-2017-0074

Dancygier, B., and Vandelanotte, L. (Editors) (2017b). Special Issue: Viewpoint Phenomena in Multimodal Communication. *Cogn. Linguistics*, 28 (3).

Dancygier, B., Lee, D., Lou, A., and Wong, K. Forthcoming "*Standing Together by Standing Apart*: Distance, Safety, and Fictive Deixis in the COVID-19 Storefront Communication," in *COVID 19: Metaphor and Metonymy Across Languages And Cultures*. Editors X. Wen, W-L. Lu, and Z. Kövecses, (Amsterdam: John Benjamins Publishing Company).

David, O. (2016). *Metaphor in the Grammar of Argument Realization*. PhD dissertation: Berkeley: University of California.

Diessel, H. (2006). Demonstratives, Joint Attention, and the Emergence of Grammar. *Cogn. Linguistics* 17 (4), 463–489. doi:10.1515/cog.2006.015

Duchan, J. F., Bruder, G. A., and Hewitt, L. E. (Editors) (1995). *Deixis in Narrative: A Cognitive Science Perspective* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Dudis, P. G. (2004). Body Partitioning and Real-Space Blends. *Cogn. Linguistics* 15 (2), 223–238. doi:10.1515/cogl.2004.009

Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge: Cambridge University Press. doi:10.1017/cbo9781139174220

Fauconnier, G. (1994). *Mental Spaces*: Cambridge University Press.

Fauconnier, G., and Sweetser, E. (1996). (Editors) *Spaces, Worlds and Grammar*. (Chicago: University of Chicago Press). doi:10.6028/nist.sp.777-1996

Feyaerts, K., and Heyvaerts, G. (2021). *Welcome Back, We've Missed You!* Humanized Business Communication in Shop Window Messages during Early 2020-Lockdown. *Languages*, 6, 104. doi:10.3390/languages6020104

Fillmore, C. J. (1997). *Lectures on Deixis*, Vol. no. 65. Stanford, Calif: CSLI Publications.

Fillmore, C. (1982). "Towards a Descriptive Framework for Spatial Deixis," in Speech, Place and Action. Editors R. Jarvella and W. Klein (Chichester: John Wiley), 31–59.

Freeman, M. H. (2005). The Poem as Complex Blend: Conceptual Mappings of Metaphor in Sylvia Plath's 'the Applicant'. *Lang. Lit.* 14 (No. 1), 25–44. doi:10.1177/0963947005048874

Hanks, W. F. (1990). *Referential Space: Language and Lived Space Among the Maya*. Chicago: University of Chicago Press

Janzen, T. (2004). Space Rotation, Perspective Shift, and Verb Morphology in ASL. *Cogn. Linguistics* 15 (2), 149–174. doi:10.1515/cogl.2004.006

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites*, Vol. 1. Stanford, CA: Stanford University Press.

Langacker, R. W. (1991). *Foundations of Cognitive Grammar: Descriptive Applications*, Vol. 2 (Stanford, CA: Stanford University Press).

Levinson, S. C. (2008). "Deixis," in *The Handbook of Pragmatics*. Editors L. R. Horn and G. Ward, 97–121. doi:10.1002/9780470756959.ch5

Matlock, T. (2004). *"The Conceptual Motivation of Fictive Motion," in* Motivation In Grammar. Editors G. Radden and R. Dirven (Amsterdam: John Benjamins), 221–248.

Pascual, E. (2006). Fictive Interaction within the Sentence: A Communicative Type of Fictivity in Grammar. *Cogn. Linguistics* 17 (2), 245–267. doi:10.1515/cog.2006.006

Pascual, E. (2014). *Fictive Interaction: The Conversation Frame in Thought, Language, and Discourse*. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/hcp.47

Pascual, E., and Sandler, S. (Editors) (2016). *The Conversation Frame: Forms and Functions of Fictive Interaction* (Amsterdam & Philadelphia: John Benjamins).

Rubba, J. (1996). in *"Alternate Grounds in the Interpretation of Deictic Expressions," in* Spaces, Worlds And Grammar. Editors G. Fauconnier and E. Sweetser (Chicago: University of Chicago Press), 227–261.

Ruppenhofer, J., and Michaelis, L. A. (2010). A Constructional Account of Genre-BasedArgument Omissions. *Constructions and Frames* 2 (2), 158–184.

Sanders, J., and Redeker, G. (Editors) (1996). in Perspective and the Representation of Speech and Thought in Narrative Discourse. Spaces, Worlds and Grammar. (Chicago: University of Chicago Press), 290–317. doi:10.6028/nist.sp.777-1996

Semino, E. (1997). *Language and World Creation in Poems and Other Texts*. London and New York: Longman.

Short, M., Semino, E., and Wynne, M. (2002). Revisiting the Notion of Faithfulness in Discourse Presentation Using a Corpus Approach. *Lang. Lit.* 11 (4), 325–355. doi:10.1177/096394700201100403

Stukenbrock, A. (2014). Pointing to an 'empty' Space: Deixis Am Phantasma in Face-To-Face Interaction. *J. Pragmatics* 74, 70–93. doi:10.1016/j.pragma.2014.08.001

Stukenbrock, A. (2020). Deixis, Meta-Perceptive Gaze Practices, and the Interactional Achievement of Joint Attention. *Front. Psychol.* 11, 1779. doi:10.3389/fpsyg.2020.01779

Sweetser, E., and Stec, K. (2016). "Maintaining Multiple Viewpoints with Gaze," in Viewpoint And the Fabric Of Meaning. Editors B. Dancygier, W-L. Lu, and A. Verhagen (Berlin: Mouton-De Gruyter), 237–258. doi:10.1515/9783110365467-011

Talmy, L. (1996). *Toward a Cognitive Semantics: Concept Structuring Systems. Fictive Motion in Language and 'ception'*. (Cambridge, MA: MIT Press), Vol. 1, 99–175.

Talmy, L. (2020). Targeting in Language: Unifying Deixis and Anaphora. *Front. Psychol.* 11, 2016. doi:10.3389/fpsyg.2020.02016

Tomasello, M. (1995). *"Joint Attention as Social Cognition," in* Joint Attention: Its Origins And Role In Development. Editors C. Moore and P. J. Dunham (Hillsdale, N.J: Lawrence Erlbaum Associates).

Turner, M., Avelar, M., and Mendes de Oliveira, M. (2019). Blended Classic Joint Attention and Multimodal Deixis. *Signo* 44 (79), 03–09. doi:10.17058/signo.v44i79.12710

Vandelanotte, L. (2004). Deixis and Grounding in Speech and Thought Representation. *J. Pragmatics* 36 (3), 489–520. doi:10.1016/j.pragma.2003.10.003

Vandelanotte, L. (2009). *Speech and Thought Representation in English: A Cognitive-Functional Approach*. *Topics in English Linguistics*, 65. BerlinNew York: Mouton de Gruyter.

Vandelanotte, L., and Dancygier, B. (Editors) (2017). Special Issue: *Multimodal Artefacts And the Texture Of Viewpoint*. *J. Pragmatics* 122. doi:10.1016/j.pragma.2017.10.011

West, D. E. (2013). Cognitive and Linguistic Underpinnings of Deixis Am Phantasma: Bühler's and Peirce's Semiotic. *Sss* 41 (1), 21–41. doi:10.12697/sss.2013.41.1.02

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.

Copyright © 2021 Dancygier. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 16 August 2021
doi: 10.3389/fpsyg.2021.701486

Consequences of Sensory Modality for Perspective-Taking: Comparing Visual, Olfactory and Gustatory Perception

*Elsi Kaiser**

Department of Linguistics, University of Southern California, Los Angeles, CA, United States

Perspective-taking is fundamental for language comprehension, including the interpretation of subjective adjectives (e.g., *fun, tasty*, and *amazing*). To understand these adjectives, one needs to know whose opinion is being conveyed—in other words, who is the attitude-holder or perspectival center. Although the perspective-sensitivity of subjective adjectives has received considerable attention in prior work in formal semantics, potential effects of sensory modality (e.g., sight, taste, and smell) on the process of attitude-holder identification have not been systematically investigated. This paper reports a series of studies testing whether interpretation of subjective adjectives depends on whether they refer to the visual, olfactory (smell) vs. gustatory (taste) domains. The results provide evidence that sensory modality has a significant impact on the process of identifying the attitude-holder. This outcome suggests that perspective-sensitivity is highly context-dependent, and the observed modality effects align well with the biological and social properties of sight, taste, and smell.

Keywords: perspective-taking, subjective adjectives, predicates of personal taste, senses, psycholinguistics

OPEN ACCESS

Edited by:
Jorrig Vogels,
University of Groningen, Netherlands

Reviewed by:
Francesca Strik Lievers,
University of Genoa, Italy
Carolyn O'Meara,
National Autonomous University of Mexico, Mexico

***Correspondence:**
Elsi Kaiser
emkaiser@usc.edu

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

Received: *28 April 2021*
Accepted: *05 July 2021*
Published: *16 August 2021*

Citation:
Kaiser E (2021) Consequences of Sensory Modality for Perspective-Taking: Comparing Visual, Olfactory and Gustatory Perception. Front. Psychol. 12:701486. doi: 10.3389/fpsyg.2021.701486

INTRODUCTION

Subjective opinions are fundamental to human cognition and perception (e.g., Markus and Zajonc, 1985; Jarvis and Petty, 1996), and language contains a wide range of subjective expressions, including a class of adjectives known as predicates of personal taste (PPTs), such as *fun*, *tasty*, *disgusting*, *amazing*, that reflect opinions. Intuitively, these subjective adjectives differ from objective adjectives, such as *wooden*, *organic* or *Finnish*. If two people disagree about an objective statement such as "This apple is organic," one of them must be in the wrong. However, people can disagree about a subjective statement such "This apple is tasty" without anyone being in the wrong [e.g., Koelbel (2004) on faultless disagreement]: one person might find an apple to be tasty while another finds the same apple too sour. As the old adage goes, there's no arguing about taste. Thus, to fully understand subjective adjectives, a comprehender needs to know whose opinion/attitude is being conveyed. Understanding that a certain food is tasty to a cat is different from certain food is tasty to a human, for example. In theoretical analyses, PPTs are often described as making reference to a judge, attitude-holder or evaluator [e.g., Lasersohn (2005, 2007), Potts (2007), Stephenson (2007), Patel-Grosz (2012), but see also Pearson (2013), Coppock (2018)]. Thus,

intuitively, a fundamental aspect of successfully comprehending a PPT involves identification of the intended attitude-holder[1].

Although PPTs have received considerable attention in theoretical semantics and philosophy, to the best of my knowledge, semantically-oriented theories of PPTs do not explicitly make systematic distinctions based on sensory modality—in other words, whether the evidence on which the subjective opinion is based stems from visual, auditory, gustatory, olfactory, or tactile experience. Broadly speaking, under these accounts, a sentence such as "It was disgusting" is analyzed the same way semantically regardless of whether it refers to the taste, smell or visual appearance of a pizza slice, for example. These accounts are not incompatible with differences between sensory modalities, they simply do not make explicit predictions about them. This leaves open the (more pragmatic) question of whether comprehenders' interpretations of subjective adjectives—in particular, who is the intended attitude-holder—differ based on sensory modality. If yes, this can inform our understanding of the interplay between non-linguistic real-world knowledge and the interpretation of linguistic elements. Thus, one of the aims of the present paper is to systematically investigate whether modality matters in how language users identify the intended attitude-holder of PPTs.

Moreover, although the notion of perspective-taking has received extensive attention in psycholinguistic research [see e.g., Brown-Schmidt and Heller (2018) for a recent review], perspective-taking of the type involved with subjective predicates has received less attention. Most prior psycholinguistic work has focused on objective situations—contexts where one person has access to factually-correct information that the other person cannot access (see e.g., Keysar et al., 2000; Hanna et al., 2003; Heller et al., 2008). In the case of subjective opinions, everyone can have access to the same objective information but still arrived at different subjective opinions [and can all be equally "in the right," e.g., Koelbel (2004)]. The present paper aims to inform our understanding of perspective-taking in subjective contexts.

This paper reports three experiments testing if identification of the attitude-holder of subjective adjectives (specifically PPTs) in English is influenced by (i) whether the adjective is presented in a modality-neutral way or associated with a sensory modality, and (ii) whether different sensory modalities differ in how they impact the process of attitude-holder identification. Before taking a closer look at the hypotheses, let us first consider why one might expect sensory modalities to differ with respect to attitude-holder identification. With this goal in mind, the next section reviews work on the biological properties of different senses, their perceived subjectivity, and the ease of accessing sensory stimuli in different modalities.

[1]I use the term "attitude-holder" as a non-theoretical, descriptive term for the individual whose perspective/attitude is reflected by the subjective adjective or other subjective element. I set aside generic readings, where the attitude-holder is conceptualized as "people in general" [see e.g., Moltmann (2010), Pearson (2013) for discussion], as they are not salient in the contexts investigated here.

SENSORY MODALITIES

Vision, Taste, and Smell

It is well-known that the five traditional senses (sight/vision, hearing/audition, taste/gustation, touch/feel and smell/olfaction) are fundamentally different, not only in their biological underpinnings but also in terms of their (i) perceived level of subjectivity and (ii) spatiotemporal properties and associated differences in ease of perceptual access. The subsequent sections review the relevant properties of vision, taste and smell, the senses investigated in this paper[2]. We hope that this paper can serve as a foundation for future work on other sensory modalities.

Perceived Subjectivity

Vision is commonly viewed as the dominant sense in most human cultures and languages [e.g., San Roque et al. (2015), but contra Aikhenvald and Storch (2013)]. Biologically, vision is a highly specialized sense in humans, and by some estimates, up to 50% of the cortex is involved in visual functions (Palmer, 1999). Research on sensory dominance effects suggests that visual input tends to dominate over auditory input if the two are in conflict (Colavita, 1974; Sinnett et al., 2007; Spence, 2009)—humans may have a biologically hardwired preference to rely on vision [but see Aglioti and Pazzaglia (2010)].

As regards the perceived level of subjectivity, vision is typically viewed as providing relatively *objective* information: Sweetser (1990) claims that vision is "our primary source of objective data about the world" (p.39)[3]. Along similar lines, Korsmeyer (1999) calls vision "phenomenally objective" (p.25). Not surprisingly, visual evidence is often considered as more reliable than auditory or other kinds of evidence [e.g., in grammaticalized evidentiality systems, Willett (1988), Aikhenvald (2004)]. In essence, we typically think of visual sensory experiences as eliciting relatively uniform sensory percepts across people.

In contrast to vision, the senses of taste and smell are typically regarded as conveying more *subjective* information and as involving more variable percepts across people (e.g., Viberg, 1984, 2001; Chafe and Nichols, 1986; Sweetser, 1990; Dubois, 2007). Evidence for a higher level of subjectivity for taste and smell relative to vision comes from both neurological data and lexical patterns. On the linguistic side, it has been observed that linguistic terms related to smell and taste (e.g., Buck, 1949; Krifka, 2010; Levinson and Majid, 2014; Winter, 2016) have a strong subjective component. Even the simple utterances "it's tasty" and "it's smelly" both convey subjective opinions [in opposite

[2]I do not investigate hearing, touch, proprioception, or skin senses (e.g., temperature). This is due to challenges associated with incorporating them into the experiment: As explained below, items in the different conditions only differed with respect to the modality-indicating verb. All other aspects, including adjectives, were held constant. This ensures that any differences between modalities are not due to specific adjectives. Given these constraints, it was not possible to create plausible items involving more than three senses while also using the semantically relevant kind of adjectives.

[3]The link between vision and objective information has been claimed to explain why, in many Indo-European languages, verbs of vision develop cognition-related meanings (e.g., Sweetser, 1990). However, there are Australian languages where verbs of *audition* develop such meanings (e.g., Evans and Wilkins, 2000). Thus, there is crosslinguistic and cross-cultural variation in this regard.

directions, see e.g., Krifka (2010)]. As Winter (2016) notes, there are many vision-based expressions that convey largely objective information (e.g., *striped, green*, and *round*), whereas words linked to taste and smell tend to be more subjective (e.g., *smelly, pungent*, and *delicious*).

Even after we abstract away from particular lexical items, non-linguistic neurological experiments provide evidence for the idea that sensations of taste and smell are inherently more linked to subjective evaluation than visual percepts. For example, the amygdala (linked to the olfactory bulb) exhibits increased blood flow for pleasant or unpleasant smells and tastes, but *not* for visual (or auditory) stimuli (e.g., Royet et al., 2000). As noted by Phillips and Heining (2002), neural evidence indicates that "emotion processing and perception of odors and flavors have similar neural bases" and that "olfactory and gustatory stimuli seem to be processed to a significant extent in terms of their emotional content, even if not presented in an emotional context" (p. 204). Broadly speaking, neural evidence indicates that smell and taste have similar neural bases and are more subjective than vision.

In sum, vision differs from taste and smell in terms of the perceived level of subjectivity: the visual modality is regarded as conveying more objective, non-opinion-dependent information, while taste and smell convey more subjective, opinion-based information. This core asymmetry between vision on the one hand, and taste and smell on the other, is summarized by Caballero and Paradis (2015) who note that "in contrast to the relatively objective and stable nature of visual elements in the world, the perceptions of smell, taste and touch are highly subjective and variable across human beings" (e.g., Viberg, 1984, 2001; Chafe and Nichols, 1986). Thus, in contrast to the visual domain (where Person A will tend to assume that she has roughly the same visual experience as Person B when they focus their visual attention on the same thing), in the domains of taste and smell A is less likely to assume that she has the same gustatory or olfactory experience as B when they eat or smell the same thing.

Access to the Relevant Perceptual Experience

In addition to differing levels of perceived subjectivity, vision, taste and smell differ in terms of the spatial relation that holds between the experiencer and the stimulus. This has consequences for who can have the relevant perceptual experience to be considered a potential attitude-holder, in a way that groups the modalities differently from what we saw in the preceding section.

Vision is traditionally viewed as a distal sense. Indeed, we can see things that are (relatively) far away, and no direct physical contact is needed between the visual stimulus and the perceiver. Early Greek philosophers already noted that with vision, "there is no evident contact between the perceived object and the organs of perception" (Korsmeyer, 1999, p.12). These characteristics mean that multiple people can easily experience the same visual stimulus. Indeed, unless our eyes are closed, sighted individuals are constantly exposed to on on-going stream of visual stimuli. As noted by San Roque et al. (2015), "As a distal sense, it seems likely that sight is one of the most readily and regularly shared perceptual experiences among interlocutors" (p. 50). They also note that visual cues are generally viewed as the basic foundation for joint attention (e.g., Moore and Dunham, 1995). In essence, in the domain of vision, multiple people can easily experience a visual stimulus and thus are potential attitude-holders for a PPT that describes the resulting percept.

TABLE 1 | Vision, taste, and smell grouped by level of subjectivity and ease of perceptual access.

Subjectivity	vision << smell, taste *less subjective* *more subjective*
Access to the relevant perceptual experience	vision < smell << taste *less constrained access* *more constrained access*

In contrast to vision, taste is a proximal sense. To experience a gustatory stimulus, close contact—specifically, an event of something getting into one's mouth—is necessary. As noted by Elder et al. (2017, p. 878), "physical distance for taste is quite low because a stimulus must be within one's mouth in order to be sensed." Although humans with normal vision are constantly experiencing visual stimuli during their waking hours, we typically only experience gustatory stimuli while eating. Thus, unlike vision, perceptual access to the relevant experience in the gustatory domain is limited to those who have a specific kind of proximate contact with the stimulus.

What about smell? The olfactory modality resembles vision in that—unlike with taste—no direct physical contact is needed between the stimulus and the experiencer, although greater physical proximity is required with smell than with vision: "[U]nlike with taste and touch, the stimulus can be sensed without any contact with the body. It simply must be close enough for the molecules to reach the nose" (Elder et al., 2017, p. 878). In sum, although access to olfactory experiences is not as unconstrainted as access to visual experiences, it is easier for multiple people to be potential attitude-holders for a PPT describing an olfactory experience as compared to a gustatory experience[4].

Table 1 summarizes how these three sensory domains differ in (i) level of subjectivity and (ii) ease of access to the relevant perceptual experience.

As can be seen in **Table 1**, the olfactory modality has a mixed status in that it patterns like taste in terms of conveying subjective information, but is almost as unconstrained as vision in in terms of ease of perceptual access.

Linguistically speaking, the olfactory domain has been found to be unusual in another way as well: In many languages, the dedicated vocabulary for olfactory experiences is very limited (e.g., Yeshurun and Sobel, 2010) and people often struggle to name smells [but, importantly, this is not the case in all languages or cultures, Majid et al. (2018)]. Thus, while the sensory domains

[4]This section has a non-linguistic point-of-view—it is about sensory perception *independent* of linguistic encoding. In the linguistic domain, sensory verbs differ in terms of agentivity and intentionality/volition. E.g., *to look* conveys an agentive, intentional action; *to see* does not. In terms of their linguistic argument structure, *to smell* and *to taste* allow for both an agentive and non-agentive/experiencer-type interpretation (e.g., Gisborne, 2010). I discuss this more in the section entitled "Potential Complication: On the Meanings of Smell".

of vision and taste differ both in terms of perceived subjectivity and ease of perceptual access, the olfactory domain has a more mixed status, and is also linguistically less robustly encoded in English and many other languages. However, it is important to acknowledge that my experiments test U.S.-born native English speakers. Given crosslinguistic differences in the (in)effability of senses (e.g., Majid, 2021), the findings reported here should not be construed as linguistic universals, and further crosslinguistic experiments are needed.

Identifying the Attitude-Holder of Subjective Adjectives

So far, we have focused on the differences between sensory modalities. Let us now take a closer look at a class of linguistic expressions, *predicates of personal taste* (PPTs), that can be used to express sensory experiences in different modalities and that have attracted extensive attention in theoretical semantics and philosophy [e.g., Lasersohn (2005), Stephenson (2007), Pearson (2013) and many others; see also Solt (2018) for recent experimental work].

Prior theoretical work on PPTs has approached them in a largely modality-neutral way. Potential differences between sensory modalities have received little if any direct attention. Researchers have mostly focused on sentences of the form "noun is adjective" (ex.1), where the sensory modality is not linguistically explicit. Sentences like (2), where the verb explicitly pins down the relevant sensory modality, have not been extensively or systematically discussed.

(1) The muffin was disgusting.

(2) The muffin looked/smelled/tasted disgusting.

Prior theoretical work on PPTs has tended to focus on the fundamental question of how to linguistically represent the fact that the meaning of these adjectives is judge-dependent—in other words, their interpretation is in some way relativized/anchored to the opinion or perspective of an attitude-holder. A number of different analyses have been proposed, but the present paper does not aim to distinguish between different formal accounts of judge dependence (e.g., Anand and Korotkova, forthcoming; Lasersohn, 2005; Glanzberg, 2007; Stephenson, 2007; Anand, 2009; Cappelen and Hawthorne, 2009; Patel-Grosz, 2012; Pearson, 2013; Snyder, 2013; Bylinina, 2014; Coppock, 2018; Rudin and Beltrama, 2019; Zakkou, 2019; Willer and Kennedy, 2020).

As regards the question of *attitude-holder identification* (when a comprehender encounters a PPT, how does s/he figure out whose opinion the PPT expresses?), prior work agrees that the first-person speaker is the default attitude-holder [aka the judge, e.g., Lasersohn (2005) on autocentric perspective].

Perspective-Shifting Away From the Default Attitude-Holder

Crucially, with PPTs, the first-person speaker is not the only possible attitude-holder [see e.g., Lasersohn (2005); for discussion regarding a range of perspective-sensitive elements see e.g., Karttunen and Zaenen (2005), Wang et al. (2005), Stephenson (2007), Amaral et al. (2008), Harris and Potts (2009), Kaiser (2015), Korotkova (2016), and Kaiser and Herron Lee (2017, 2018)]. For an example of shifting, let's imagine that ex. (3) is an excerpt from a novel. It has a first-person narrator, as indicated by the presence of "I" in the first clause. In addition, the excerpt mentions Eliza (whom I refer to as the "character"). In this context, given that Eliza is described as seeing the muffin, she could potentially also be construed as the attitude-holder in addition to (or instead of) the first-person narrator [For related experimental data, see Kaiser (2015), Kaiser and Herron Lee (2017, 2018)]. In other words, it seems possible in (3) to *shift away* from the default perspectival interpretation where the first-person narrator is the attitude-holder to an interpretation where the character is the attitude-holder (Note that in (3), *looked* is used in the final sentence, and both the narrator and the character can presumably see the muffin; I discuss this below).

(3) When I came into the room, Eliza saw the muffin on the platter. It looked disgusting.

The general question of perspective-shifting in fictional narrative contexts has been investigated by philosophers, linguists, literary narratologists and psychologists [see e.g., Banfield (1973), McHale (1978), Clark and Gerrig (1990), Fludernik (1993), Redeker (1996), Schlenker (2004), Lasersohn (2005), Sharvit (2008), Harris (2012), Eckardt (2015), Kaiser (2015), Maier (2015), Salem et al. (2015), Hinterwimmer (2017), Abrusán (2020), and many others; see also Bortolussi and Dixon (2003) and Klages et al. (2020) on perspective-shifting].

Crucially, prior work on narratives shows that not only is it possible to interpret the character as an attitude-holder of a PPT (such as *disgusting* in ex.3), but that subjective expressions like PPTs can in fact serve as *explicit cues to perspective-shift* from the narrator to the character: There exists a large body of narratological evidence indicating that subjective linguistic expressions in general act as cues to shift from the perspective of the narrator to the perspective of a character. Many of the more literary investigations did not look systematically at the specific class of PPTs, but they cite subjective adjectives such as *poor*, *dear*, *terrific, marvelous, awful,* and *stupendous* as examples of elements that are cues to perspective shift [see e.g., McHale (1978, p. 269) and Fludernik (1993, p. 26)]. More generally, in addition to subjective adjectives, other kinds of subjective expressions are also known to trigger perspective shift, including interjections and exclamations like *alas* and *oh*, expressions of uncertainty like *probably*, *perhaps*, as well as epithets like *that idiot, the jerk* and so on [e.g., McHale (1978), Fludernik (1993), see also Banfield (1973)].

These findings corroborate the intuition that in a narrative context like (3), the character can be construed as an attitude-holder of the PPT *disgusting*[5]. However, because prior work

[5]There exists a large literature on different kinds of perspective-shifting phenomena in language, including free indirect discourse [free indirect speech, e.g., Sharvit (2008), Eckardt (2015)], protagonist projection (e.g., Abrusán, 2020), viewpoint shift [e.g., Hinterwimmer (2017); see also Hinterwimmer (2019)],

has not systematically investigated different sensory modalities, it is not yet known whether the likelihood of perspective shift (from narrator to character) is increased when a PPT is explicitly presented as involving a sensory experience, nor is it known whether the likelihood of perspective shift is modulated by the particular type of sensory modality.

Constraints on Identifying the Attitude-Holder

Even in narrative contexts, it is not the case that any character can freely be construed as the attitude-holder of a subjective adjective. There are constraints that guide this process. Of central relevance for the process of attitude-holder identification is the observation made in prior theoretical work that use of a PPT indicates that the attitude-holder must have *first-hand experience of the relevant kind* [see e.g., Pearson (2013), Ninan (2014) on the acquaintance inference, Bylinina (2014), Gunlogson and Carlson (2016), McNally and Stojanovic (2017), and Rudin and Beltrama (2019), inter alia]. In essence, for something to be judged fun or tasty, the person making this judgement must have the relevant experience. Usually, this someone is the default first-person speaker. For example, if I say (1), this suggests that I have the right kind of first-person experience (presumably direct gustatory, visual or olfactory experience; the prior theoretical semantics typically work does not make claims about specific modalities) on which to base my statement.

In other words, it has been claimed that PPTs entail that the attitude-holder is an *experiencer*, "a sentient individual who perceives the property in question" (McNally and Stojanovic, 2017, p. 24). The significance of the sentient experiencer is also discussed by Bylinina (2014): "A direct statement about someone's internal state can be made only if the judge parameter is set to the same value as the experiencer of this internal state" (Bylinina, 2014, p. 58). The attitude-holder/experiencer relation has been explored in depth, from a variety of perspectives, in several recent papers (e.g., Sæbø, 2009; Pearson, 2013; Ninan, 2014; Gunlogson and Carlson, 2016; Kennedy and Willer, 2016; Willer and Kennedy, 2020).

However, prior work in this tradition does not systematically distinguish between situations where an adjective is presented without a particular sensory modality (e.g., *It was disgusting*) vs. situations where an adjective is associated with a sensory modality (e.g., *It looked/smelled/tasted disgusting*). Nor does prior work in this vein make claims about differences between sensory domains. Thus, this raises the question: In narrative contexts, with sentences like (3) that have two possible attitude-holders, how does the requirement for the attitude-holder to have the relevant first-hand experience guide the process of attitude-holder identification when different sensory modalities are involved?

Aims of the Present Work

The need for the attitude-holder to have the relevant kind of subjective, first-hand experience, coupled with the fact that sensory modalities differ in terms of their (i) level of subjectivity and (ii) ease of access to the relevant perceptual experience, suggests that sensory modality may guide the process of identifying the attitude-holder of subjective adjectives. However, despite the large body of work on the fundamental physiological and physical differences between sensory modalities (see the section entitled "Perceived Subjectivity"), this question has not been systematically addressed in prior linguistically-oriented research.

The question of whether and how the interpretation of PPTs depends on sensory modality has (i) implications for our understanding of perspective-taking—including whether information explicitly linked to a sensory modality is a stronger cue to perspective shift—as well as (ii) implications for theories of subjective adjectives. In existing semantic work, many analyses of PPTs' judge dependence seem to implicitly or explicitly center on the adjective itself. However, if we find that the attitude-holder of the *same* adjective [e.g., *disgusting* in (2)] can be interpreted differently depending on sensory modality, this suggests that accounts that treat attitude-holder identification as determined purely by the semantic properties of a specific class of adjectives are not sufficient. Instead, this kind of outcome would be more amenable to pragmatically-oriented theories of PPTs that allow for contextual, top-down effects to play a role, because such accounts could be extended to encompass differences between the senses even when the adjective itself is held constant.

To fill this empirical gap and to address the issues sketched out above, this paper reports three experiments that test whether sensory modality has an impact on how English-speaking comprehenders identify the attitude-holder of a subjective adjective in narrative contexts with two possible attitude-holders: a first-person narrator and a character mentioned in the story.

We consider three hypotheses about the effects of sensory modality on the process of attitude-holder identification. In the rest of this section, I introduce these hypotheses: the Sensory Experience Hypothesis, the Subjectivity-based Hypothesis and the Inference-based Hypothesis. The first focuses on the basic question of whether presenting a subjective adjective in a modality-neutral way vs. associated with a particular modality influences who is interpreted as the attitude-holder. The latter two make specific predictions regarding sensory modalities and are rooted in the differences reviewed in the sections entitled "Perceived Subjectivity" and "Access to the Relevant Perceptual Experience".

To test these hypotheses, I used two-sentence sequences like ex.(4a-c), which describe a character having perceptual experiences in the visual, olfactory or gustatory domains respectively and present the critical PPT in predicative position using the matching sensory verb [See Anand and Korotkova (forthcoming) on attributive PPTs]. Thus, the sensory modality is expressed by means of the verb whose subject the character is (*Eliza saw/tasted/smelled*) and by means of the verb in the second sentence with the PPT (*it looked/smelled/tasted disgusting*). This ensures that it is clear to the reader that the character has a perceptual experience that "matches" the modality expressed in the PPT-containing sentence. I also compare these conditions to a baseline condition (ex.4d) where no perceptual experience is

psychonarration and narrated perception [e.g., Fludernik (1993), p. 311 for an overview]. The present work does not distinguish these potentially overlapping phenomena, as the differences between them are not relevant for the central aims of this paper.

described and the PPT is presented with the verb *was*.

(4a) *Vision condition*
When I came into the room, Eliza saw the muffin on the platter. It looked disgusting.
Whose opinion is it that the muffin looked disgusting?
The narrator's OR Eliza's

(4b) *Smell condition*
When I came into the room, Eliza smelled the muffin on the platter. It smelled disgusting.
Whose opinion is it that the muffin smelled disgusting?
The narrator's OR Eliza's

(4c) *Taste condition*
When I came into the room, Eliza tasted the muffin on the platter. It tasted disgusting.
Whose opinion is it that the muffin tasted disgusting?
The narrator's OR Eliza's

(4d) *Baseline*
When I came into the room, Eliza put the muffin on the platter. It was disgusting.
Whose opinion is it that the muffin was disgusting?
The narrator's OR Eliza's

In ex.(4a), the vision condition, only Eliza is linguistically specified as seeing the muffin (she is the subject of the verb *see*), but both the narrator and Eliza are in the same room. Thus, we can infer they can both see the muffin[6].

In ex.(4b), the smell condition, only Eliza is linguistically specified as smelling the muffin. However, again, given that she and the narrator are in the same room, both can be plausibly inferred as having the experience of smelling the muffin. However, the availability of the narrator as an attitude-holder is expected to be weaker than in the vision condition, because smell is more constrained by proximity than vision (see the section entitled "Access to the Relevant Perceptual Experience").

In contrast, in ex.(4c), the taste condition, only Eliza is linguistically specified as tasting the muffin. Thus, comprehenders can infer that only Eliza has the relevant gustatory experience.

In ex.(4d), the baseline condition, Eliza is not linguistically described as seeing, tasting or smelling the muffin—she simply puts in on a plate.

In all conditions, the basic syntactic structures are the same. The only difference is the verb in the second clause (*Eliza saw/smelled/tasted/put…*) and the verb in the final sentence (*it looked/smelled/tasted/was disgusting*). After each text, participants were asked a question about the attitude-holder of the subjective adjective. The question used a verb that matches the sensory modality expressed in the preceding sentence. This allows us to test whether different sensory modalities influence the extent to which PPTs are able to trigger perspective-shifting away from the default attitude-holder—the first-person narrator—toward a character. Let us now turn to the predictions made by the three hypotheses about the likelihood of perspective-shift in each condition.

TABLE 2 | Predictions of the three hypotheses.

	Predicted likelihood of perspective shift to character's perspective (i.e., character is the attitude-holder)
General hypothesis:	
Sensory Experience Hypothesis	(i) *baseline << sensory experience*
Additional hypotheses about differences between sensory modalities:	
Subjectivity-based Hypothesis	(ii) *baseline << vision << {smell, taste}*
Inference-based Hypothesis	(iii) *baseline << vision < smell << taste*

On a general level, it is possible that associating a subjective adjective with a specific sensory modality (regardless of what that modality is) will influence the process of attitude-holder identification, as compared to the same subjective adjective being presented in a context where no sensory modality is mentioned. I refer to this as the **Sensory Experience Hypothesis**. According to this hypothesis, describing the character as involved in *any kind* of sensory perception (e.g., *Eliza smelled/tasted/saw the muffin…*) makes perspective-shift to the character relatively more likely than in the baseline condition (e.g., *Eliza put the muffin…*). The predictions of this hypothesis regarding likelihood of perspective shift are in row (i) of **Table 2** (Note that this hypothesis makes no predictions about differences *between* sensory modalities).

More specific predictions about how different sensory modalities impact attitude-holder identification are made by the Subjectivity-based Hypothesis and the Inference-based Hypothesis.

According to the **Subjectivity-based Hypothesis,** the process of attitude-holder identification in contexts involving narrative fiction is guided by level of subjectivity (see the section entitled "Perceived Subjectivity"). Under this view, more subjective information acts as a stronger cue to shift from a narrator perspective to a character perspective than less subjective information. This prediction has its roots in the large body of research on perspective-shifting in narratives, which shows that *subjective linguistic expressions* function as a cue for readers to shift to the perspective of the character (see the section entitled "Perspective-Shifting Away From the Default Attitude-Holder").

Given that the visual modality is more objective than taste and smell (as discussed in the section entitled "Perceived Subjectivity"), the Subjectivity-based Hypothesis predicts that in the smell and taste conditions (4b, 4c), comprehenders are more likely to perspective-shift from the narrator to the character—i.e., to interpret *disgusting* as conveying Eliza's opinion—as compared to the visual condition (4a). The baseline condition is predicted to trigger the lowest rate of perspective-shift. The

[6]The first sentence (*Eliza saw the muffin*) uses *to see* instead of *to look*, to make the argument structural properties of the perception verbs (*saw/smelled/tasted*) as comparable as possible. The lexical semantics of the three verbs in the experiments all allow for an experiencer-based interpretation, unlike *look at* which is hard-wired for only an agentive reading (e.g., Kopytko, 1990; Gisborne, 2010) and deserves to be investigated in future work. Thus, the verbs in the current studies are unified in not requiring their subjects to be intentional agents. I discuss the verbs more in the section on "Potential Complication: On the Meanings of Smell".

predictions for the Subjectivity-based Hypothesis are in row (ii) of **Table 2**.

The predictions of the Subjectivity-based Hypothesis diverge from those of the **Inference-based Hypothesis.** This hypothesis posits that the process of attitude-holder identification is guided by comprehenders' inferences about who has access to the relevant perceptual experience. As discussed in the section entitled "Access to the Relevant Perceptual Experience", (a) taste experiences require physical contact between the stimulus and the perceiver and thus are more constrained than either vision or smell, and (b) although neither smell nor vision require physical contact, smell is more constrained by proximity requirements than vision. If attitude-holder identification is shaped by inferences about who plausibly has access to the relevant sensory experience to be an attitude-holder, we expect that taste will clearly differ from both vision and smell, and that smell can also differ from vision.

In ex.(4c) with *taste*, Eliza is expected to be interpreted as the attitude-holder, as she is the only one with access to the relevant perceptual experience (tasting the muffin). Thus, taste should elicit the highest rate of perspective-shifts to the character. What about smell? Unlike the Subjectivity-based Hypothesis which groups taste and smell together, the Inference-based Hypothesis predicts that smell (ex.4b) patterns more like vision (ex.4a) than like taste, due to smell and vision having relatively less constrained perceptual access properties than taste. Smell is presumably somewhat more constrained than vision, though, given that it typically involves greater physical proximity to the stimulus. Thus, the Inference-based Hypothesis predicts more shifts from the narrator to the character's perspective with taste than with vision or smell, as well as more shifts with smell than with vision. The baseline condition is predicted to elicit the lowest number of perspective shifts to the character. These predictions are in row (iii) of **Table 2** (The predictions of the three hypotheses regarding potential differences between sensory modalities are relative, not absolute).

The Inference-based Hypothesis has its roots in the general view that a central part of language processing has to do with context-sensitive inferences based on real-world knowledge that comprehenders make to arrive at a coherent interpretation of the discourse (e.g., Hobbs, 1979; Kehler, 2002). For example, Hobbs (1979) argues that pronoun interpretation is not governed by an independent mechanism (as many others have argued) but rather is a side-effect of comprehenders using real-world knowledge and reasoning to make inferences about how the components of a discourse fit together in a coherent way. Under the Hobbsian view, no special mechanism is needed for pronoun interpretation, beyond independently-needed reasoning and inferencing abilities that are rooted in our real-world knowledge. Thus, broadly speaking, the Inference-based Hypothesis can be viewed as a "cousin" of the coherence-based approach to pronoun interpretation. Thus, although the core assumption of the Inference-based Hypothesis (that we need to take seriously comprehenders' real-world inferences when considering aspects of language processing) has not previously been systematically tested in the domain of PPT interpretation, it is amply supported by prior work in other areas of language.

EXPERIMENT 1

Experiment 1 tests whether and how information about sensory modality guides the process of attitude-holder identification. Experiment 1 used sequences like ex.(4a-d) and manipulated whether a character in the narrative has a perceptual experience in the visual, auditory or olfactory modality, to see whether this influences the likelihood of perspective shifting from the default first-person narrator to the character.

Methods

Participants

Participants were recruited *via* Amazon Mechanical Turk and completed the experiment *via* the Qualtrics web interface. For all the studies reported in this paper, MTurk participants had to have a U.S. IP address, at least 1,000 previously-approved HITs and 98% or greater HIT approval rate. Participants received USD 1.50.

For Experiment 1, 56 native English speakers were included in the final data analysis. We only included those who self-reported being born in the U.S., speaking English as their first language (one person was excluded because of this), and having normal/corrected-to-normal vision and hearing (no one was excluded because of this) and who made no errors on four unambiguous catch trials (15 people excluded)[7]. In addition, three people were excluded to balance the number of participants per list.

In all experiments reported in this paper, exclusion criteria were pre-specified before data analyses on the target trials were conducted[8]. The research reported in this paper was reviewed and approved by the USC Institutional Review Board.

Design and Materials

Participants read two-sentence sequences (ex.4, repeated as ex.5), presented as extracts from novels, and answered questions about them. Each target started with a subordinate clause preamble that mentions the speaker/narrator by means of a first-person pronoun and describes the narrator as arriving at/entering the inferred location of the character. This clause is followed by a main clause that mentions a character by name. This set-up explicitly makes available two possible candidate attitude-holders (the narrator and the character e.g., Eliza, George, Amanda, Tim) for the "whose opinion" question that was presented after each target (Note that the question disambiguates "it"

[7]We asked about (non-correctable) vision and hearing impairments, because auditory and visual impairments can limit individuals' access to English-language input. The catch trials were two-sentence sequences like the targets, but did not describe sensory experiences. All four catch trials were followed by a forced-choice question that had a clear correct answer, For example, "Sandy heard that Destiny and Julie are cruel to their pets. Destiny dragged the kitten with the string" was followed by the question "Who had the string?" and the answer choices were "Destiny" and "Sandy".

[8]A potential concern with MTurk is the decrease in data quality that researchers have recently observed (e.g., Chmielewski and Kucker, 2020; Kennedy et al., 2020). Given these concerns, we used catch trials to exclude participants, in line with the advice of Chmielewski and Kucker. Our exclusion numbers may seem high, but are in line with or lower than Chmielewski and Kucker's: They found 38–62% of MTurk participants failing at least one data quality validity indicator in summer/fall 2018 and spring 2019.

as referring to the muffin/relevant object, not the platter or something else). This question was presented on the same screen as the two-sentence sequence, to avoid a memory load, and was a two-alternative force-choice question. The answers provide a measure of who participants think is the attitude-holder of the PPT.

(5) When I came into the room, Eliza {saw/smelled/tasted/put} the muffin on the platter. It {looked/smelled/tasted/was} disgusting.
Whose opinion is it that the muffin {looked/smelled/tasted/was} disgusting?
The narrator's OR Eliza's

The verbs were used to manipulate the senses involved in the item (vision, smell, taste, or no sense/baseline). Within an item, the PPT itself was kept constant in all conditions—this ensures that potential differences between the conditions cannot be attributed to the lexical semantics of particular adjectives. The adjectives were selected based on prior semantic work, and chosen so that they would be felicitous in the domains of smell, taste and vision (e.g., something can look, taste, or smell disgusting or amazing); both negatively-valenced and positively-valenced adjectives were included[9]. In each condition, the particular sensory domain was specified by the verbs in both the first and the last sentences, except for the baseline condition, where it was underspecified in both sentences. In the baseline condition, the verb *put* is used to describe the action done by the character, and the verb *was* is used in the second sentence.

The study included 24 target items, which used 12 different adjectives (specifically, predicates of personal taste; each used twice, see **Appendix**) and 24 different food items, as well as 42 filler items. The items were presented to participants in a Latin-Square design, so that no participant saw more than one version of each target. Variants where the preamble clause mentions the third-person character instead of the first-person narrator (e.g., *When she came into the room…*) were also included in the design, but are not reported here: They are not relevant for the perspective-shifting questions investigated in this paper because they do not explicitly introduce another potential attitude-holder.

Procedure

Participants were recruited from Amazon Mechanical Turk and completed the study at their own pace, over the internet. Participants read two-sentence sequences and answered questions about them (ex.5). The items were presented in writing. Each item was presented on a separate screen, but the critical sentences were displayed on the same screen as the multiple-choice question. Participants were told to imagine they were reading extracts from novels, and the term "narrator" was presented as part of the instructions.

[9]Given that these studies focus on PPT adjectives (a semantically restricted set), it was not possible to specifically match for levels of sensorimotor strength, but crucially all adjectives could be felicitously used with the three sensory domains.

Data Analysis

Data was analyzed using R (R Core Team, 2013). To compare the conditions to each other in order to assess effects of sensory modality on the process of attitude-holder identification, we fit logistic mixed effects regression models (glmer, lme4 1.1-20, Bates et al., 2015) to our data and used the emmeans package (emmeans 1.5.0, Lenth, 2018) to obtain Bonferroni-corrected pairwise comparisons. The proportion of "character's opinion" responses (1, 0) was used as the dependent variable; it is the inverse of the proportion of "narrator's opinion" responses. "Condition" was entered as a fixed effect into the model. As random effects, the models included random intercepts for subjects and items—as well as by-subject and by-item random slopes for the effect of condition when justified by model comparison—unless this resulted in singularity or non-convergence, in which case the model was further simplified [For each model, we started with the maximal random effect structure for subjects and items, and used model comparison to identify the maximal random effect structure justified by the design and supported by the data. Only random effects that contributed significantly to the model ($p < 0.05$) were included (Baayen et al., 2008)]. From-chance analyses were conducted using intercept-only logistic regression models.

Results

The proportion of "character's opinion" and "narrator's opinion" responses are shown in **Figure 1**. It's immediately clear that the baseline condition (no sensory modality specified) elicited mostly narrator responses and fewer than 25% character responses. This fits with the existing claims from the theoretical literature that the speaker (or writer) is the default attitude-holder of the PPT. Indeed, the proportion of character's opinion responses is significantly lower than chance (beta $= -1.415$, SE $= 0.0298$, $z = -4.739$, $p < 0.0001$).

The default preference to interpret the first-person narrator as the PPT attitude-holder vanishes in the other three conditions. Once the character in the narrative is described as the subject of a sensory verb (regardless whether it is seeing, smelling or tasting), that character becomes the preferred attitude-holder. Regardless of which sensory modality is specified, all three conditions elicit a higher-than chance rate of character responses (*taste*: beta $=$ 2.881, SE $= 0.764$, $z = 3.769$, $p < 0.001$, *smell*: beta $= 1.808$, SE $=$ 0.393, $z = 4.601$, $p < 0.0001$, *see*: beta $= 0.8699$, SE $= 0.328$, $z =$ 2.653, $p < 0.01$).

When the conditions are compared directly to each other, the baseline condition elicits less character responses (and more narrator responses) than all of the other conditions (see **Table 3** for statistical details). This supports the Sensory Experience Hypothesis.

In addition, a closer look at the different sensory modality conditions shows that the rate of character responses is higher (and the rate of narrator lower) in the smell and taste conditions than the see condition (see **Table 3**). However, the taste and smell conditions do not differ significantly from each other (although smell elicits a numerically lower proportion of shifts to the character's perspective). Thus, although all three sensory conditions show a preference to interpret the character as the

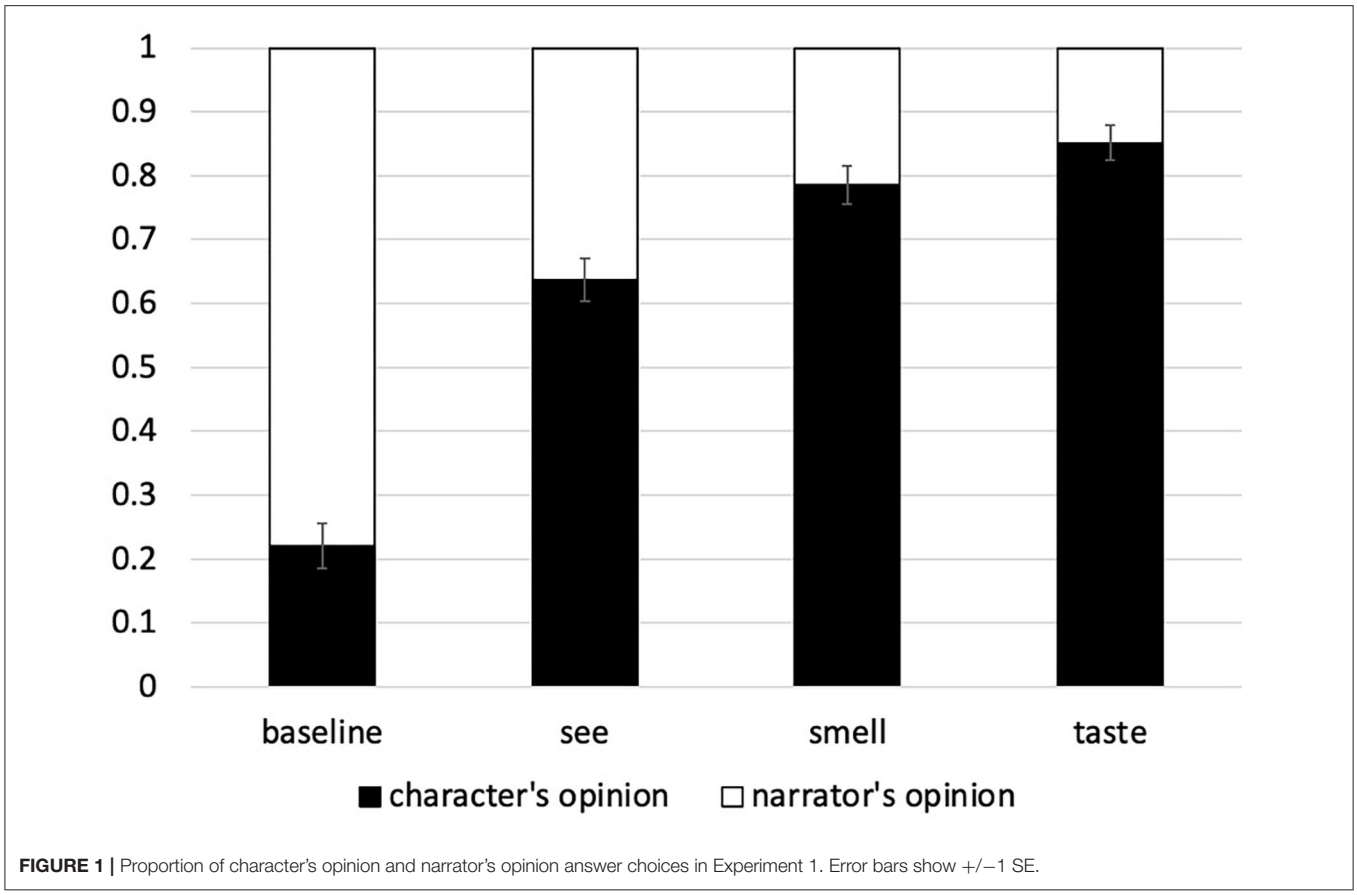


FIGURE 1 | Proportion of character's opinion and narrator's opinion answer choices in Experiment 1. Error bars show +/−1 SE.

TABLE 3 | Pairwise comparisons for Experiment 1.

	Estimate	SE	*z* ratio	*p*
Baseline—see	−2.482	0.311	−7.990	<0.0001
Baseline—smell	−3.487	0.350	−9.960	<0.0001
Baseline—taste	−4.080	0.385	−10.606	<0.0001
See—smell	−1.005	0.294	−3.422	0.0037
See—taste	−1.598	0.323	−4.943	<0.0001
Smell—taste	−0.0593	0.332	−1.790	0.4412

P-value adjustment: Bonferroni method for six tests.
Shading indicates significance at $p < 0.05$.

attitude-holder (rather than the narrator), this preference is stronger with taste and smell than with see, yielding the ranking *baseline* < *vision* < *{smell, taste}*, in line with the Subjectivity-based Hypothesis.

Discussion

The results of Experiment 1 show that sensory modality has a significant impact on the process of identifying the attitude-holder of predicates of personal taste (PPTs). When no sensory modality is specified (baseline), the narrator is the preferred attitude-holder. As soon as a sensory modality is specified in the context, we see more shifts to the character's perspective. These findings support the Sensory Experience Hypothesis: The same PPT is interpreted differently depending on whether it is presented in a modality-neutral way (baseline) or explicitly associated with a sensory modality (the other three conditions).

We also find significant differences between the three sensory modalities: Contexts involving the gustatory and olfactory modalities elicit more shifts to the character's perspective than contexts involving the visual modality: *vision* < *{smell, taste}*. Even though smell is numerically in between vision and taste, statistically it does not differ from taste. This outcome fits best with the Subjectivity-based Hypothesis, according to which the process of attitude-holder identification is guided by the level of subjectivity associated with each sensory domain.

Potential Complication: On the Meanings of Smell

A potential concern with Experiment 1 arises from the polysemy of the transitive verb *smell* in English (and many other languages): When used in a transitive sentence, as in the clause *Eliza smelled the muffin*, the verb *to smell* can have an agentive interpretation (e.g., a person sniffs the muffin on purpose) or an experiencer interpretation (e.g., a person simply breathes the air and thereby becomes aware of a smell), e.g., Kopytko (1990), Gisborne (2010), and Dziwirek (2016). Gisborne (2010) describes two meanings in terms of the agent vs. experiencer distinction; similarly, Kopytko (1990) uses the labels [+active, + intent] and [-active, -intent] for these two meanings (*To taste* is semantically

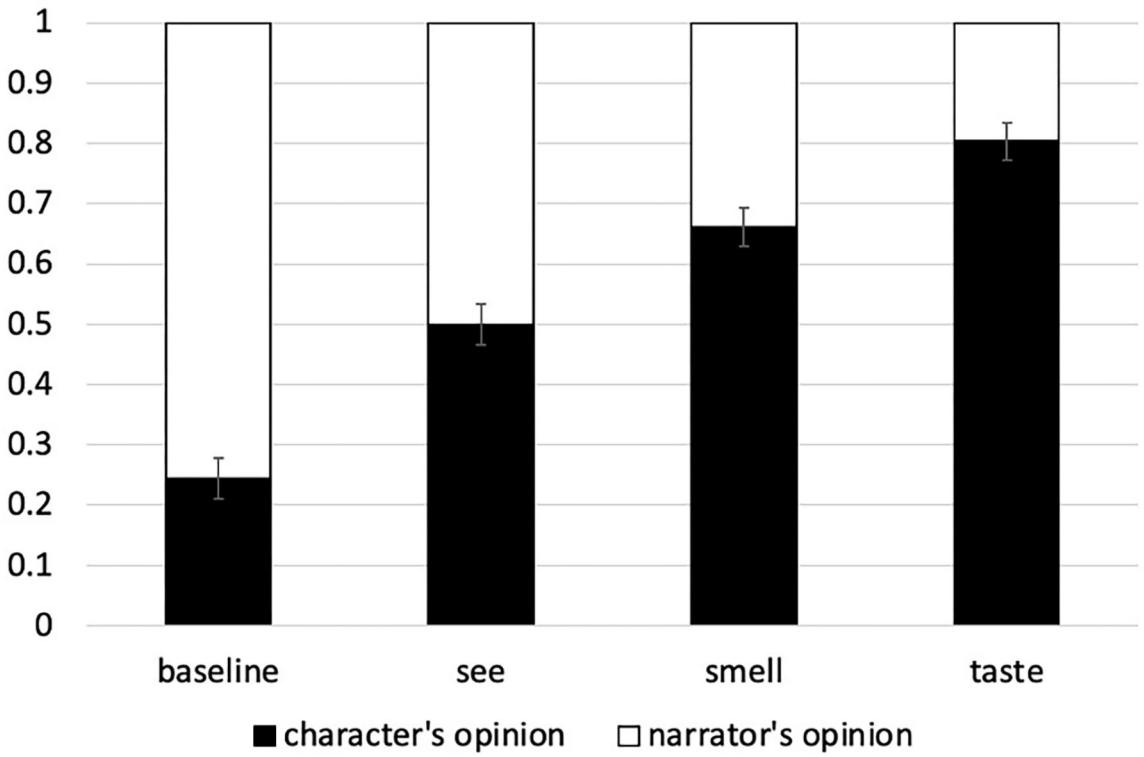


FIGURE 2 | Proportion of character's opinion and narrator's opinion answer choices in Experiment 2.

ambiguous in the same way, but in the contexts tested in this paper, the agentive interpretation is more salient. Thus, I focus here on *to smell*).

If the process of attitude-holder identification is guided by inferences about who has the relevant kind of experience (as posited by the Inference-based Hypothesis), these two meanings of the verb *smell* yield different predictions: If *smell* is interpreted as having the agentive meaning, the character is the most likely attitude-holder, as she is the one linguistically presented as the agent of the verb *smell*. In contrast, if *smell* has the experiencer meaning, both the character and the narrator are potential attitude-holders: Even the narrator, some distance away and not presented as the syntactic agent of smelling, can experience the smell.

If participants in Experiment 1 were interpreting the verb *smell* as having an agentive meaning, then—under the Inference-based Hypothesis—this would have boosted the character responses and made the smell condition pattern like the taste condition. This is indeed what we found. Thus, the conclusion that the results of Experiment 1 support the Subjectivity-based Hypothesis is too hasty, as the same outcome is also predicted by the Inference-based Hypothesis if the verb *to smell* is construed as having an agentive meaning[10]. So, rather than stemming from more subjective modalities triggering perspective shifting (in line with the Subjectivity-based Hypothesis), the results could instead be due to the polysemy of the verb *to smell* allowing for an interpretation that triggers an inference which favors the character as the attitude-holder (in line with the Inference-based Hypothesis).

EXPERIMENT 2

Experiment 2 addresses the issue left open by Experiment 1 by making the experiencer meaning of the verb *to smell* more available. This was done by adjusting the stimuli to boost the availability of the narrator as an attitude-holder, by adding a speaker-oriented intensifier that modifies the PPT (e.g., *totally disgusting*) [see e.g., Athanasiadou (2007) and Rhee (2016)].

If the proportion of character responses in the smell condition of Experiment 1 was boosted to taste-like levels by the polysemy of the verb *smell* (in particular by its agentive meaning), the Inference-based Hypothesis predicts that once the narrator is made more available as an attitude-holder in Experiment 2, a difference will emerge between the taste and smell conditions. This is because once the narrator is boosted as a salient attitude-holder, the experiencer meaning of the verb *smell* (a person breathes and becomes aware of a smell) is also expected to become more available, given that the narrator's smelling can be inferred to be experiencer-oriented. In this case, we expect a lower rate of character opinion responses. Crucially, because

[10]Using *sniff* instead of *smell* does not solve this concern, as *sniff* is typically interpreted as only having the agentive meaning.

this experiencer meaning is not available with the verb *taste* in the kinds of contexts we tested, the Inference-based Hypothesis predicts a difference between the taste and smell conditions in Experiment 2[11].

To boost the availability of the narrator as an attitude-holder, Experiment 2 uses adverbial intensifiers (e.g., *totally, absolutely*). Adverbial intensifiers are an ideal tool for our aim of making the narrator's perspective more prominent, because a large body of work shows that they function as signs of the speaker's opinions/attitudes [e.g., Athanasiadou (2007), Waksler (2012), Rhee (2016), Beltrama (2018), see also Biber and Finegan (1988)]. For example, Beltrama notes that "the use of *totally* contributes to strengthening the speaker's commitment toward the utterance" (Beltrama, 2018, p. 119–220). More generally, Athanasiadou (2007) notes that these kinds of intensifiers "tend to be subjective in character and show involvement on the part of the speaker" (p. 560), and Rhee (2016) adds that this class "encodes evaluation or reflects the speaker's positionality" (p. 399). Based on this prior work, it is reasonable to expect that these intensifiers can make the first-person narrator more available as an attitude-holder.

Thus, in Experiment 2 adverbial intensifiers were added to the PPTs, with the goal of making the narrator's perspective more salient. In analyzing the results of this experiment, we first check whether the addition of the intensifier indeed had the predicted effect of making the narrator more available as an attitude-holder, and if so, whether this results in a difference emerging between the smell and taste conditions as predicted by the Inference-based Hypothesis.

Method

Participants

Recruitment, payment and MTurk requirements were as in Experiment 1. Fifty-six new participants who had not done Experiment 1 were included in the final analysis. The exclusion criteria were the same as Experiments 1. In Experiment 2, three people were excluded for not being U.S.-born native English speakers and 20 for making errors on catch trials. Three additional people were excluded to balance the number of participants per list.

Design and Materials

The design was the same as Experiment 1, except that now, in all target items, the subjective adjective in the final clause was preceded by an intensifier (e.g., *totally, absolutely, really, extremely*), as shown in ex.(6).

(6) When I came into the room, Eliza {saw/smelled/tasted/put} the muffin on the platter. It {looked/smelled/tasted/was} really disgusting.
Whose opinion is it that the muffin {looked/smelled/tasted/was} really disgusting?
The narrator's
Eliza's

Procedure

The procedure was the same as Experiment 1.

Predictions

If the narrator becomes more available as an attitude-holder due to the presence of intensifiers, then, according to the Inference-based Hypothesis, the experiencer-based meaning of the verb *to smell* should become more available and a difference between the smell and taste conditions should emerge. This is because experiencing a taste percept requires physical contact between the taster and the stimulus [something that only the character does in narratives like ex.(6)], whereas on the experiencer-based meaning of *to smell*, both the character and the narrator can experience the smell percept.

In contrast, the Subjectivity-based Hypothesis predicts that the results of Experiment 2 will pattern like Experiment 1, because the inherent (non-linguistic) level of subjectivity associated with the gustatory, olfactory, and visual modalities is not impacted by the addition of a speaker-oriented adverb.

Results

The proportion of trials on which participants answered that the PPT reflects the opinion of the character is shown in **Figure 2**. As in Experiment 1, the proportion of narrator responses is the inverse of the character responses (due to the two-alternative forced-choice design). Like Experiment 1, in the baseline condition the proportion of character responses is significantly below chance (beta $= -2.05$, SE $= 0.55$, $z = -3.73$, $p < 0.001$), in contrast to the three sensory conditions: The proportion of character responses is significantly above chance in the taste and smell conditions (taste: beta $= 2.235$, SE $= 0.553$, $z = 4.04$, $p < 0.0001$), smell: beta $= 1.122$, SE $= 0.382$, $z = 2.935$, $p < 0.001$), and at chance in the see condition (beta $= 0.016$, SE $= 0.33$, $z = 0.048$, $p > 0.96$).

Before directly comparing the conditions to each other, let's first check whether the presence of intensifiers increases the availability of the narrator as an attitude-holder in the expected conditions, namely smell and see (i.e., contexts where the narrator could plausibly have access to the relevant experience). Indeed, the rate of narrator responses with smell and see is higher in Experiment 2 than Experiment 1 (*smell* Exp 2 vs. Exp 1: beta $=$ 0.922, SD $= 0.462$, $z = 1.996$, $p = 0.046$; *see* Exp 2 vs. Exp 1: beta $=$ 0.899, SD $= 0.463$, $z = 1.941$, $p = 0.052$). There are no significant differences between Experiments 1 and 2 in the baseline or the taste conditions, as expected (p's > 0.3). The differences between Experiments 1 and 2 confirm that presence of an intensifier does indeed boost the likelihood of the first-person narrator being interpreted as the attitude-holder in exactly those contexts where the narrator can also be inferred to be a plausible experiencer.

Furthermore, when we compare the conditions to each other, we find clear differences between all four conditions. First, as in Experiment 1, the baseline condition elicits significantly fewer

[11] These predictions could be recast in terms of agentivity, in a way that I regard as compatible with the Inference-based Hypothesis: If we make the agentive meaning of the verb *to smell* less available—in a context where the verb *to taste* only has an agentive meaning—we predict a decrease in perspective-shifts to the character's viewpoint. I regard an agentivity-based account as a sub-case of the inference-based account, because the effects of agentivity can be plausibly attributed to comprehenders' inferences being sensitive to who is linguistically presented as the agent of a particular verb.

TABLE 4 | Pairwise comparisons for Experiment 2.

	Estimate	SE	*z* ratio	*p*
Baseline—see	−1.695	0.305	−5.558	<0.0001
Baseline—smell	−2.688	0.329	−8.171	<0.0001
Baseline—taste	−3.731	0.370	−10.098	<0.0001
See—smell	−0.994	0.270	−3.564	0.0022
See—taste	−2.037	0.314	−6.493	<0.0001
Smell—taste	−1.043	0.304	−3.425	0.0037

P-value adjustment: Bonferroni method for six tests.

character responses than all three of the sensory experience conditions, in line with the Sensory Experience Hypothesis (statistical details are in **Table 4**). Second, as in Experiment 1, the taste and smell conditions elicit more character responses than the vision condition. However, now we also find significant differences between taste and smell: the rate of character responses is higher (and the rate of narrator responses lower) in the taste condition than in the smell condition. Thus, for the proportion of character's perspective interpretations, we find the following ranking: *baseline* < *see* < *smell* < *taste*. The distinction that now emerges between smell and taste is not predicted by the Subjectivity-based Hypothesis, but is compatible with the Inference-based Hypothesis.

Discussion

Experiment 2 set out to test a potential concern left open by Experiment 1, namely that the taste and smell conditions patterning together in terms of attitude-holder identification might be due *not* to similar levels of subjectivity (as proposed by the Subjectivity-based Hypothesis) but instead due to the polysemy of the verb *to smell* [agentive vs. experiencer meanings, e.g., Gisborne (2010)] boosting the rate of character responses, thereby making the smell condition look like the taste condition. Experiment 2 used adverbal intensifiers to boost the availability of the narrator as an attitude-holder, thus highlighting the experiencer meaning of the verb *to smell*, to see if this would reveal a difference between smell and taste conditions, as predicted by the Inference-Based Hypothesis.

Indeed, the results show that now, a significant difference between the smell and taste conditions emerges, exactly in the direction predicted by the Inference-based Hypothesis: The same subjective adjective is more likely to be interpreted as having the character as the attitude-holder in the taste condition than in the smell condition. According to the Inference-based Hypothesis, this is exactly what we expect, given that experiencing a taste percept requires the physical act of tasting [something that only the character does in narratives like ex.(6)] whereas we can infer that both the character and the narrator can experience the smell percept in these contexts. Crucially, the results of Experiment 2 are not predicted by the Subjectivity-based Hypothesis. According to this hypothesis, the outcome of Experiment 2 should have been just like Experiment 1, since the inherent (non-linguistic) level of subjectivity associated with the gustatory, olfactory and visual modalities does not change between experiments.

Multiple Potential Attitude-Holders

However, a potential concern with both Experiments 1 and 2 is that participants had to select a single attitude-holder when answering the forced-choice question about who is the attitude-holder (either the character or the narrator). There was no way to indicate an interpretation where *both* the narrator and the character share the opinion expressed by the subjective adjective. This is potentially problematic for the see and smell conditions, where it is possible for both the narrator and the character to have access to the relevant perceptual experience, and thus both could be inferred to be potential attitude-holders. Because Experiments 1 and 2 did not allow participants to report this kind of interpretation and forced them to opt for a binary response, one may wonder whether this constraint was distorting the results. To address this concern, in Experiment 3 participants had greater flexibility in indicating who they felt were the relevant attitude-holders: A third answer-choice was added that offers *both* the character and the narrator as possible attitude-holders. This allows us to assess whether the pattern predicted by the Inference-based Hypothesis—specifically the claim that vision and smell pattern more alike than taste—arises even when participants are free to select both the character and the narrator as attitude-holders.

EXPERIMENT 3

Methods

Participants

Recruitment, payment and MTurk requirements were as in Experiments 1 and 2. Sixty-four new participants who had not done Experiment 1 or 2 were included in the final analysis. The exclusion criteria were the same as above. Four people were excluded for not being U.S.-born native English speakers, one for reporting a hearing impairment and 26 for making errors on catch trials.

Design and Materials

The design and materials were the same as Experiments 1 and 2, except that now, the multiple choice question after each item had three answer choices (ex.7): Participants now had the additional option of selecting *both* the narrator and character as attitude-holders.

(7) When I came into the room, Eliza {saw/smelled/tasted/put} the muffin on the platter. It {looked/smelled/tasted/ was} disgusting.
Whose opinion is it that the muffin {looked/smelled/tasted/ was} disgusting?
The narrator's
Eliza's
Both the narrator and Eliza have this opinion.

Procedure

The procedure was the same as Experiments 1 and 2.

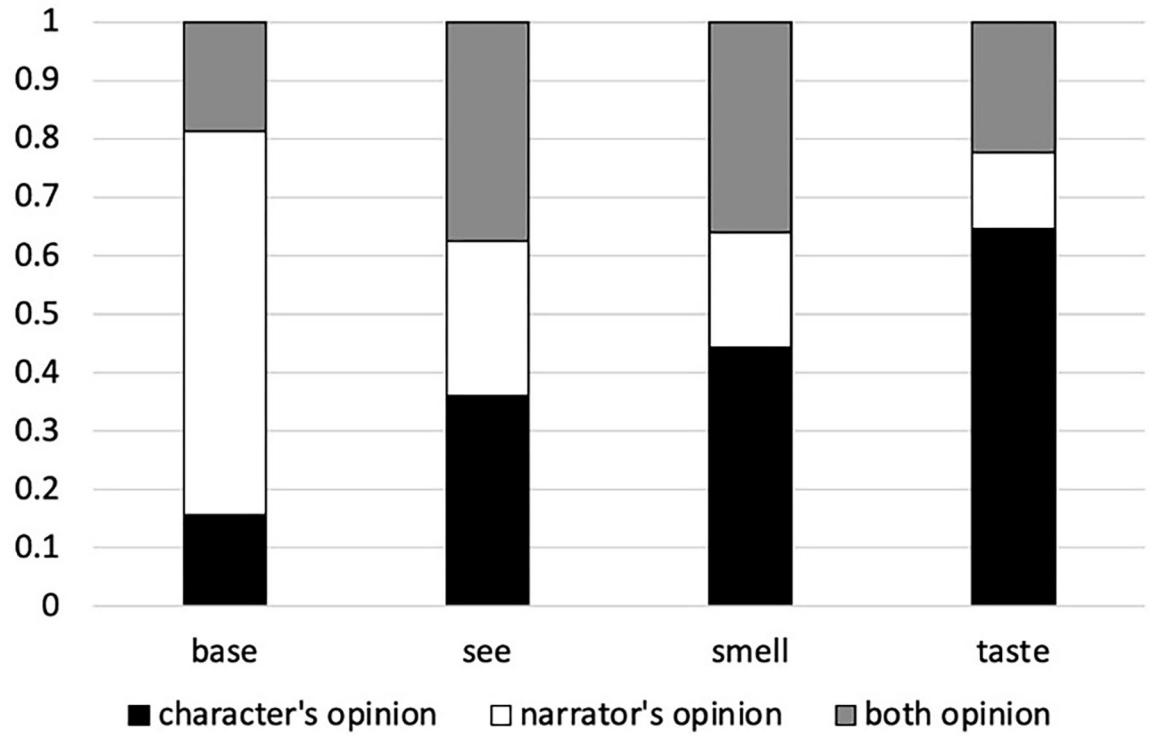


FIGURE 3 | Proportion of character's opinion, narrator's opinion and both answer choices in Experiment 3.

Data Analysis

Data analyses were conducted in the same way as in Experiments 1 and 2, except that—since the dependent variable now has three levels—we conducted three analyses: One analysis on the proportion of *character's opinion* responses (with character's opinion responses coded as 1 and all other responses as 0), a second on the proportion of *narrator's opinion* responses (with narrator's opinion responses coded as 1 and all other responses as 0), and a third on the proportion of *both* responses (with both responses coded as 1 and all other responses as 0).

Predictions

The broad predictions about perspective-shift are the same as for Experiments 1 and 2: The Subjectivity-based Hypothesis predicts that the taste and smell conditions will pattern alike, differently from vision. In contrast, the Inference-based Hypothesis predicts that taste and smell will differ, with smell falling in-between taste and vision.

Results

Figure 3 shows the proportion of character's opinion, narrator's opinion and both opinion responses for each of the four conditions. The outcomes of the statistical analyses are reported in **Tables 5A**–**C**. As in Experiments 1 and 2, the baseline condition elicits the most narrator responses: When no sensory modality is specified, participants are most likely to interpret the narrator as the attitude-holder of the subjective adjective. Indeed, the proportion of narrator responses is significantly higher in the baseline condition than in the other three conditions (**Table 5B**). This fits with the Sensory Experience Hypothesis. Next, we consider the differences between the three sensory modalities.

TABLE 5A | Character's opinion responses: pairwise comparisons for Experiment 3.

	Estimate	SE	*z* ratio	*p*
Baseline—see	−1.868	0.339	−5.505	<0.0001
Baseline—smell	−2.511	0.349	−7.194	<0.0001
Baseline—taste	−4.166	0.406	−10.251	<0.0001
See—smell	−0.643	0.288	−2.233	0.1533
See—taste	−2.297	0.332	−6.921	<0.0001
Smell—taste	−1.654	0.317	−5.216	<0.0001

P-value adjustment: Bonferroni method for six tests.

See vs. Taste

There are more narrator responses and both responses with see than taste (**Tables 5B**,**C**) and more character responses with taste than see (**Table 5A**)—echoing the high rate of perspective-shifts to the character with taste that we saw in Experiments 1 and 2. This difference is predicted by both the Subjectivity-based Hypothesis and the Inference-based Hypothesis.

TABLE 5B | Narrator's opinion responses: pairwise comparisons for Experiment 3.

	Estimate	SE	*z* ratio	*p*
Baseline—see	2.496	0.298	8.377	<0.0001
Baseline—smell	3.075	0.325	9.468	<0.0001
Baseline—taste	3.827	0.371	10.307	<0.0001
See—smell	0.582	0.303	1.923	0.3269
See—taste	1.331	0.340	3.919	0.0005
Smell—taste	0.749	0.347	2.162	0.1836

P-value adjustment: Bonferroni method for six tests.

TABLE 5C | Both (character and narrator's) opinion responses: pairwise comparisons for Experiment 3.

	Estimate	SE	*z* ratio	*p*
Baseline—see	−1.409	0.286	−4.920	<0.0001
Baseline—smell	−1.306	0.286	−4.565	<0.0001
Baseline—taste	−0.319	0.295	−1.079	1.000
See—smell	0.103	0.254	0.404	1.000
See—taste	1.090	0.275	3.967	0.0004
Smell—taste	0.988	0.275	3.593	0.0020

P-value adjustment: Bonferroni method for six tests.

Smell vs. Taste

The smell condition elicits fewer character responses, more both responses and numerically more narrator responses[12] than the taste condition (**Tables 5A–C**). Broadly speaking, as in Experiment 2, taste elicits more perspective-shifts to the character than smell. Thus, these two conditions do not pattern alike, contrary to the predictions of the Subjectivity-based Hypothesis and in line with the Inference-based Hypothesis.

Smell vs. See

Numerically, the smell condition elicits more character responses (smell: 44.27% vs. see: 35.94%), fewer narrator interpretations (smell: 19.79% vs. see: 26.56%)[13] and comparable numbers of both responses as the see condition (35.94% vs. 37.5%). Although these differences do not reach significance, they resemble the finding in Experiment 2 that the smell condition is in-between the see and the taste conditions in terms of the likelihood of perspective-shifts to the character.

The comparable numbers of both responses fit with the inference-based view that the character and the narrator are both plausible attitude-holders in the smell and see conditions. In line with this line of thinking, the proportion of both responses is significantly higher in the see and smell conditions than in the taste or the baseline condition (**Table 5C**)—which is what the Inference-based Hypothesis leads us to expect.

Discussion

Experiment 3 set out to address a potential concern with Experiments 1 and 2, namely that by forcing participants to choose only one attitude-holder, we may inadvertently have distorted the data. Thus, in Experiment 3 three answer choices were provided, so that participants could select the narrator, the character or *both* as attitude-holders.

As a whole, the results of Experiment 3 are more in line with the Inference-based Hypothesis—as well as the general Sensory Experience Hypothesis—than with the Subjectivity-based Hypothesis: In a flexible situation where participants are able to select both the character and the narrator as potential attitude-holders, we replicate the earlier differences between the baseline condition vs. all three sensory conditions (predicted by the Sensory Experience Hypothesis) and the differences between smell and taste (in line with the Inference-based Hypothesis and contrary to the predictions of the Subjectivity-based Hypothesis). Moreover, the gradient numerical differences between see, smell and taste (in order of increasing perspective-shift to the character) and the high numbers of both responses with see and smell fit best with the Inference-based Hypothesis.

GENERAL DISCUSSION

The interpretation of subjective adjectives such as *amazing, disgusting, interesting*, and *enticing* depends on an attitude-holder with the relevant kind of experience. To judge, say, a cake as tasting or smelling amazing, the attitude-holder must have tasted or smelled the cake. However, although the importance of the relevant kind of first-person experience has been acknowledged in prior work on attitude-holder identification, potential differences between sensory modalities have not been systematically investigated. The present work investigates whether and how the process of identifying the attitude-holder is influenced by the sensory modality of the subjective experience. Short narratives were used to test if sensory modality has an effect on whether a predicate of personal taste (PPT) is interpreted as reflecting the point-of-view of the default attitude-holder (first-person narrator), or shifted to the point-of-view of a character in the narrative. The experiments manipulated whether the PPT referred to a visual, olfactory or gustatory experience (e.g., *Eliza smelled/tasted/saw the muffin … The muffin looked/smelled/tasted disgusting*) or was presented in a modality neutral way (e.g., *Eliza put the muffin … The muffin was disgusting*).

Perspectival Consequences of Being Linguistically Realized as a Sensory Experiencer

Let us first consider the Sensory Experience Hypothesis. According to this hypothesis, presenting the character as the subject of any kind of sensory perception verb (e.g., *Eliza smelled/tasted/saw the muffin…*) makes a perspective-shift from

[12]The narrator responses do not differ significantly after Bonferroni correction (**Table 5B**), but before Bonferroni correction, the smell condition elicits more narrator responses than the taste condition ($p = 0.03$).

[13]The differences in the proportion of character responses and narrator responses between the see and the smell conditions are not significant after strict Bonferroni correction (see **Tables 5A**,**B**), but they are significant or marginal without Bonferroni (smell vs. vision character responses $p = 0.03$, narrator responses $p = 0.054$).

the narrator to the character more likely than if the character is not presented as a sensory experiencer (e.g., *Eliza put the muffin…*). Indeed, all three experiments support this hypothesis: Mention of any sensory modality—whether it be visual, olfactory or gustatory—triggers more perspective-shifting than the baseline condition where no sensory experience is mentioned. Though this finding does not contradict current semantic theories of subjective adjectives, it is not directly predicted by them. As a whole, the finding that mention of any sensory modality increases the likelihood of perspective-shift highlights the importance of contextual factors and the role that pragmatic inferences play in the interpretation of PPTs.

Because the same PPTs were used in all version of an item, these results cannot be attributed to the lexical semantics of particular adjectives. By using the same adjective in all versions of an item, the present work differs from most prior linguistic work on sensory vocabulary which has tended to focus on specific lexical items related to different senses (e.g., *pungent* for smell, *delicious* for taste). Using the same subjective adjectives in all four conditions (e.g., *disgusting*, *amazing*), the present work shows that the effect of sensory domain is not restricted to the lexical semantics of certain adjectives. Furthermore, this approach has the advantage of not requiring us to identify certain adjectives as being linked to a certain sensory modality. As Winter (2016) notes, sensory terms are often multimodal (p. 976), and assigning sensory modalities to adjectives is non-trivial. In the present work, I sidestep this concern by using the verbs to unambiguously indicate what the modality is.

Differences Between Sensory Modalities

Having established that explicit presentation as a sensory experiencer boosts the likelihood of being identified as the attitude-holder, we can now ask whether there are further differences between the sensory domains of taste, smell and vision. According to the Subjectivity-based Hypothesis, more subjective information acts as a stronger cue to shift to the character's perspective than less subjective information. Since the visual modality is regarded as conveying more objective information than taste or smell (see the section entitled "Perceived Subjectivity"), the Subjectivity-based Hypothesis predicts that comprehenders are more likely to perspective-shift from the narrator to the character with taste and smell than with vision (*vision* < *smell, taste*). In contrast, the Inference-based Hypothesis posits that the process of attitude-holder identification is guided by comprehenders' inferences about who has access to the relevant perceptual experience. Based on differences between modalities regarding the need (or lack thereof) of physical contact and proximity to the stimulus (see the section entitled "Access to the Relevant Perceptual Experience"), this hypothesis predicts that perspective-shift is more likely with taste than smell or vision (physical contact vs. no physical contact needed). This hypothesis is also compatible with more perspective-shifts with smell than vision: Although neither smell nor vision require physical contact, perceiving smells is more governed by physical proximity than perceiving visual input (*vision* < *smell* << *taste*).

Furthermore, because the Inference-based Hypothesis derives the differences between conditions from comprehenders' inferences about who is most likely to have the relevant experience, contextual factors that modulate these inferences can also play a role. The Subjectivity-based hypothesis does not predict sensitivity to contextual factors, because the inherent (non-linguistic) subjectivity of the different sensory modalities stems from their biological and neural properties, which do not change in different linguistic contexts.

Put together, the results of the three experiments presented here provide more support for the Inference-based Hypothesis than for the Subjectivity-based Hypothesis. The taste condition triggers more perspective-shifting to the character than the vision condition in all three studies, as predicted by both hypotheses. At first glance, the results of Experiment 1 appear to show attitude-holder identification in the smell condition patterning like the taste condition, but Experiment 2 shows that once we address the polysemy associated with the agentive and experiencer meanings of the verb *smell*, the smell condition no longer patterns like the taste condition. Experiment 3 corroborates the result from Experiment 2 by showing that even if participants have the possibility of selecting multiple attitude holders, the smell condition diverges from the taste condition. These findings support the Inference-based Hypothesis for English. Overall, there is no convincing evidence of smell and taste consistently patterning more alike than smell and vision, contrary to the predictions of the Subjectivity-based Hypothesis.

The conclusion that inferences about who has access to the relevant sensory experience guide the process of attitude-holder identification fits with the general view that much of language processing has to do with inferences based on real-world knowledge that comprehenders make to arrive at a coherent interpretation of the discourse [Hobbs (1979), see also Kehler (2002)][14]. This coherence-based approach captures many key aspects of pronoun use and interpretation and is supported by a growing body of pronoun-focused research [see e.g., Kehler and Rohde (2013)]. Although the present paper does not argue for a direct equivalence between pronoun interpretation and attitude-holder identification, it is (to the best of my knowledge) the first to provide experimental evidence that the process of attitude-holder identification with subjective predicates is sensitive to inferences rooted in the real-world differences between sensory modalities coupled with information from the linguistic context.

On the Variability of Smell

Compared to the visual and gustatory condition, the results for the olfactory conditions in all three experiments in this paper are relatively more variable. While this may at first glance seem unexpected, there are multiple reasons why the process of attitude-holder identification in the smell condition can be more variable than with vision or taste. First, on a non-linguistic level, the sensory modality of smell is in-between taste and vision in

[14]Future work on the Inference-based Hypothesis could use sensorimotor norms (e.g., Lynott and Connell, 2009; Lynott et al., 2020) to investigate how fine-grained ratings about the sensory properties of subjective adjectives modulate the process of attitude-holder identification.

terms of not requiring direct physical contact but still requiring some level of physical proximity (as discussed in the section entitled "Access to the Relevant Perceptual Experience"), the level of which could depend on the strength and type of the smell as well as other contextual factors. This is likely to make inferences about potential attitude-holders more variable with smell than with vision or taste. Another potential source of variation is the fact that, on the linguistic level, in English the verb *to smell* is ambiguous between an agentive, intentional act of smelling vs. a non-agentive, non-volitional act of experiencing a smell (see the section entitled "Potential Complication: On the Meanings of Smell"). These two shades of meaning add further variation to the inferences that comprehenders can draw, as we already saw in the differences between Experiments 1 and 2. Third, it is possible that inferences about smell being less stable than vision or taste stems from the same (not-yet-fully-understood) reason that underlies the impoverished nature of the lexicon for smell in English and many other languages and people's struggles with naming smells [see the section entitled "Access to the Relevant Perceptual Experience", e.g., Yeshurun and Sobel (2010)]. Future research on languages with richer lexicons for smell would shed light on this.

Crucially, this kind of variability between the smell conditions and the other conditions is fully compatible with the Inference-based Hypothesis, as this hypothesis allows for multiple factors to influence the kinds of inferences made by comprehenders—in fact, this is a key prediction made by the Inference-based Hypothesis. This contrasts with the Subjectivity-based Hypothesis, under which the level of subjectivity attributed to each modality is fixed, and thus the process of attitude-holder identification is not expected to change based on contextual factors.

Implications for Semantic Theories of Predicates of Personal Taste (PPTs)

Experiments 1, 2, and 3 show that the same subjective adjective can be interpreted differently depending on whether it is presented in a modality-neutral way (baseline condition) or explicitly associated with a sensory modality (the other three conditions). These results pose challenges for analyses of judge dependence that are purely centered on the adjective, as they suggest that accounts that only focus on the adjective itself and do not take the context (including the rest of the sentence) into account are not sufficient.

Many prior semantic accounts of PPTs seem to be largely focused on the semantic representation of the adjective itself [e.g., Lasersohn's (2005) judge-based analysis] and thus do not appear to straightforwardly predict the results reported here. Although these accounts are not incompatible with my results, if one wants to maintain a purely semantic approach, it seems that some kind of additional mechanism—perhaps one that is able to operate directly on the assignment of the judge variable/parameter, if we are following an approach that uses such a variable or parameter—would be needed to derive the differences found in Experiments 1, 2, and 3. Under a view where the lexical entries of PPTs have a special structure involving a judge parameter (e.g., Lasersohn, 2005; Bylinina, 2014), it seems that (under a purely semantic approach) to capture the results reported in this paper one might need to complicate the lexical entries of these adjectives.

The results reported in the present paper fit more easily with approaches to PPTs that seem more amenable to pragmatic, top-down effects stemming from the differences between the senses. One such approach has been proposed by Kennedy and Willer (2016), who did not look at sensory modalities but who make the point, more generally, that subjectivity is a highly context-sensitive, pragmatic phenomenon that "is not to be explained strictly in terms of any particular semantic parameter, implicit argument, or lexical underspecification" (p. 292). In more recent work, Willer and Kennedy (2020) suggest that experiential attitudes (resulting from the experiences that allow people to make judgements about taste and other subjective matters) can ultimately be captured in terms of a "functionalist analysis," such that "they are to be characterized in terms of the role they play in the *cognitive system* of which they are a part and, specifically, in terms of *their causal relations to sensory stimulations*, other mental states, and behavior" (p. 9, italics added). Thus, Willer and Kennedy acknowledge the need for a linguistic account of subjective expressions to interface with sensory experiences and a broader cognitive system. Moreover, in a different line of research, Rudin and Beltrama (2019, p. 96–99) sketch out an approach that argues against subjective predicates having special lexical semantics and instead emphasizes the role of world knowledge and different kinds of inferences. It seems that these kinds of approaches are, at least on a broad level, compatible with the results reported here regarding the differences between sensory modalities, which I suggest are best captured by the Inference-based Hypothesis.

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 the University of Southern California Institutional Review Board. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This material was partly based upon work supported by the National Science Foundation under Grant No. BCS-1749612.

ACKNOWLEDGMENTS

Thanks to Catherine Wang and Gwenyth Portillo-Wightman for help with stimulus creation and setting up the experiments. Earlier versions of some of this research were presented at the 2018 Annual Meeting of the Berkeley Linguistic Society, the 2018 CUNY Conference on Human Sentence Processing at UC Davis, as well as workshops on Semantic Approaches to Fiction, Literature, and Narrative at the University of Groningen and on Evaluative Meanings at the 40th Annual Conference of the German Linguistic Society (DGfS). An earlier version of some of this research is included in the Proceedings of 2018 Berkeley Linguistic Society Meeting.

REFERENCES

Abrusán, M. (2020). The spectrum of perspective shift: protagonist projection versus free indirect discourse. *Linguist. Philos.* 20:9300. doi: 10.1007/s10988-020-09300-z

Aglioti, S., and Pazzaglia, M. (2010). Sounds and scents in (social) action. *Trends Cogn. Sci.* 15, 47–55. doi: 10.1016/j.tics.2010.12.003

Aikhenvald, A. (2004). *Evidentiality*. Oxford: Oxford University Press.

Aikhenvald, A., and Storch, A. (2013). "Linguistic expression of perception and cognition: a typological glimpse," in *Perception and Cognition in Language and Culture*, eds Y. Alexandra, Y. Aikhenvald and A. Storch (Leiden: Brill), 1–46. doi: 10.1163/9789004210127

Amaral, P., Roberts, C., and Smith, E. A. (2008). Review of 'The logic of conventional implicatures' by Chris Potts. *Linguist. Philo.* 30, 707–749. doi: 10.1007/s10988-008-9025-2

Anand, P. (2009). *Kinds of Taste.* Ms: UCSC.

Anand, P., and Korotkova, N. (forthcoming). How to theorize about subjective meaning: a lesson from "de re". *Lingust. Philos*.

Athanasiadou, A. (2007). On the subjectivity of intensifiers. *Lang. Sci.* 29, 554–565. doi: 10.1016/j.langsci.2007.01.009

Baayen, R., Davidson, D., and Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Banfield, A. (1973). Narrative style and the grammar of direct and indirect speech. *Found. Lang.* 10, 1–39.

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

Beltrama, A. (2018). Totally between subjectivity and discourse. Exploring the pragmatic side of intensification. *J. Semant.* 35, 219–261. doi: 10.1093/semant/ffx021

Biber, D., and Finegan, E. (1988). Adverbial stance types in English. *Disc. Proces.* 11, 1–34. doi: 10.1080/01638538809544689

Bortolussi, M., and Dixon,. P. (2003). *Psychonarratology: Foundations for the Empirical Study of Literary Response*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511500107

Brown-Schmidt, S., and Heller, D. (2018). "Perspective taking during conversation," in *Oxford Handbook of Psycholinguistics,* eds G. Gaskell and S.A. Rueschemeyer. (Oxford: Oxford University Press), 551–574. doi: 10.1093/oxfordhb/9780198786825.013.23

Buck, C. D. (1949). *A Dictionary of Selected Synonyms in the Principal Indo-European Languages: A Contribution to the History of Ideas*. Chicago, IL: University of Chicago Press.

Bylinina, L. (2014). *The Grammar of Standards.* LOT Dissertation Series.

Caballero, R., and Paradis, C. (2015). Making sense of sensory perceptions across languages and cultures. *Funct. Lang.* 22, 1–19. doi: 10.1075/fol.22.1.01cab

Cappelen, H., and Hawthorne, J. (2009). *Relativism and Monadic Truth.* Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199560554.001.0001

Chafe, W., and Nichols, J. (1986). *Evidentiality: The linguistic Coding of Epistemology*. Norwood, NJ: Ablex.

Chmielewski, M., and Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Soc. Psychol. Personal. Sci.* 11, 464–473. doi: 10.1177/1948550619875149

Clark, H., and Gerrig, R. (1990). Quotations as demonstrations. *Language* 66, 764–805. doi: 10.2307/414729

Colavita, F. (1974). Human sensory dominance. *Percept. Psychophys.* 16, 409–412. doi: 10.3758/BF03203962

Coppock, E. (2018). Outlook-based semantics. *Linguisti. Philos.* 41, 125–164, doi: 10.1007/s10988-017-9222-y

Dubois, D. (2007). "From psychophysics to semiophysics: categories as acts of meaning. A case study from olfaction and audition back to colors," in *Speaking of Colors and Odors*, eds M. Plümacher and P. Holz (Amsterdam: Benjamins), 167–185. doi: 10.1075/celcr.8.10dub

Dziwirek, K. (2016). Smell in Polish: lexical semantics and cultural values. *J. Slavic Linguist.* 24.2. 273–301 doi: 10.1353/jsl.2016.0012

Eckardt, R. (2015). *The Semantics of Free Indirect Discourse: How Texts Allow Us to Mind-read and Eavesdrop*. Leiden/Boston, MA: Brill. doi: 10.1163/9789004266735

Elder, R. S., Schlosser, A. E., Poor, M., and Xu, L. (2017). So close I can almost sense it: the interplay between sensory imagery and psychological distance. *J. Consumer Res.* 44, 877–894. doi: 10.1093/jcr/ucx070

Evans, N., and Wilkins, D. (2000). In the mind's ear: the semantic extensions of perception verbs in Australian languages. *Language* 76, 546–592 doi: 10.2307/417135

Fludernik, M. (1993). *The Fictions of Language and the Languages of Fiction: The Linguistic Representation of Speech and Consciousness*. London: Routledge.

Gisborne, N. (2010). *The Event Structure of Perception Verbs.* Oxford University Press. doi: 10.1093/acprof:oso/9780199577798.001.0001

Glanzberg, M. (2007). Context, content, and relativism. *Philos. Stud.* 136, 1–29. doi: 10.1007/s11098-007-9145-5

Gunlogson, C., and Carlson, G. (2016). "Predicates of experience," in *Subjective Meaning: Alternatives to Relativism*, eds C. Meier and J. van Wijnbergen-Huitink (Berlin: de Gruyter), 169–200.

Hanna, J., 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, J., and Potts, C. (2009). Perspective-shifting with appositives and expressives. *Linguist. Philos.* 32, 523–552. doi: 10.1007/s10988-010-9070-5

Harris, J. A. (2012). *Processing Perspectives.* (Ph.D. thesis). University of Massachusetts Amherst, Amherst, MA, United States.

Heller, D., Grodner, D. and Tanenhaus, M. K. (2008). The role of perspective in identifying domains of references. *Cognition* 108, 831–836. doi: 10.1016/j.cognition.2008.04.008

Hinterwimmer, S. (2017). "Two kinds of perspective taking in narrative texts," in *Proceedings of Semantics and Linguistic Theory (SALT),* 27, 282–301 doi: 10.3765/salt.v27i0.4153

Hinterwimmer, S. (2019). Prominent protagonists. *J. Pragmat.* 154, 79–91. doi: 10.1016/j.pragma.2017.12.003

Hobbs, J. R. (1979). Coherence and coreference. *Cogn. Sci.* 3, 67–90. doi: 10.1207/s15516709cog0301_4

Jarvis, W. B. G., and Petty, R. E. (1996). The need to evaluate. *J. Personal. Soc. Psychol.* 70, 172–194. doi: 10.1037/0022-3514.70.1.172

Kaiser, E. (2015). "Perspective-shifting and free indirect discourse: experimental investigations," in *Proceedings of Semantics and Linguistic Theory (SALT),* 25, 346–372. doi: 10.3765/salt.v25i0.3436

Kaiser, E., and Herron Lee, J. (2017). "Experience matters: a psycholinguistic investigation of predicates of personal taste," in *Proceedings of Semantics and Linguistic Theory (SALT)*, 27, 323–339. doi: 10.3765/salt.v27i0.4151

Kaiser, E., and Herron Lee, J. (2018). "Predicates of personal taste and multidimensional adjectives: an experimental investigation," in *Proceedings of the 35th West Coast Conference On Formal Linguistics (WCCFL 35)* (Somerville, MA: Cascadilla Proceedings Project) 224–231.

Karttunen, L., and Zaenen, A. (2005). "Veridicity," in *Annotating, Extracting and Reasoning About Time and Events, No. 05151 in Dagstuhl Seminar Proceedings*,

eds G. Katz, J. Pustejovsky and F. Schilder (Dagstuhl: Internationales Begegnungs- und Forschungszentrum für Informatik).

Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI Publications.

Kehler, A., and Rohde, H. (2013). A Probabilistic reconciliation of coherence-driven and Centering-driven theories of pronoun interpretation. *Theoreti. Linguist.* 39, 1–37. doi: 10.1515/tl-2013-0001

Kennedy, C., and Willer, M. (2016). "Subjective attitudes and counterstance contingency," in *Proceedings of Semantics and Linguistic Theory (SALT) 26*, 913–933. doi: 10.3765/salt.v26i0.3936

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P., Jewell, R., and Winter, N. (2020). The shape of and solutions to the MTurk quality crisis. *Polit. Sci. Res. Methods* 8, 614–629. doi: 10.1017/psrm.2020.6

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

Klages, J., Holler, A., Kaiser, E., and Weskott, T. (2020). Commonalities and differences in the interpretation of predicates of personal taste vs. relational locative expressions: some theoretical considerations and experimental evidence. *Open Lib. Human.* 6, 1–30. doi: 10.16995/olh.513

Koelbel, M. (2004). Faultless disagreement. *Proc. Aristotelian Soc.* 104, 55–73. doi: 10.1111/j.0066-7373.2004.00081.x

Kopytko, R. (1990). Verbs of sensory cognition: a contrastive analysis of a lexical field in the lexicon of Polish and English. *Contrastive Linguist.* 25, 59–70.

Korotkova, N. (2016). *Heterogeneity and Uniformity in the Evidential Domain*. (PhD dissertation) UCLA, Los Angeles, California, United States.

Korsmeyer, C. (1999). *Making Sense of Taste : Food and Philosophy*. Ithaca, NY: Cornell University Press.

Krifka, M. (2010). "A note on the asymmetry in the hedonic implicatures of olfactory and gustatory terms," in *Between the Regular and the Particular in Speech and Language,* eds S. Fuchs, P. Hoole, C. Mooshammer and M. Zygis (Frankfurt am Main: Peter Lang), 235–245.

Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguisti. Philos.* 28, 643–686. doi: 10.1007/s10988-005-0596-x

Lasersohn, P. (2007). Expressives, perspective and presupposition. *Theoret. Linguist.* 33, 223–230. doi: 10.1515/TL.2007.015

Lenth, R. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.0*. Available online at: https://CRAN.R-project.org/package=emmeans

Levinson, S. C., and Majid, A. (2014). Differential ineffability and the senses. *Mind Lang.* 29, 407–427. doi: 10.1111/mila.12057

Lynott, D., and Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behav. Res. Methods* 41, 558–564. doi: 10.3758/BRM.41.2.558

Lynott, D., Connell, L., Brysbaert, M., Brand, J., and Carney, J. (2020). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behav. Res. Methods* 52, 1271–1291. doi: 10.3758/s13428-019-01316-z

Maier, E. (2015). Quotation and unquotation in free indirect discourse. *Mind Lang.* 30, 345–373. doi: 10.1111/mila.12083

Majid, A. (2021). Human olfaction at the intersection of language, culture, and biology. *Trends Cogn. Sci.* 25, 111–123. doi: 10.1016/j.tics.2020.11.005

Majid, A., Burenhult, N., Stensmyr, M., De Valk, J., and Hansson, B. S. (2018). Olfactory language and abstraction across cultures. *Philos. Trans. Royal Soc. London Ser. B Biol. Sci.* 373:20170139. doi: 10.1098/rstb.2017.0139

Markus, H., and Zajonc, R. B. (1985). "The cognitive perspective in social psychology," in *Handbook of Social Psychology, 3rd Edn.* eds G. Lindzey and E. Aronson (New York, NY: Random House), 137–229.

McHale, B. (1978). Free indirect discourse: a survey of recent accounts. *J. Descript. Poetics Theory Literature* 3, 249–278.

McNally, L., and Stojanovic, I. (2017). "Aesthetic adjectives," *The Semantics of Aesthetic Judgments*, ed J. Young (Oxford: Oxford University Press), 17–37. doi: 10.1093/acprof:oso/9780198714590.003.0002

Moltmann, F. (2010). Relative truth and the first person. *Philos. Stud.* 150, 187–220. doi: 10.1007/s11098-009-9383-9

Moore, C., and Dunham, P. (1995). *Joint Attention: Its Origins and Role in Development*. Hillsdale, NJ: Erlbaum.

Ninan, D. (2014). "Taste predicates and the acquaintance inference," in *Proceedings of Semantics and Linguistic Theory (SALT) 24*, 290–309. doi: 10.3765/salt.v24i0.2413

Palmer, S. (1999). *Vision Science: Photons to Phenomenology*. Cambridge: Bradford.

Patel-Grosz, P. (2012). *(anti-)Locality at the Interfaces*. (PhD dissertation), MIT, Cambridge, MA, United States.

Pearson, H. (2013). A judge-free semantics for predicates of personal taste. *J. Semant.* 30, 103–154. doi: 10.1093/jos/ffs001

Phillips, M. L., and Heining, M. (2002). "Neural correlates of emotion perception: from faces to taste," in *Olfaction, Taste, and Cognition,* eds C. Rouby, B. Schaal, D. Dubois, R. Gervais and A. Holley (Cambridge: Cambridge University Press), 196–208. doi: 10.1017/CBO9780511546389.018

Potts, C. (2007). The expressive dimension. *Theoret. Linguist.* 33, 165–198. doi: 10.1515/T. L.2007.011

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Redeker, G. (1996). "Free indirect discourse in newspaper reports," in *Linguistics in the Netherlands 1996*, eds C. Cremers and M. den Dikken (Amsterdam: Benjamins), 221–232. doi: 10.1075/avt.13.21red

Rhee, S. (2016). On the emergence of the stance-marking function of English adverbs: a case of intensifiers. *Linguist. Res.* 33, 395–436 doi: 10.17250/khisli.33.3.201612.003

Royet, J. P., Zald, D., Versace, R., Costes, N., Lavenne, F., Koenig, O., et al. (2000). Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: a positron emission tomography study. *J. Neurosci.* 20, 7752–7759. doi: 10.1523/JNEUROSCI.20-20-07752.2000

Rudin, D., and Beltrama, A. (2019). "Default agreement with subjective assertions," in *Proceedings of Semantics and Linguistic Theory (SALT)* 29, 82–102. doi: 10.3765/salt.v29i0.4597

Sæbø, K. J. (2009), Judgment ascriptions. *Linguist. Philos.* 32, 327–352. doi: 10.1007/s10988-009-9063-4

Salem, S., Weskott, T., and Holler, A. (2015). "Does narrative perspective influence the spatial point of view in readers' mental representation of a text? – a psycholinguistic study on free indirect discourse," talk presented at the *Workshop on Perspective Taking, Annual Conference of the German Linguistic Society (DGfS)*. Leipzig.

San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M. et al. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cogn. Linguist.* 26, 31–60 doi: 10.1515/cog-2014-0089

Schlenker, P. (2004). Context of thought and context of utterance (a note on free indirect discourse and the historical present). *Mind Lang.* 19, 279–304. doi: 10.1111/j.1468-0017.2004.00259.x

Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguist. Philos.* 31, 353–395. doi: 10.1007/s10988-008-9039-9

Sinnett, S., Spence, C., and Soto-Faraco, S. (2007). Visual dominance and attention: the Colavita effect revisited. *Percept. Psychophys.* 69.673–686. doi: 10.3758/BF03193770

Snyder, E. (2013). Binding, genericity, and predicates of personal taste. *Inquiry* 56, 278–306. doi: 10.1080/0020174X.2013.784485

Solt, S. (2018). "Multidimensionality, subjectivity and scales: experimental evidence," in *The Semantics of Gradability, Vagueness, and Scale Structure: Experimental Perspectives*, eds E. Castroviejo, L. McNally and G. Weidman Sassoon (Cham: Springer), 59–91. doi: 10.1007/978-3-319-77791-7_3

Spence, C. (2009). Explaining the Colavita visual dominance effect. *Progr. Brain Res.* 176, 245–258. doi: 10.1016/S0079-6123(09)17615-X

Stephenson, T. (2007). Judge dependence, epistemic modals, and predicates of personal taste. *Linguist. Philos.* 30, 487–525. doi: 10.1007/s10988-008-9023-4

Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620904

Viberg, Å. (1984). The verbs of perception: a typological study. *Linguistics* 21, 23–62. doi: 10.1515/ling.1983.21.1.123

Viberg, Å. (2001). "Verbs of perception," in *Language Typology and Language Universals*, eds M. Haspelmath, E. König, W. Oesterreicher and W. Raible (Berlin: de Gruyter), 1294–1309.

Waksler, R. (2012). "Super, uber, so, and totally: over-the-top intensification to mark subjectivity in colloquial discourse," in *Subjectivity in Language and*

Discourse, eds N. Baumgarten, I. Du Bois, and J. House (Leiden:Brill), 15–32. doi: 10.1163/9789004261921_003

Wang, L., Reese, B., and McCready, E. (2005). The projection problem of nominal appositives. *Snippets* 10, 13–14. Available online at: https://www.ledonline.it/snippets/allegati/snippets10005.pdf

Willer, M. and Kennedy, C. (2020). Assertion, expression, experience. *Inquiry*. doi: 10.1080/0020174X.2020.1850338

Willett, T. (1988). A cross-linguistic survey of the grammaticalization of evidentiality. *Stud. Lang.* 12, 51–97. doi: 10.1075/sl.12.1.04wil

Winter, B. (2016). Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon. *Lang. Cogn. Neurosci.* 31, 975–988. doi: 10.1080/23273798.2016.1193619

Yeshurun, Y., and Sobel, N. (2010). An odor is not worth a thousand words: from multidimensional odors to unidimensional odor objects. *Ann. Rev. Psychol.* 61, 219–241. doi: 10.1146/annurev.psych.60.110707.163639

Zakkou, J. (2019). Denial and retraction: a challenge for theories of taste predicates. *Synthese* 196, 1555–1573. doi: 10.1007/s11229-017-1520-y

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 Kaiser. 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.

APPENDIX

The target stimuli used in the experiments. The adverbial intensifiers shown in parentheses were only used in Experiment 2.

1. Right after I entered the room, Emily {saw/smelled/tasted/put} the salad in the bowl. It {looked/smelled/tasted/was} (extremely) revolting.
2. After I entered the room, Tim {saw/smelled/tasted/put} the cookie on the shelf. It {looked/smelled/tasted/was} (really) dreadful.
3. When I got home, Joe {saw/smelled/tasted/put} the fish in the pan. It {looked/smelled/tasted/was} (absolutely) horrifying.
4. When I entered the room, Jane {saw/smelled/tasted/put} the pretzel in the bowl. It {looked/smelled/tasted/was} (truly) dreadful.
5. As soon as I stepped into the room, Alex {saw/smelled/tasted/put} the baked potato on the counter. It {looked/smelled/tasted/was} (truly) delicious.
6. Right after I walked in, Lisa {saw/smelled/tasted/put} the cake on the table. It {looked/smelled/tasted/was} (absolutely) amazing.
7. Right after I arrived, Tina {saw/smelled/tasted/put} the baguette on the rack. It {looked/smelled/tasted/was} (really) interesting.
8. As soon as I walked in, Luke {saw/smelled/tasted/put} the donut in the basket. It {looked/smelled/tasted/was} (extremely) pleasant.
9. After I arrived at the party, Kevin {saw/smelled/tasted/put} the cupcake in the box. It {looked/smelled/tasted/was} (incredibly) enticing.
10. As soon as I arrived, Charlotte {saw/smelled/tasted/put} the pancake on the plate. It {looked/smelled/tasted/was} (totally) delightful.
11. As soon as I came home, Amanda {saw/smelled/tasted/put} the quesadilla in the skillet. It {looked/smelled/tasted/was} (truly) amazing.
12. As soon as I went into the room, Chris {saw/smelled/tasted/put} the cookie in the jar. It {looked/smelled/tasted/was} (extremely) interesting.
13. When I came into the room, Eliza {saw/smelled/tasted/put} the muffin on the platter. It {looked/smelled/tasted/was} (really) disgusting.
14. After I got home, Billy {saw/smelled/tasted/put} the hamburger on the tin foil. It {looked/smelled/tasted/was} (extremely) unpleasant.
15. When I returned home after work, George {saw/smelled/tasted/put} the pork in the pot. It {looked/smelled/tasted/was} (totally) horrifying.
16. When I arrived, Sarah {saw/smelled/tasted/put} the ice cream in the bowl. It {looked/smelled/tasted/was} (really) tantalizing.
17. After I got home from school, Michael {saw/smelled/tasted/put} the stir-fry in the dish. It {looked/smelled/tasted/was} (very) pleasant.
18. As soon as I came to the party, Todd {saw/smelled/tasted/put} the biscuit on the stove. It {looked/smelled/tasted/was} (absolutely) delightful.
19. As soon as I walked into the room, Mary {saw/smelled/tasted/put} the pizza in the tupperware. It {looked/smelled/tasted/was} (very) tantalizing.
20. When I arrived at the gathering, Ted {saw/smelled/tasted/put} the hot dog on the tray. It {looked/smelled/tasted/was} (really) enticing.
21. When I returned to the living room, Michelle {saw/smelled/tasted/put} the popcorn in the container. It {looked/smelled/tasted/was} (really) delicious.
22. After I entered the office, Bob {saw/smelled/tasted/put} the bagel on the counter. It {looked/smelled/tasted/was} (truly) disgusting.
23. Right after I arrived at the tea party, Linda {saw/smelled/tasted/put} the brownie on the baking sheet. It {looked/smelled/tasted/was} (absolutely) revolting.
24. Right after I entered, Denise {saw/smelled/tasted/put} the pasta in the pot. It {looked/smelled/tasted/was} (incredibly) unpleasant.

frontiers in Communication

ORIGINAL RESEARCH
published: 18 August 2021
doi: 10.3389/fcomm.2021.612733

Epistemic Perspectives and Communicative Acts

*Anton Benz**

Leibniz-Centre General Linguistics, Berlin, Germany

Searle (Speech Acts, 1969) introduced his famous distinction between constitutive and regulative rules that together define felicity conditions of speech acts. Regulative rules are normative rules, whereas constitutive rules define what counts as a performance of a speech act. In this paper we demonstrate with the example of assertions and referential uses of definite description that simple regulative rules can be given to speech acts that hold only on a core of well-behaved utterance situations. From this core, extended uses can be derived based on epistemic paths that are defined by the epistemic perspectives of speaker and hearer. As the use of speech acts get extended to a wider class of utterance situations, conflicts with the constitutive rules can emerge. We show that the extended uses are nevertheless felicitous. We represent epistemic relations in a possible worlds framework, and take an interactional approach that considers speech acts as part of joint communicative acts.

Keywords: speech acts, common knowledge, epistemic perspective, referential acts, assertions

OPEN ACCESS

Edited by:
Petra Hendriks,
University of Groningen, Netherlands

Reviewed by:
Jonathan Ginzburg,
Université Paris Diderot, France
Robert Van Rooij,
University of Amsterdam, Netherlands

***Correspondence:**
Anton Benz
benz@leibniz-zas.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *30 September 2020*
Accepted: *18 June 2021*
Published: *18 August 2021*

Citation:
Benz A (2021) Epistemic Perspectives and Communicative Acts. Front. Commun. 6:612733. doi: 10.3389/fcomm.2021.612733

1 INTRODUCTION

Pragmatics is often defined as the study of language in context (see Korta and Perry, 2020, Sec. 4), and, in particular, the study of *the relation of signs to interpreters* (Morris, 1938, p. 6). *Context* is a multifaceted concept that includes, among other things, the physical environment, social relations, the dialogue history, and epistemic states of the interlocutors. In this article, we address the dependencies between felicity conditions of communicative acts and the epistemic relations between interlocutors, i.e., their knowledge about each other and the facts of the world. Central to our approach will be the assumption that communicative acts are organized as action–response pairs (*joint projects,* Clark, 1996) that need to be coordinated between speaker and hearer.

Suppose one undertakes it to define felicity conditions of, for example, the speech act of *asserting*, then the question arises whether the requirement that speakers know that p, if they assert p, is part of the definition of the speech act, or merely a normative rule imposed by general requirements about cooperative communication (Grice, 1975). Searle (1969) introduced the famous distinction between *regulative* and *constitutive* rules that govern the use of speech acts. Constitutive rules are defining rules that say which linguistic utterances count as performances of a certain speech act type. Regulative rules are normative rules that say how a speech act *should* be performed. In Searle's classification, regulative rules include constraints on the speaker's or hearer's information state. For example, the act of asserting is subject to the regulative rule that speakers must believe what they assert to be true (Searle, 1969, p. 66). As constitutive rules state requirements particular of certain speech acts, and regulative rules general requirements of rational behavior, one would like constitutive requirements to be weak, and regulative rules to be powerful and applicable to as wide a range of speech acts as possible. In this article, we concentrate on two communicative acts that seem, at first, little related to each other: the illocutionary act of *asserting* and the locutionary act of

referring to a specific object with a definite description. We assume that they are constituted by the following minimal rules:

1) **Assertion.** The utterance of a sentence expressing proposition φ is a legitimate communicative act given the state of affairs represented by a model m if, and only if φ is true in m.
2) **Referential definite description.** The utterance of a definite description '*the* φ' with the aim of referring to a referent r is a legitimate communicative act given the state of affairs represented by a model m if, and only if $\varphi(r)$ is true in m.

Clearly, these rules on their own cannot guarantee the felicity of their communicative acts. Clark and Marshall (1981) argued that successful referential uses of definite descriptions require that $\varphi(r)$ is common knowledge between speaker and hearer. Furthermore, there should only be one object for which $\varphi(r)$ holds. For assertions the speaker should know that φ holds, and the hearer should believe it at least possible. Otherwise, asserting φ may lead to false beliefs, or fail to convince the hearer of the truth of φ. However, the speaker may assert φ exactly for the purpose of creating such a false belief, and the hearer may see through it and keep quiet. Assuming that '*tell*' refers to an act of assertion, the following examples seem to be in conflict with the constitutive rule (1).

3) Leo told me that it is snowing in the Alps, but I knew that the snowing had stopped.
4) Leo told me that it is snowing in the Alps, but I knew that she is lying.

If (1) is correct, then no utterance of a sentence with propositional content φ should count as an assertion if φ is false. However, in (3) it seems fine to report that someone (Leo) asserted a proposition φ (snowing in the Alps) although the person reporting this act knows that φ is false. Example (4) shows that an utterance with meaning φ can be reported as an assertion even if the person uttering it is known to disbelieve φ. This shows that constitutive rules cannot be understood as semantic meaning components of reported assertions such that '*A told B that* φ.' would mean that *there is an event e which is an utterance event with speaker A and addressee B and propositional content* φ *for which rule* (1) *holds*. Nevertheless, the constituting rules must play some role in reported utterance events.

With Searle, we assume that constitutive and regulative rules define speech acts as social institutions. They are a form of conventional linguistic behavior. We postulate that this behavior is defined for a core of perfect communicative situations in which interlocutors can entertain only true beliefs and are assumed to be fully cooperative. The constitutive rules only apply here. From this core, communicative acts are extended to more complex and possibly non-cooperative utterance situations via epistemic paths that involve changing perspectives between interlocutors. For example, in (3) the speaker S who reports Leo's utterance believes that from Leo's perspective constitutive rule (1) is satisfied, and, hence, that the utterance can be called an assertion from Leo's perspective. We assume that the path from S's to Leo's perspective allows S to call Leo's utterance an '*assertion*', although the constitutive rule (1) is violated from S's own perspective. In (4), the constitutive rule (1) is violated from both the speaker S's perspective and from Leo's perspective. However, Leo must think that from S's perspective it is satisfied. Hence, it is the path from S to Leo to S that allows S, or us as readers of (4), to classify Leo's utterance as assertion. However, paths can be more convoluted than suggested by (3) and (4) alone. Suspicions may introduce circular paths and mutual mistrust in the validity of constitutive rules. We show also for these situations how epistemic paths can justify the classification of utterances as assertions.

We present a theory that explains how epistemic paths can give rise to felicitous joint communicative acts that extend beyond the epistemic core of perfect utterance situations. In contrast to Searle, we take an interactional perspective on speech acts (see Clark, 1996) where speaker and hearer have each to perform their own required act: the speaker performs an utterance act that is followed by an appropriate response of the hearer. We introduce two epistemic felicity constraints that decide whether a joint communicative act is consistent with the interlocutor's beliefs: a *licensing* constraint and a *uniqueness* constraint. *Licensing* requires that the joint act is possible from the interlocutor's perspectives, and *uniqueness* that the hearer's response is uniquely determined by the speaker's utterance act. We will see how the constraints eliminate infelicitous communicative acts when joint acts are extended to new epistemic situations.

In the next section, we present a general format for the representation of constitutive rules for speech acts. We then consider referential uses of definite descriptions in more detail and demonstrate how epistemic paths allow extended uses outside the communicative core situations. In particular, we consider the examples discussed by Clark and Marshall (1981) that are supposed to show that felicitous references to an object r with *the* φ require that $\phi(r)$ is common knowledge. We show that this has only to be true for communicative core situations. In **Section 5**, we introduce the formal model. We represent utterance situations and epistemic states in a possible worlds framework of knowledge and belief (e.g., Hintikka, 1962; Barwise, 1989; Fagin et al., 1995; Gerbrandy, 1998; Baltag et al., 2008), building up, in particular, on (Benz, 2008, 2012). We construct the class of situations in which referential uses of definite descriptions are felicitous, first for the core situations that satisfy common knowledge of true beliefs and cooperativity, then for situations that show an internal hierarchical structure in which the utterance situation is connected to a core situations only via epistemic paths. Throughout, we discuss examples of assertions and referential uses of definite descriptions in parallel. Finally, in **Section 6**, we return to the introductory examples, and discuss wider ramifications of the proposed account for speech act theory.

2 REPRESENTING CONSTITUTIVE RULES AS JOINT PROJECTS

Searle (1969, Sec. 2.5) illustrates the difference between constitutive and regulative rules with the rules of Chess. The

rules of Chess are a paradigmatic example of constitutive rules, the main purpose of which it is to define what counts as a move of the game. In addition to constitutive rules there may also be regulative rules, for example, that the players should not smoke and abstain from distracting behavior. However, these rules do not define chess. As an example of a constitutive rule, Searle (1969, p. 34) cites the rules for *checkmate*. In general, these rules take into account only the position of pieces on the chess board. Some rules may also take into account the game history, for example, the rule of *castling*. For example, moving the White King from his start position[1] two squares to the right and the Rook from its start position to the left of the King counts as legal chess move called *castling kingside* if King and Rook had not moved before, none of the squares between them are occupied, and the King does not move out of, through, or into check. If this rule that defines the legal move of kingside castling in chess were given to a program that checks the moves of players, then any violation of its conditions would mean that the program would reject the move as a move of chess. Nevertheless, we can, without contradiction, make statements as in (5) and (6), which are analogous to (3) and (4).

5) Leo castled kingside, but I knew that the King had moved before.
6) Leo castled kingside, but I knew that she is cheating.

As in the case of speech acts, exploiting the different perspectives of people involved can explain why one can call a move '*castling*' although it violates its defining rules. In (5), the move may seem legal from Leo's perspective, or from the perspective of an observer who does not know the history of the game. In (6), the move may seem legal from an outside observer's perspective, or the violation may go unnoticed from the opponent's perspective. Also in (7), the speaker can describe what he did as castling kingside.

7) I castled kingside. Luckily, my opponent didn't remember that the King had moved before.

The speaker could not say '*I moved the pieces as if I castled kingside,*' or '*I pretended to castle kingside.*' He has to say that he castled kingside, although one could say that he pretended to perform a legal move.

There seem to be the same pragmatic mechanisms at work that widen the meaning of '*castling*' and the meaning of '*asserting*'. However, playing chess differs in important respects from conversation. Chess is a game with strictly opposed players, whereas we assume with Grice (1975) that an unmarked conversational situation is one where speakers and hearers are cooperative. Chess is a game without private information, i.e., whatever happens in the game as well as the positions of the pieces on the board are shared knowledge between players. In a typical dialogue situation the knowledge of speaker and hearer differ. And performing a certain speech act, for example, *asserting*, requires the speaker to have more knowledge than the hearer. A further difference is that chess games can be described as sequences of moves by the White and Black players. It has been argued forcibly by, for example, Clark (1996) that conversation is a sequence of joint coordinated actions, i.e., that each communicative act performed by the speaker needs a corresponding communicative act on the hearer side to be completed. These pairs of communicative acts have been called '*joint projects*' Clark (1996).[2] We follow this line of research and represent communicative acts as triples consisting of a model m, a communicative act a performed by the speaker, and a response act r by the hearer. Hence, each joint project is a set of triples $\langle m, a, r\rangle$. We call the triples joint communicative acts. For assertions, we assume that the speaker's act is an utterance of a sentence s with some propositional content φ, and that the hearer reacts with a grounding act that updates the common ground with the fact that the speaker asserted φ. The constitutive rule (1) for assertions then translates into the following representation (8).

8) **Assertion.** Let $\mathcal{M}$ be a set of models, $\mathcal{L}$ a set of sentences of a given language, and Φ a set of logical forms for sentences of $\mathcal{L}$. Asserting sentence s with propositional content φ in situation m is a legal communicative act if φ is is true in m. We identify the joint project of asserting s with meaning φ with $p_{ass}(s, \varphi) = \{\langle m, s, \varphi\rangle \mid m \in \mathcal{M} \text{ and } m \models \varphi\}$.

Note that each sentence s defines its own joint project. Hence, the classification into joint projects is more fine-grained than the classification into speech acts. This is also true of the following representation of referential uses of definite descriptions. We assume that each pair consisting of a description *the* φ and a referent r define their own joint project.

9) **Referential definite description.** Let $\mathcal{M}$ be a set of models, and Φ a set of logical forms. The utterance of a definite description *the* φ with intended referent r in a situation m is a legal communicative act if $\varphi(r)$ is is true in m. We identify the joint project of referring to object r with definite description *the* φ with $p_{def}(\textit{the}\ \varphi, r) = \{\langle m, \textit{the}\ \varphi, r\rangle \mid m \in \mathcal{M} \text{ and } m \models \varphi(r)\}$.

The constitutive rules have to be accompanied by regulative rules. Together they define the felicity conditions of a speech act. Here, we only consider felicity conditions that pertain to the epistemic perspectives of speaker and hearer. As mentioned before, we consider two constraints called *licensing* and *uniqueness*.

10) **Licensing.** Let p be a given joint project. An utterance act a is epistemically licensed for the speaker, if from the speaker's perspective doing a can initiate the joint project p in all

[1]The chess board is an 8×8-square with columns named a to h, and rows numbered 1 to 8. The White King's start position is e1, and the Kingside Rook's start position h1.

[2]'*Joint projects*' can be seen as a generalization of the notion of *adjacency pairs* (Schegloff and Sacks, 1973).

possible state of affairs m and for all possible belief states of the hearer. An utterance act a is epistemically licensed for the hearer, if from the hearer's perspective doing a can initiate the joint project p in at least one possible state of affairs m and for at least one belief states of the speaker.

11) **Uniqueness.** Let p be a given joint project. An utterance act a satisfies the uniqueness condition for p from the speaker's or hearer's perspective, if it holds for all their possible states of affairs m for which doing a can initiate any joint project that a leads to the same hearer response r such that the joint communicative act $\langle m, a, r\rangle$ belongs to p.

If we ask for a deeper reason for these constraints, then the answer is the requirement that interlocutors should not gamble. If *licensing* is violated, then the speaker believes that the attempted speech act may fail. For the hearer, a violation of *licensing* means that the speaker's utterance act is inconsistent with the hearer's beliefs. If *uniqueness* is violated, then it is unclear how to respond to the utterance act. This complete ban on gambling may be too strict a requirement for realistic utterance situations, but, as a logical idealization, it is appropriate for our purposes.[3]

3 THE REFERENTIAL USE OF DEFINITE DESCRIPTIONS

In this section we consider referential uses of definite descriptions. As mentioned before, there are two closely related problems about the interpretation of communicative acts: the classification problem and the meaning problem. The classification problem stems from the fact that utterances and uses of definite descriptions can be classified as assertions or referential uses although their constitutive conditions are not satisfied. Referential uses of definite descriptions provide examples that are particularly suitable for studying the role of epistemic paths in the classification problem.

In their influential study, Clark and Marshall (1981, C&M) discuss a series of examples that show that reference to an object r with definite description *the* φ can fail although any finite sequence of conditions *the speaker beliefs that* $\varphi(r)$, *the hearer believes that* $\varphi(r)$, *the speaker believes that the hearer believes that* $\varphi(r)$, *the hearer believes that the speaker believes that* $\varphi(r)$, *etc* are true. C&M concluded that successful referential uses of definite description require *common knowledge* of $\varphi(r)$. The relevant examples consist of short stories about two protagonists who read the early edition of a newspaper together and discuss the fact that it says that *A Day at the Races*, a movie with the Marx Brothers, is showing that night at the local cinema Roxy. Then, one of the protagonists, or both learn individually that the movie has been changed to *Monkey Business*. The stories always end with one of the protagonists asking '*Have you ever seen the movie showing at the Roxy tonight?*' The question is then whether the definite description '*the movie showing at the Roxy tonight*' refers to *Monkey Business*. Version 4 of their examples reads as follows:

> On Wednesday morning Ann and Bob read the early edition of the newspaper and discuss the fact that it says that *A Day at the Races* is playing that night at the Roxy. Later, Ann sees the late edition, notes that the movie has been corrected to *Monkey Business*, and marks it with her blue pencil. Still later, as Ann watches without Bob knowing it, he picks up the late edition and sees Ann's pencil mark. That afternoon, Ann sees Bob and asks, "Have you ever seen the movie showing at the Roxy tonight?" (Clark and Marshall, 1981, p. 13, Version 4)

Here, Bob must reason as follows: Ann knows that *Monkey Business* is playing tonight. But she thinks I believe that we both are mutually convinced that *A Day at the Races* is showing. So she must think that I think she refers to *A Day at the Races*. Hence, knowing that *Monkey Business* is showing, and knowing that the speaker knows that *Monkey Business* is showing is not enough to ensure successful reference to *Monkey Business*.

More and more complicated examples can be constructed that show that any finite sequence of sentences '*Ann knows that Bob knows that . . . that Monkey Business is showing that night*' is not enough to ensure reference to *Monkey Business*. Clark & Marshall arrive at the conclusion that both participants need to know that all sentences of the form (12) have to be true in order to secure reference to *Monkey Business*.

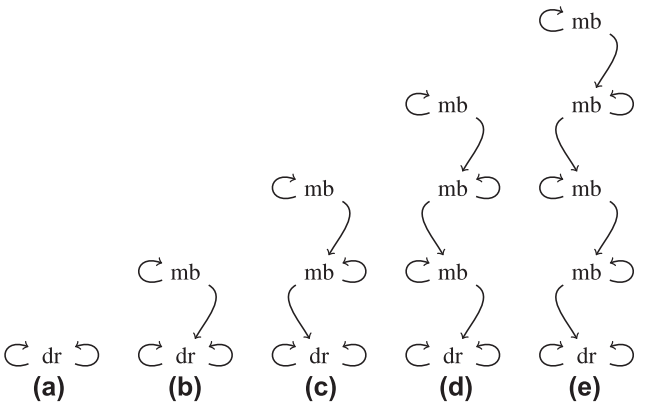
12) X_1 knows that X_2 knows that X_3 knows that . . . that X_n knows that *Monkey Business* is showing tonight.

Here, the X_i's are dialogue participants, and n is any natural number. This condition is equivalent to: It must be *common knowledge* that *Monkey Business* is showing. Common knowledge of $\varphi(r)$ entails all sentences of (12). **Table 1** shows graphical representations of the different epistemic states considered by C&M. Underlying is a possible worlds representation of beliefs, which will be defined in **Section 5**.

We are interested in the question: What does the definite description *the* φ = '*The movie showing at the Roxy tonight?*' actually refer to? Each of C&M's scenarios starts with Ann and Bob reading together that *A Day at the Races* is showing. This initial epistemic state is represented by a) in **Table 1**. We can distinguish a reading that is based on *public* information, and one that is based on *private* information. In the a) and the b) situation, where Bob thinks to be in situation a), *the* φ obviously refers to *A Day at the Races*, which is based on shared *public* belief. In situation c), the answer is not as clear cut. Bob may answer '*No, I've never seen A Day at the Races. But, you know, the program has been corrected. Monkey Business is showing.*', because he thinks that Ann thinks that it is public knowledge that *A Day at the Races* is showing. Bob

[3]This does not mean that we consider ambiguity and lack of understanding to be minor dialogue problems. The *uniqueness* constraint is a regulatory, i.e., normative constraint. Normative rules can be violated with or without intend. Crucial for us is that their logical consequences can be studied without considering repair strategies that apply in case of violations.

TABLE 1 | Different information states considered by Clark and Marshall (1981, C&M, pp. 11–14). Abbreviations: dr: *A Day at the Races* is showing, mb: *Monkey Business* is showing. Each node represents a possible world. Arrows $w \rightarrow v$ say that v is an epistemic possibility in world w; arrows to the left point to possible worlds of the speaker's information state ('Ann'), arrows to the right to possible worlds of the hearer's information state ('Bob'). b) is Version 2 of C&M; c) is C&M's Version 3; d) is their Version 4; and e) is their Version 5.



may also answer '*Yes, I have. You know, the program has been corrected and Monkey Business is showing. I saw the movie last year on TV*.' This interpretation of the definite description is based on Bob's *private* beliefs about which movie is showing. This reading involves a repair, as Bob must think that Ann will, at first, interpret the Yes-answer as a confirmation of the proposition that Bob has seen *A Day at the Races*. We are only interested in the interpretation based on public information, that does not involve a repair. In C&M's more complex scenarios, the two readings seem both to be available. We, therefore, modify C&M's examples in a way that favors the *public* reading. The modified examples show that the public reading is available although the conditions about beliefs in (12) may be violated for arbitrarily large n.

In the following scenarios, the question is always what is the referent of *the φ = the movie showing at the Roxy tonight*? Version 1 in (13) is the basic scenario in which $\varphi(mb)$ is common knowledge between Ann and Bob. The epistemic states are represented graphically in **Table 2**.

13) **Version 1.** On Wednesday morning Ann and Bob read the newsletter on Ann's computer and discuss the fact that it says that *Monkey Business* is playing at the Roxy that night. Later Ann decides that she wants to stay at home. She calls Bob and asks, '*Do you want to watch the movie showing at the Roxy tonight on Netflix with me?*'

In the next version, the beliefs of Ann and Bob have not changed, but the truth of $\varphi(mb)$ is not given.

14) **Version 2.** On Wednesday morning Ann and Bob read the newsletter on Ann's computer and discuss the fact that it says that *Monkey Business* is playing at the Roxy that night. Later, a correction was sent saying that, in fact, *A Day at the races* is playing. Neither Ann nor Bob notice the correction. Later Ann decides that she wants to stay at home. She calls Bob and asks, '*Do you want to watch the movie showing at the Roxy tonight on Netflix with me?*'

In Version 3, Ann learns that $\varphi(dr)$, but Bob's beliefs are unchanged. This does not block the reference to *Monkey Business*.

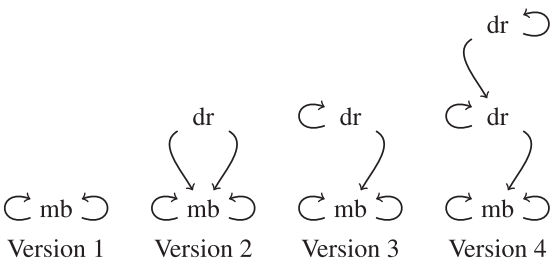
15) **Version 3.** On Wednesday morning Ann and Bob read the newsletter on Ann's computer and discuss the fact that it says that *Monkey Business* is playing at the Roxy that night. Later, a correction was sent saying that, in fact, *A Day at the races* is playing. Only Ann notices the correction. She doesn't like *A Day at the races*. She knows that Bob would love to see it, but that he couldn't have noticed the correction. She calls Bob and asks, '*Do you want to watch the movie showing at the Roxy tonight on Netflix with me?*'

In Version 4, both Ann and Bob learn that $\varphi(dr)$, and Bob learns that Ann learns it. Again, this does not block reference to *Monkey Business*.

16) **Version 4.** On Wednesday morning Ann and Bob read the newsletter on Ann's computer and discuss the fact that it says that *Monkey Business* is playing at the Roxy that night. Later, a correction was sent saying that, in fact, *A Day at the races* is playing. Ann notices the correction. Later, Bob reads her email and notices the correction, and notices also that Ann has read it. Bob would love to see *A Day at the races* but he knows that Ann doesn't like it at all. He wants to please her, but doesn't want her to know that he reads her mail without her knowing it, he calls Ann and asks, '*Do you want to watch the movie showing at the Roxy tonight on Netflix with me?*'

In this manner, more and more complex epistemic states can be created in which it holds that X_1 knows that X_2 knows that X_3 knows that . . . that X_n knows that *A Day at the Races* is showing, and, hence, in which X_1 knows that X_2 knows that . . . that X_n knows that *Monkey Business* is not showing.

TABLE 2 | Different information states considered in (13)–(16). Abbreviations: dr: *A Day at the Races* is showing, mb: *Monkey Business* is showing. Each node represents a possible world. Arrows $w \rightarrow v$ say that v is an epistemic possibility in world w; arrows to the left point to possible worlds of the speaker's information state, arrows to the right to possible worlds of the hearer's information state.



Nevertheless, reference to *Monkey Business* is possible. This leads to te following paradox: Clark and Marshall (1981) showed that successful reference to a referent r requires that it is common knowledge that $\varphi(r)$ holds; common knowledge of $\varphi(r)$ entails all conditions of the form (12); however, the procedure for constructing (13)–(16) shows that all these sentences can be false, and, still, the referential act can be successful. How is this possible?

The graphs in **Tables 1** and **2** point to a solution. The complex states constructed by C&M and by us embed a basic situation in which common knowledge of $\varphi(r)$ is satisfied. In this basic situation, *the* φ refers to r on the basis of public information. The interpretation of *the* φ as r then travels upwards along epistemic paths to the real situation, and licenses this interpretation although the constituting rules are not satisfied. What if we ask about the deeper reason for the interpretation's ability to travel along epistemic paths? If we consider Version 2, then we see that the situation is *indistinguishable* from Version 1 from the perspective of both interlocutors. Hence, the interlocutors should behave identically in both situation. In Version 3, the situation is distinguishable from the core situation in Version 1, but it is still *indistinguishable* for the hearer. This allows the speaker to behave identically in both situations. In Version 4, both interlocutors can see that the situation is different from that of Version 1. However, for the speaker it is *indistinguishable* from that of Version 3, which immediately explains why she behaves identically in both situations, and why the hearer can behave identically in Version 3 and 4. *Behaving identically* means here that the speaker chooses the same utterance act and the hearer the same interpretation. This solution will be further worked out in the following, and transferred to similar problems with assertions.

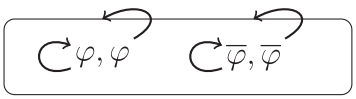
4 ASSERTIONS

In this section, we consider assertions in situations that show epistemic relations between speaker and hearer similar to those seen with the modified Clark & Marshall scenarios. The examples are taken from (Benz, 2008) and slightly modified for the present discussion. There are two possible states of affairs, either it snows in the Alps (φ), or it does not snow in the Alps ($\overline{\varphi}$). In each example, the classification of the speech act of uttering '*It is snowing in the mountains.*' is of interest. Version 1 in (17) shows the basic situation equivalent to Version 1 for definite descriptions in (13). Version 2 is a case in which the constitutive truth-condition is violated due to a false belief of the speaker. In Version 3 the speaker is lying outright, and Version 4 is a case in which the lie is detected by the hearer. Although the utterance of φ in versions 2–4 violates the constitutive condition of assertions as formulated in (1) and (8), we could, in each case, classify it as an instance of an assertion.

17) **Version 1.** Helga calls up her son Stephan who lives in a small town in the Alps and asks him whether he wants to visit her in Munich. Stephan answers: '*It is snowing in the mountains. So I don't want to drive now.*'
Version 2. Helga calls up her son Stephan and asks him whether he wants to visit her in Munich. Stephan answers: '*It is snowing in the mountains. So I don't want to drive now.*' But he has not checked the weather for some time, and it is now raining and the streets are clear.
Version 3. Helga calls up her son Stephan and asks him whether he wants to visit her in Munich. Stephan has a new girl-friend and prefers to stay at home this weekend. He answers: '*It is snowing in the mountains. So I don't want to drive now.*' However, he knows that it is not snowing and that the streets are clear.
Version 4. Helga calls up her son Stephan and asks him whether he wants to visit her in Munich. Stephan answers: '*It is snowing in the mountains.*' Helga has just talked to her daughter, who lives next to Stephan, and learned from her that it is not snowing and that the streets are clear. She also learned that Stephan has a new girl-friend and prefers to stay at home this weekend.

The epistemic relations in the basic situation is shown in (18). There are two states of the world: one in which it snows in the Alps (φ), and in which the speaker knows that it snows and says so, and another in which it does not snow ($\overline{\varphi}$), and in which the speaker knows it and says so. The hearer does not know which world is the real one, and all this is common knowledge. The state of the world and the possible utterance are coded as a pair of two formulas. In the basic Version 1, the two formulas are identical.

18) The epistemic relations in Version 1 of (17):



In (8), the joint project of asserting a sentence s with meaning φ has been defined as the set of all triples $\langle m, s, \varphi\rangle$ where m is a model that represents the state of the world and makes φ true. The idea behind this representation is that the state of affairs not only represents what is true about the world (the model m), but also what the possible future utterance events are. In (18), if φ is true in m, then the speaker can utter, according to constitutive rule (1), a sentence s with meaning φ, and thereby initiate the corresponding joint project. In the graph in (18), the state of affairs is represented by a pair of formulas, e.g., $\langle\varphi, \varphi\rangle$. The pair φ, φ represents an instance $\langle m, s, \varphi\rangle$ of the joint project of asserting s with meaning φ. The graphs in **Table 3** show the epistemic relations for Versions 2–4 of (17). They are all instances in which the proposition supported by the outer state of affairs and the meaning of the sentence uttered by the speaker are different from each other.

A comparison between **Tables 2** and **3** shows that the graphs are structurally identical except for their respective basic versions. The same reasoning that explains why the use of a definite description *the* φ can count as a referential act with target *Monkey Business* in the situations represented by the top nodes of Versions 2–4 in **Table 2** explains why the utterance of '*It is snowing in the mountains.*' can count as an assertoric act with propositional content φ in the situations represented by the top nodes of Versions 2–4 in **Table 3**. The classification as assertoric act travels along the epistemic path leading from the top node down to the basic situation that properly licenses the assertion.

TABLE 3 | Information states in versions 2 to 4 of (17). The nodes of the graph are pairs of formulas, where the first formula says whether φ or $\overline{\varphi}$ is true in the world, and the second formula represents an utterance by the speaker.

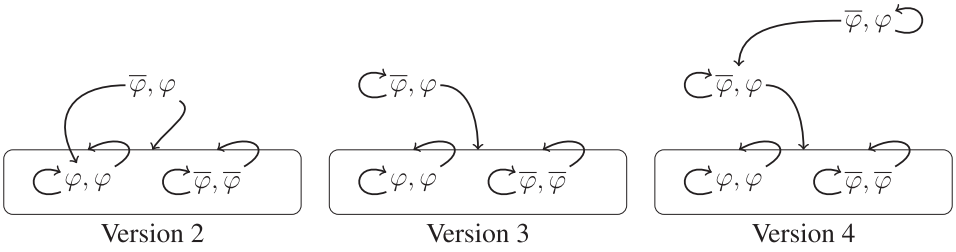
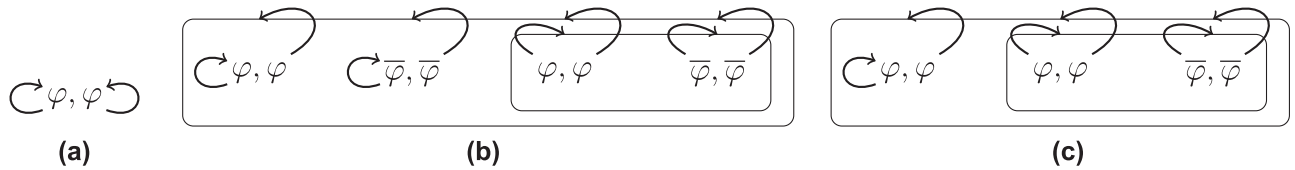


TABLE 4 | Different basic situations for Version 1 of (17).



We will introduce the mathematical framework which allows us to handle these examples precisely in **Section 5**. Before we turn to formal representations, we have a closer look at the structure of the epistemic graphs. They can be divided into a base and a hierarchical structure building up on it. The hierarchical part shows descending paths. The bases can differ in their internal structure. As we have seen, the bases for assertions in (18) and that for referential uses of definite descriptions shown in **Table 2** have different structure. **Table 4** shows three further possibilities for the base of the assertoric speech act. The first a) is a copy of the base for the referential use of definite descriptions. As a plausible base for Version 1 of (17) it is ruled out by an additional pragmatic constraint that says that the speaker should not say what is already common belief. However, we do not formalize this constraint so that a) remains a theoretical possibility. In setting up the epistemic graph in (18), we made the assumption that it is known that the speaker knows whether it is snowing, or not. This assumption does not follow, however, from Version 1 of (17). **Table 4b and c** show two possibilities where the hearer thinks it possible that the speaker does not know whether it snows. There are even more possibilities. For example, by bending the hearer's edge going out from the rightmost $\overline{\varphi}, \overline{\varphi}$-world back to this world, we would have a licit epistemic graph that allows for an assertion of φ in the leftmost φ, φ-world. We will discuss more examples once we have introduced formal representations.

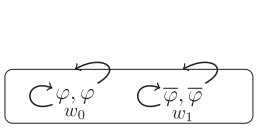
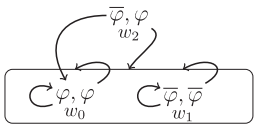
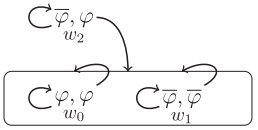
Our task is to explain why a certain utterance can be classified as an assertion in a given node in an epistemic graph. We adopt the following strategy: once it is explained why this classification is justified in a base situation, the classification can travel upward through the hierarchical part of the graph. This means, we can divide our considerations into that of the basic level and that of the higher hierarchical levels. Once the classification problem is solved for the base, the solution for the hierarchical part follows. One characteristic of the bases is the absence of descending paths. This means that all nodes in the bases are connected with each other. This leads to circular structures. We therefore consider circular structures separated from hierarchical ones.

5 THE MODEL

As explained in **Section 2**, we adopt a Clark (1996) perspective and represent *communicative acts* not as isolated acts but as coordinated joint projects consisting of a linguistic act by the speaker and a response by the hearer. A joint project consists of triples $\langle m, a, r\rangle$, where m represents the outer facts of the world, a the speaker's act, and r the hearer's response. This representation is, in general, more fine-grained than the traditional classification of speech acts. For example, we defined the joint project of referring to an object r with definite description *the* φ as the set of all triples $\langle m, \textit{the}\ \varphi, r\rangle$ for which m makes $\varphi(r)$ true ($m \models \varphi(r)$). The *referential use of definite descriptions* then consists of many such joint projects. It consists of all joint projects p for which there is a one-place predicate $\varphi(\,.\,)$ and an object r such that $p = \{\langle m, \textit{the}\ \varphi, r\rangle \mid m \models \varphi(r)\}$. Similarly, we defined the joint project of asserting a sentence s with meaning φ as the set of all triples $\langle m, s, \varphi\rangle$ for which $m \models \varphi$. The phenomenon of *assertive utterances* is then represented by the set of all joint projects p for which there is a sentence s with reading φ such that $p = \{\langle m, s, \varphi\rangle \mid m \models \varphi\}$.[4] In the previous section, we simplified the notation. For example, in Version 1 of (17) there are two state of affairs, one in which it is snowing and one in which it is not snowing. We identified them with two formulae, φ and $\overline{\varphi}$. There are two sentences s = '*It is snowing in the mountains*' and $\overline{s}$ = '*It is not snowing in the mountains*', which were again identified with φ and $\overline{\varphi}$ respectively. Hence, there were two joint projects involved: $p = \{\langle \varphi, s, \varphi\rangle\}$ and $\overline{p} = \{\langle \overline{\varphi}, \overline{s}, \overline{\varphi}\rangle\}$.

[4]The representation of the hearer's response by a formula φ is not essential here. We could have represented the same joint project as $p = \{\langle m, s, [\![s]\!]\rangle \mid m \in [\![s]\!]\}$, which would have made the connection to formal semantics even clearer.

TABLE 5 | Worlds and their epistemic graphs for scenarios in (17).

Version 1	Version 2	Version 3
		
$w_0 = \langle\langle\varphi, s, \varphi\rangle, \{w_0\}, \{w_0, w_1\}\rangle$	$w_0 = \langle\langle\varphi, s, \varphi\rangle, \{w_0\}, \{w_0, w_1\}\rangle$	$w_0 = \langle\langle\varphi, s, \varphi\rangle, \{w_0\}, \{w_0, w_1\}\rangle$
$w_1 = \langle\langle\overline{\varphi}, \overline{s}, \overline{\varphi}\rangle, \{w_1\}, \{w_0, w_1\}\rangle$	$w_1 = \langle\langle\overline{\varphi}, \overline{s}, \overline{\varphi}\rangle, \{w_1\}, \{w_0, w_1\}\rangle$	$w_1 = \langle\langle\overline{\varphi}, \overline{s}, \overline{\varphi}\rangle, \{w_1\}, \{w_0, w_1\}\rangle$
	$w_2 = \langle\langle\overline{\varphi}, s, \varphi\rangle, \{w_0\}, \{w_0, w_1\}\rangle$	$w_2 = \langle\langle\overline{\varphi}, s, \varphi\rangle, \{w_2\}, \{w_0, w_1\}\rangle$

5.1 Possible Worlds and Epistemic Relations

The joint projects do not represent epistemic relations between interlocutors and interlocutors and the world. We adopt a possible worlds representation in which beliefs are modeled as sets of epistemically possible worlds. A world has the form:[5]

19) **Possible world:** A *possible world* w is a triple $w = \langle\langle m, a, r\rangle, S, H\rangle$, where $\langle m, a, r\rangle$ is an element of some joint project p, and S and H are sets of possible worlds representing, respectively, the speaker's and hearer's beliefs.

We write S^w for the speaker's information state, and H^w for the hearer's information state in world w. Furthermore, we write $\langle m^w, a^w, r^w\rangle$ or d^w for the joint communicative act represented by w. Example definitions of worlds and their epistemic graphs are shown in **Table 5**.

In standard set theory, there is no w_0 that could satisfy equation $w_0 = \langle\langle\varphi, s, \varphi\rangle, \{w_0\}, \{w_0, w_1\}\rangle$ due to the Axiom of Foundation. We therefore turn to a variant of set theory with Anti-Foundation Axiom (AFA) developed by Aczel (1988). This theory has been used extensively for modeling circular structures (Barwise, 1989; Barwise and Etchemendy, 1989; Barwise and Moss, 1996; Gerbrandy and Groeneveld, 1997; Gerbrandy, 1998; Benz, 2008). We do not go into the intricacies of this theory. We need one important property: in AFA-set theory every system of equations has a unique solution. For example, the equations for the different graphs shown in **Table 5** are systems of equations. We can consider the names of worlds $w_0, w_1, \ldots$ as *variables* for which we seek a solution. A solution is a function that maps the variables to ordinary (non-well-founded) sets that satisfy the equations. As we have said, every such system of equations has a unique solution in AFA-set theory. Hence, the worlds shown in **Table 5** are well-defined set-theoretic entities. The property also allows for simple representations of *belief updates*. Propositions can be identified with sets of possible worlds. If an interlocutor X learns that a proposition φ holds, then this can be represented by intersecting the set of worlds that represent X's beliefs with the set of worlds representing the meaning of φ. We say then that X's beliefs have been *updated* with φ. If the proposition is mutually learned, then each interlocutor has to update not only his/her own belief set, but also the belief sets representing the beliefs of others, and this update has to be iteratively applied to each other's beliefs. In terms of systems of equations, this can be modeled by first writing down the original system of equations, and then intersecting all belief sets occurring in the system with the set representing φ. Finally, the modified system of equations has to be solved again. The solution then represents the updated system of beliefs. The results of updating the worlds in **Table 5** with φ, i.e., with $\{w_0\}$, are shown in **Table 6**. The results for Version 1 and 2 follow immediately from the definition. However, Version 3 is not yet accounted for. If we update with $\{w_0\}$, then w_2 should be eliminated from the speaker's belief state, and, therefore, we should expect the empty, i.e., contradictory, belief state after updating w_2. We will see later how to account for the result shown in **Table 6**.

The update that we just described can be represented by a formal update operator $\star$. It models the effect of *mutual learning* some information Y.[6] In **Eq. 1**, $w\star Y$ denotes the update of beliefs in a world w with Y, and in **Eq. 2**, $X\star Y$ the update of a belief set X with Y.

$$w\star Y := \langle\langle m, a, r\rangle, S\star Y, H\star Y\rangle \tag{1}$$

$$X\star Y := \{v^\star Y | v \in X \cap Y\} \tag{2}$$

The graph of Version 1 of **Table 6** represents $w_0\star\{w_0\}$ with w_0 as in Version 1 of **Table 5**, and the graph of Version 2 of **Table 6** represents $w_2\star\{w_0\}$ with w_2, w_0 defined as in Version 2 of **Table 5**. World w_2 survives in Version 2 as only worlds in belief states are eliminated. If a system of equations represents a belief state, i.e., a set of possible worlds, then updating the system of equations with information Y is equivalent to removing all variables w_i from *both sides* of the system for which the solution $s(w_i)$ is not an element of Y.

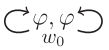
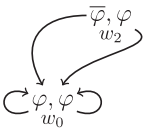
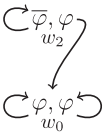
We are now in a position to explain an important modeling decision. Why do possible worlds represent joint communicative acts $\langle m, a, r\rangle$, and not only the state of affairs m? Let us consider Version 3 in **Table 5**, and let us change the definition of worlds such that only the outer state of affairs is represented. Then Version 3 is represented by the following system of equations:

20) $w_0 = \langle\varphi, \{w_0\}, \{w_0, w_1\}\rangle, w_1 = \langle\overline{\varphi}, \{w_1\}, \{w_0, w_1\}\rangle,$
$w_2 = \langle\overline{\varphi}, \{w_2\}, \{w_0, w_1\}\rangle$

[5]For the relation between possible worlds as defined in (19) and *Kripke*–frames for modal logics see **Appendix A.2**.

[6]The notation with $\star$ follows (Barwise and Moss, 1996).

TABLE 6 | Worlds and their epistemic graphs for scenarios in (17) after updating with $\varphi = \{w_0\}$.

Version 1	Version 2	Version 3
		
$w_0 = \langle\langle\varphi, s, \varphi\rangle, \{w_0\}, \{w_0\}\rangle$	$w_0 = \langle\langle\varphi, s, \varphi\rangle, \{w_0\}, \{w_0\}\rangle$ $w_2 = \langle\langle\overline{\varphi}, s, \varphi\rangle, \{w_0\}, \{w_0\}\rangle$	$w_0 = \langle\langle\varphi, s, \varphi\rangle, \{w_0\}, \{w_0\}\rangle$ $w_2 = \langle\langle\overline{\varphi}, s, \varphi\rangle, \{w_2\}, \{w_0\}\rangle$

If we replace w_2 by w_1, then (20) turns into (21):

21) $w_0 = \langle\varphi, \{w_0\}, \{w_0, w_1\}\rangle$, $w_1 = \langle\overline{\varphi}, \{w_1\}, \{w_0, w_1\}\rangle$

Every solution that solves (21) also provides a solution for (20). As every system of equations has only one solution, it follows that the solutions for w_1 in (21) and for w_2 in (20) must be identical. A graphical representation corresponding to that of Version 3 in **Table 5** accompanied by the equation in (20) can easily create the illusion of w_1 being distinct from w_2. If we represent not only the state of affairs but also possible interactions in worlds, then w_1 and w_2 become distinct. In (Benz, 2008) the distinction between w_1 and w_2 was achieved by including the speaker's goal in the structure of possible worlds. By including sequences $\langle m, a, r\rangle$ we achieve the same effect: we represent the speaker's intention, the intention to evoke response r by doing a in situation m. Without intentions, we could not distinguish lies from honest assertions.

Possible worlds defined by systems of equations can represent utterance situations one at a time. It would be desirable to have definitions of whole classes of utterance situations that share certain characteristics. To avoid the necessary apparatus, we continue on a case by case basis.[7]

We need some additional concepts. First, we introduce the notion of an *epistemic path*. An epistemic path from w_1 to w_{n+1} is a sequence $\langle w_0, X_0, \ldots, w_n, X_n, w_{n+1}\rangle$ with the property: for all i, X_i is either S^{w_i} or H^{w_i}, and $w_{i+1} \in X_i$.

The *transitive hull* of a world w is the set of worlds that includes w itself and all worlds that can be reached from w via a connecting epistemic path. Let w be a possible world. We first construct sets of worlds that are reachable in $0, \ldots, n$ steps:

$$T^0 = \{w\},\ T^{n+1} = T^n \cup \bigcup\{X^v \mid v \in T^n \wedge X = S, H\}. \tag{3}$$

The transitive hull of w is then defined as the union of all T^ns:

$$\overline{T}(w) := \bigcup_n T^n. \tag{4}$$

It can be verified that $\overline{T}(w)$ is the set of all worlds that are reachable via an epistemic path from w. For example, in Version 1 of **Table 5**, $\overline{T}(w_0) = \{w_0, w_1\} = \overline{T}(w_1)$, and in Versions 2 and 3, we find again $\overline{T}(w_0) = \{w_0, w_1\} = \overline{T}(w_1)$, and $\overline{T}(w_2) = \{w_0, w_1, w_2\}$. Hence, $\overline{T}(w_0) = \overline{T}(w_1) \subsetneq \overline{T}(w_2)$. This shows the hierarchical structure of w_2, and helps distinguishing worlds in the base of a graph where it holds for all v, w that $\overline{T}(v) = \overline{T}(w)$, and the worlds w which are higher up in the graph, for which it holds that there is a $v \in \overline{T}(w)$ such that $\overline{T}(v) \subsetneq \overline{T}(w)$.

Finally, we introduce two formal properties of possible worlds w:

$$\forall v \in \overline{T}(w) \forall X = S, H:\ X^v \neq \varnothing \wedge \forall u \in X^v\, X^u = X^v \qquad \text{introspection} \tag{5}$$

$$\forall v \in \overline{T}(w): v \in S^v \cap H^v \qquad \text{truthfulness} \tag{6}$$

The first property entails that interlocutors know what they know, and know what they do not know. This is sometimes considered too strong an assumption about beliefs. We assume it here for convenience. The other property says that it is common knowledge that interlocutors have only true beliefs. If *truthfulness* holds for w, then every path in $\overline{T}(w)$ can be reversed, i.e., if some world v can be reached from another world u, then u can also be reached from v. In particular, *truthfulness* entails that for all $v \in \overline{T}(w): \overline{T}(v) = \overline{T}(w)$.

We always assume *introspection*, and for elements of the base of an epistemic graph, we also assume *truthfulness*.

5.2 The Base Level of an Epistemic Graph

In **Table 4** we have seen various examples of basic epistemic graphs. They have in common that all worlds are connected with each other. This is entailed by the *truthfulness* condition that we assume to hold for all well-behaved communicative situations. The idea is that we can first solve the simpler task of classifying communicative acts in well-behaved situations, and then generalize the classification to the ill-behaved ones.

[7]For example, the class of all possible worlds could be introduced as the maximal fixed–point of the set continuous operator $\Gamma X := \{\langle d, x, y\rangle \mid d \in D,\ x, y \subseteq X\}$, where D is some set of instances of joint projects.

The epistemic relations in an utterance situation is represented by an epistemic graph. The goal of this section is to show how a sub-graph can be constructed that satisfies all epistemic felicity conditions. This construction will be a fixed-point construction. We first introduce formal variants of the *licensing* and *uniqueness* conditions.

Let us consider *licensing* from the speaker's perspective. If the speaker wants to start a joint project p he has be to sure that it can be performed in all epistemically possible worlds. The speaker can only perform a single act. Hence, there must be an act a such that for all epistemically possible states of affairs m there is a response r and a world $v \in S^w$ such that $\langle m^v, a^v, r^v\rangle = \langle m, a, r\rangle \in p$. For example, if the speaker wants to assert s with meaning φ, then φ has to be true in all epistemically possible states of affairs. As information states are sets of possible worlds, not sets of state of affairs, the actual definition that follows in (9) has to be slightly more roundabout. Assume that there are several sentences $s_0, s_1, s_2, \ldots$ with different and non-exclusive meanings φ_i that the speaker knows to be true. Then, for each joint communicative act $\langle m, s_i, \varphi_i\rangle$ there is a world w_i in the speaker's belief state in which the joint act is performed. For this world w_i it would not be clear what it should mean that another joint act $\langle m, s_j, \varphi_j\rangle$ can be performed. So, the requirement that it must be possible to perform a joint communicative act in all the speaker's epistemically possible worlds has to be re–worded: For all possible worlds w there must exist a world v that represents the joint act and agrees with w in the state of affairs m and the speaker's and hearer's belief states. Hence, we say that two worlds w and v are *similar*, if $\langle m^w, S^w, H^w\rangle = \langle m^v, S^v, H^v\rangle$. For the following it is convenient to introduce notation for the set of worlds out of a given set X that are similar to a given world w:

$$[w]_X := \{v \in X \mid m^v = m^w \wedge S^v = S^w \wedge H^v = H^w\}. \tag{7}$$

For convenience, we also introduce notation for the set X_a of all worlds that share the same utterance act a, and the set X^p of all worlds with a joint communicative act that belongs to a given project p:

$$X_a := \{v \in X \mid a^v = a\}, \quad X^p := \{v \in X \mid \langle m^v, a^v, r^v\rangle \in p\}. \tag{8}$$

With these preparations, we can introduce the formal constraints for *licensing* and *uniqueness*. They are formulated as conditions on information states, i.e., sets of possible worlds X, that depend on a project p and an act a:

$$\mathrm{L}_{p,a}X :\Leftrightarrow \forall w \in X \exists v \in [w]_X : a^v = a \wedge \langle m^v, a^v, r^v\rangle \in p \quad \text{licensing} \tag{9}$$

$$\mathrm{U}_{p,a}X :\Leftrightarrow \forall w, v \in X_a : \langle m^v, a^v, r^v\rangle \in p \rightarrow r^v = r^w \quad \text{uniqueness} \tag{10}$$

The *uniqueness* condition says that for every state of affairs in which act a can initiate a joint communicative act it will lead to the same response. *Uniqueness* is downward entailing, i.e., $X \subseteq Y$ entails $\mathrm{U}_{p,a}Y \rightarrow \mathrm{U}_{p,a}X$, and depends only on the joint communicative acts represented in X.

We can now show how to construct a maximal sub–set of a given set X in which the epistemic felicity conditions *licensing* and *uniqueness* are mutually guaranteed to hold. Let there be a given set $\mathcal{P}$ of joint projects. Let X be a set of possible worlds such that for each $v \in \overline{T}(w)$ it holds that its joint communicative act $\langle m^w, a^w, r^w\rangle$ belongs to some project $p \in \mathcal{P}$. If X is the speaker's belief state, then she knows that the epistemic felicity conditions for initiating a certain project p with a certain act a are satisfied in the following sub–set of X:

$$F^S_{p,a}X = \left\{v \in X^p_a \,\middle|\, \mathrm{L}_{p,a}S^v \wedge \mathrm{U}_{p,a}S^v\right\}. \tag{11}$$

This is the set of all $v \in X$ with joint communicative act $\langle m^v, a^v, r^v\rangle \in p$ and utterance act $a^v = a$ that satisfy *licensing* and *uniqueness*.

The hearer, in contrast to the speaker, does not need to believe that act a initiates project p in all possible worlds. It suffices that he believes that it is consistent with his information. Hence, *licensing* can be restricted to a non-empty sub-set of his belief state:

$$F^H_{p,a}X = \left\{v \in X^p_a \,\middle|\, \exists Y \subseteq H^v\left(\emptyset \neq Y \wedge \mathrm{L}_{p,a}Y\right) \wedge \mathrm{U}_{p,a}H^v\right\}. \tag{12}$$

We can construct the set of possible worlds in which the epistemic felicity conditions are mutual knowledge by an iterative process of eliminating worlds that do not satisfy them. The construction proceeds in parallel for all acts a and joint projects $p \in \mathcal{P}$. We start with a set X of possible worlds for which *truthfulness* holds and *transitivity* holds, i.e., for each $w \in X$ it holds that $\overline{T}(w) \subseteq X$. We set $F^0_\star = F^0 = X$. In the first step, we collect all worlds v in X which satisfy the speaker's epistemic felicity conditions, and *update* X with the information that they are satisfied. We do this for all joint projects $p \in \mathcal{P}$ and acts a:

$$F^1 = \bigcup_{p,a} F^S_{p,a}F^0_\star \quad \text{(selects the worlds in which the speaker's epistemic conditions are satisfied)} \tag{13}$$

$$F^1_\star = F^0_\star * F^1 \quad \text{(updates with this information)} \tag{14}$$

In the next step, this is repeated for the hearer's epistemic felicity conditions:

$$F^2 = \bigcup_{p,a} F^H_{p,a}F^1_\star \quad \text{(selects the worlds in which the hearer's epistemic conditions are satisfied);} \tag{15}$$

$$F^2_\star = F^1_\star * F^2 \quad \text{(updates with this information)} \tag{16}$$

This construction continues such that in each odd step the speaker's epistemic felicity conditions are checked, and in the even steps the hearer's:

$$F^{2n+1} = \bigcup_{p,a} F^S_{p,a}F^{2n}_\star \qquad F^{2n+1}_\star = F^{2n}_\star * F^{2n+1} \tag{17}$$

$$F^{2n+2} = \bigcup_{p,a} F^H_{p,a}F^{2n+1}_\star \quad F^{2n+2}_\star = F^{2n+1}_\star * F^{2n+2} \tag{18}$$

Fortunately, it is not necessary to repeat this infinitely often. We can show that:

$$\forall n \geq 3 : \; F^n_\star = F^3_\star. \tag{19}$$

Why should the construction stabilize after three steps? After the first step, it is common knowledge that licensing and uniqueness hold from the speaker's perspective. As belief states can only become smaller by updating, the speaker's uniqueness condition is guaranteed to hold for all following construction steps. As for each remaining world w, it holds that $w \in H^w$ due to *truthfulness*, the hearer's licensing condition is automatically satisfied. Some worlds may be removed in step two due to the hearer's uniqueness

condition. After step two, the hearer's uniqueness condition is guaranteed to hold in all subsequent construction steps. Updating in step two may introduce violations of the speaker's licensing condition. In step three, worlds that violate speaker's licensing are again removed. As both the speaker and the hearer's uniqueness conditions must hold, only the licensing conditions could remove further worlds. However, as *truthfulness* holds, hearer's licensing is entailed by the speaker's licensing condition. Hence, in step four, none of the remaining worlds can be removed.

22) **Fixed-point.** Given a set of joint projects $\mathcal{P}$ and a set X of possible worlds for $\mathcal{P}$ where *truthfulness* and *transitivity* hold, then the maximal sub-set of X in which it is common knowledge that the epistemic felicity conditions of speaker and hearer are satisfied is $\nabla_{\mathcal{P}} X := F^3_\star$.

We next consider some examples. The first (23) demonstrates several points: first, basic cases can become more complex than the ones considered before; second, there are additional modeling assumptions that have to be made; third, for visualization there is a different type of graph that is better suited for base situations; and fourth, it shows how the construction is applied for finding fixed–points for epistemically felicitous referential uses of definite descriptions.

23) **Scenario.** The following is common knowledge. Either (*mb*) *Monkey Business* or (*dr*) *A Day at the Races* is showing at the Roxy. Ann has read the program, and knows which one it is. The newsfeed that Bob uses would only announce the program if *Monkey Business* is showing. Hence, if the state of affairs (m_0) is such that *A Day at the Races* is on the program, Bob will be uncertain. If *Monkey Business* is showing (m_1, m_2, m_3), he might have read the announcement (m_2, m_3), or not (m_1). If he has read it, Ann may know that (m_3), or not (m_1, m_2).

In which situation is it mutually felicitous to refer to *Monkey Business* with *the* φ = '*The movie showing at the Roxy*'? The answer is only in m_3. We will see how this comes out. Two graphical representations and a system of equations are shown in **Table 7**. The graph in c) shows the joint projects more clearly. Vertical lines in the center column shows situations that are indiscernible for the hearer after the speaker's action (*the* φ), and vertical lines in the first column shows situations that are indiscernible for the speaker before acting. The graphs in a) and c) are equivalent.

There are two competing joint projects starting with *the* φ: The project $q = \{\langle m_0, \textit{the}\ \varphi, dr\rangle\}$ where reference to *A Day at the Races*, and a project $p = \{\langle m_i, \textit{the}\ \varphi, mb\rangle \mid i = 1, 2, 3\}$ where reference to *Monkey Business* is intended. We also assume, for reasons that will soon become clear, that there is a *do–nothing* project $l = \{\langle m_i, \varepsilon, \varepsilon\rangle \mid i = 1, \ldots, 4\}$ where no action is performed. We construct $\nabla_{\mathcal{P}} X$ for $\mathcal{P} = \{p, q, l\}$ and $X = \{w_0, w_1, w_2, w_3\}$. For now, we ignore project l. In the first construction step, we test for each project whether the speaker's felicity conditions are satisfied. It can be verified that for w_0 the conditions for q hold, and that for w_1, w_2, and w_3 the conditions for p hold. Hence, none of the worlds is eliminated. We turn to the hearer and the second construction step. The hearer's licensing condition is automatically satisfied as in each case $\{w_i\} \subseteq H^{w_i}$. However, *uniqueness* is violated for $H^{w_0} = H^{w_1}$. For w_2 and w_3 uniqueness is satisfied. Hence, the system has to be updated with $\{w_2, w_3\}$. This would lead to (24).

24) $w_2 = \langle\langle m_2, \textit{the}\ \varphi, mb\rangle, \{w_2\}, \{w_2, w_3\}\rangle$
$w_3 = \langle\langle m_3, \textit{the}\ \varphi, mb\rangle, \{w_3\}, \{w_2, w_3\}\rangle$

Clearly, *licensing* is satisfied for the speaker's information state in both w_2 and w_3. The construction stabilizes, and we arrive at the prediction that the referential act is mutually felicitous in both w_2 and w_3. This is obviously not correct. In the original w_2 the speaker Ann did not know whether Bob has read the announcement, and, hence, she thought that he may be ignorant about the movie playing. This cannot have changed by just reasoning about felicity conditions. What went wrong? When updating with $\{w_2, w_3\}$, we eliminated w_1. This means that Ann, in a situation in which she does not know whether Bob read the program $S = \{w_1, w_2\}$, would reason that Bob must have read the program (w_2) because, otherwise, he would not know to what she is referring to with *the* φ. This wishful reasoning is blocked by the *do–nothing* project l. It has the effect that none of the possible state of affairs $m_0, \ldots, m_3$ are eliminated. (25) shows the system of equations for the epistemic relations with project l.

TABLE 7 | Representations of scenario (23).

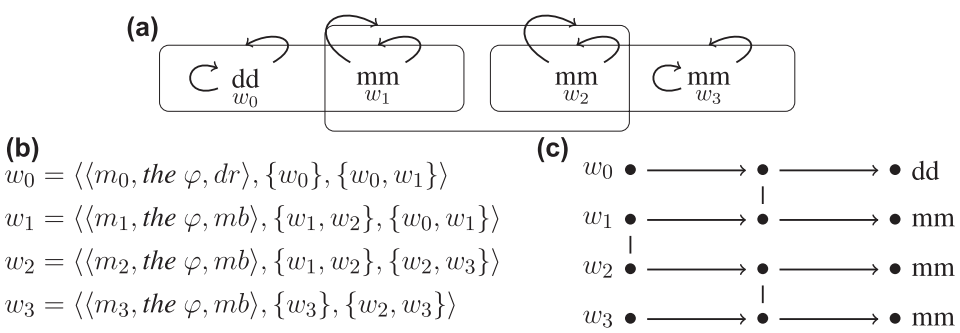
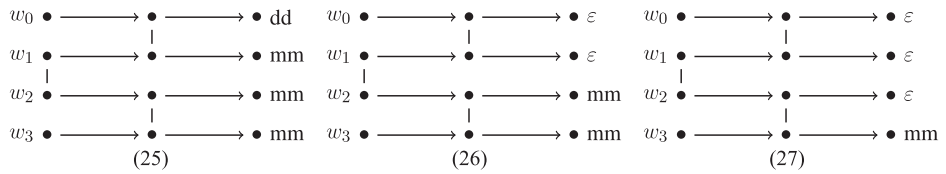


TABLE 8 | The graphs for construction steps 1, 2, and 3 for Example 23 with *do–nothing* project *l*.



25) $w_0 = \langle\langle m_0, \textit{the}\ \varphi, dr\rangle, \{w_0, w_{0,\varepsilon}\}, \{w_0, w_1, w_{0,\varepsilon}, w_{1,\varepsilon}\}\rangle$
$w_1 = \langle\langle m_1, \textit{the}\ \varphi, mb\rangle, \{w_1, w_2, w_{1,\varepsilon}, w_{2,\varepsilon}\}, \{w_0, w_1, w_{0,\varepsilon}, w_{1,\varepsilon}\}\rangle$
$w_2 = \langle\langle m_2, \textit{the}\ \varphi, mb\rangle, \{w_1, w_2, w_{1,\varepsilon}, w_{2,\varepsilon}\}, \{w_2, w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$
$w_3 = \langle\langle m_3, \textit{the}\ \varphi, mb\rangle, \{w_3, w_{3,\varepsilon}\}, \{w_2, w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$
$w_{0,\varepsilon} = \langle\langle m_0, \varepsilon, \varepsilon\rangle, \{w_0, w_{0,\varepsilon}\}, \{w_0, w_1, w_{0,\varepsilon}, w_{1,\varepsilon}\}\rangle$
$w_{1,\varepsilon} = \langle\langle m_1, \varepsilon, \varepsilon\rangle, \{w_1, w_2, w_{1,\varepsilon}, w_{2,\varepsilon}\}, \{w_0, w_1, w_{0,\varepsilon}, w_{1,\varepsilon}\}\rangle$
$w_{2,\varepsilon} = \langle\langle m_2, \varepsilon, \varepsilon\rangle, \{w_1, w_2, w_{1,\varepsilon}, w_{2,\varepsilon}\}, \{w_2, w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$
$w_{3,\varepsilon} = \langle\langle m_3, \varepsilon, \varepsilon\rangle, \{w_3, w_{3,\varepsilon}\}, \{w_2, w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$

The update in Step 2 again eliminates$\{w_0, w_1\}$, which leads to (26). Note that *licensing* and *uniqueness* are trivially satisfied for project *l*: licensing says that the project can be initiated for all state of affairs, and uniqueness says that, once initiated, it can only be completed in one way. Hence, no update can remove any of the $w_{i\varepsilon}$-worlds.

26) $w_2 = \langle\langle m_2, \textit{the}\ \varphi, mb\rangle, \{w_2, w_{1,\varepsilon}, w_{2,\varepsilon}\}, \{w_2, w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$
$w_3 = \langle\langle m_3, \textit{the}\ \varphi, mb\rangle, \{w_3, w_{3,\varepsilon}\}, \{w_2, w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$
$w_{0,\varepsilon} = \langle\langle m_0, \varepsilon, \varepsilon\rangle, \{w_{0,\varepsilon}\}, \{w_{0,\varepsilon}, w_{1,\varepsilon}\}\rangle$
$w_{1,\varepsilon} = \langle\langle m_1, \varepsilon, \varepsilon\rangle, \{w_2, w_{1,\varepsilon}, w_{2,\varepsilon}\}, \{w_{0,\varepsilon}, w_{1,\varepsilon}\}\rangle$
$w_{2,\varepsilon} = \langle\langle m_2, \varepsilon, \varepsilon\rangle, \{w_2, w_{1,\varepsilon}, w_{2,\varepsilon}\}, \{w_2, w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$
$w_{3,\varepsilon} = \langle\langle m_3, \varepsilon, \varepsilon\rangle, \{w_3, w_{3,\varepsilon}\}, \{w_2, w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$

Now, if we consider w_2, we see that licensing is not satisfied for p and S^{w_2} as there is a possible world $(w_{1,\varepsilon})$ in S^{w_2} with state of affairs m_1, for which the speaker knows that it is not possible to initiate p. Hence, w_2 is eliminated. As w_3 satisfies licensing and uniqueness, it survives. The final system of equations is shown in (27).

27) $w_3 = \langle\langle m_3, \textit{the}\ \varphi, mb\rangle, \{w_3, w_{3,\varepsilon}\}, \{w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$
$w_{0,\varepsilon} = \langle\langle m_0, \varepsilon, \varepsilon\rangle, \{w_{0,\varepsilon}\}, \{w_{0,\varepsilon}, w_{1,\varepsilon}\}\rangle$
$w_{1,\varepsilon} = \langle\langle m_1, \varepsilon, \varepsilon\rangle, \{w_{1,\varepsilon}, w_{2,\varepsilon}\}, \{w_{0,\varepsilon}, w_{1,\varepsilon}\}\rangle$
$w_{2,\varepsilon} = \langle\langle m_2, \varepsilon, \varepsilon\rangle, \{w_{1,\varepsilon}, w_{2,\varepsilon}\}, \{w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$
$w_{3,\varepsilon} = \langle\langle m_3, \varepsilon, \varepsilon\rangle, \{w_3, w_{3,\varepsilon}\}, \{w_3, w_{2,\varepsilon}, w_{3,\varepsilon}\}\rangle$

Now, the prediction is that Ann can use *the* φ for referring to *Monkey Business* only in situation w_3. If she uses it, then the system in (27) is updated with the set of all worlds that instantiate a project starting with *the* φ; i.e. it has to be updated with $\{w \mid a^w = \textit{the}\ \varphi\} = \{w_3\}$. This update leads to $w_3 = \langle\langle m_3, \textit{the}\ \varphi, mb\rangle, \{w_3\}, \{w_3\}\rangle$. This implies that the definite description not only tells Bob to pick out *Monkey Business* as the referent, it also tells him that Ann knows that he has read the program.

The definition of possible worlds becomes more complicated when adding project *l*. As project *l* is defined for all state of affairs, it does not need to be shown in graphs, except some state of affairs would otherwise be eliminated. The simplified graphs in **Table 8** represent the possible worlds defined in construction steps (25)–(27) with joint project *l* only showing when necessary.

5.3 Hierarchical Epistemic Graphs

The base level of a graph consist of worlds that satisfy *truthfulness* and *introspection*. Now, we turn to examples where the *truthfulness* condition is violated. All the examples that we have seen are represented by graphs that have a base in $\nabla\mathcal{P}$ over which a hierarchical structure is erected. All worlds in the upper structure are rooted in the base by epistemic paths reaching down to it. This section will be less technical. We will concentrate on showing different types of epistemic graphs that can be found on higher levels. We first clarify in which sense the worlds have a hierarchical structure. It is possible to distinguish different levels in this hierarchical structure, depending on how deeply the base is embedded in a world. Each level is characterized by a unique *order type* which is shared by all worlds at this level. As we have seen, the *truthfulness* condition implies that each world in the transitive hull $\overline{T}(w)$ of a world w is connected to every other world by an epistemic path, in particular, it holds that the transitive hulls of all worlds in $\overline{T}(w)$ are identical. We give these worlds the order type 0. We define the order type otp of other worlds recursively using the transitive hull.

$$\mathrm{otp}(w) = 0 \Leftrightarrow \forall v \in \overline{T}(w)\, w \in \overline{T}(v). \tag{20}$$

$$\mathrm{otp}(w) = \sup\{\mathrm{otp}(v) + 1 \mid v \in \overline{T}(w) \wedge w \notin \overline{T}(v)\} \tag{21}$$

$$\mathrm{otp}(X) = \sup\{\mathrm{otp}(v) \mid v \in X\} \tag{22}$$

The first condition says that worlds at the base have order type 0. The second, that for other worlds w the order type is the smallest ordinal that is larger than all order types of worlds from which w cannot be reached by an epistemic path.[8] The last condition introduces the order type of a set of possible worlds which is the smallest ordinal that is at least as large as the order types of all the worlds in the set. For example, in **Table 7**, the worlds w_0 and w_1 have order type 0, and w_2 has order type 1. In **Table 3**, the top worlds in Versions 1 and 2 have order type 1, and that of Version 3 order type 2, and in **Table 1** we see examples with order types increasing from 0 in a) to 4 in e).

Let us first consider Versions 2 and 3 in **Table 5** with the corresponding examples in (17). The joint project

[8]Set theoretically the *supremum* of a set of ordinals is just the union of these ordinals. The definition is maximally general and extends into the transfinite. However, in this article, we only consider worlds with finite order type.

of *asserting* sentence s with meaning φ was defined as the set of all triples $\langle m, s, \varphi\rangle$ consisting of a model m that makes φ true, the speaker's utterance s and the hearer's interpretation φ. In epistemic graphs, as in **Table 5**, such a triple was represented by the pair φ, φ, the first φ saying that m is such that φ is true, and the second φ representing the hearer's interpretation of s. Hence, in the base level of an epistemic graph, the formulae appearing in the pairs must always be identical. As Version 2 and 3 in **Table 5** show, this may no longer be the case in higher levels. To account for this possibility, we have to make the joint projects independent of the state of affairs. Let $p \in \mathcal{P}$ be a joint project, then the *extended* joint project $\tilde{p}$ is defined as

$$\tilde{p} = \{\langle m, a, r\rangle \mid \exists m' \, \langle m', a, r\rangle \in p\}. \tag{23}$$

This means, the joint communicative acts $\overline{\varphi}, \varphi$ that we see in **Table 5** are elements of the extended joint project of asserting φ. We allow extended projects to occur only on higher levels of the hierarchy. We are going to show that the *licensing* and *uniqueness* conditions can be re-used at higher levels to determine the worlds where asserting s is epistemically felicitous. For extended projects, the conditions are shown in **Eqs. (24)** and **(25)**.

$$\mathrm{L}_{\tilde{p},a}X \Leftrightarrow \forall w \in X \, \exists v \in [w]_X : \; a^v = a \wedge \langle m^v, a^v, r^v\rangle \in \tilde{p} \quad \text{licensing} \tag{24}$$

$$\mathrm{U}_{\tilde{p},a}X \Leftrightarrow \forall w, v \in X_a : \; \langle m^v, a^v, r^v\rangle \in \tilde{p} \rightarrow r^v = r^w \quad \text{uniqueness} \tag{25}$$

The conditions are unchanged, except that basic projects have been replaced by extended projects. The *licensing* condition says that the joint communicative act can be performed in all epistemically possible state of affairs, and *uniqueness* that performing it leads to a unique response for each state. The operators selecting worlds satisfying the epistemic felicity constraints stay the same, except that the basic projects are replaced by extended projects. For convenience, they are shown in **Eq. 26** and **Eq. 27**.

$$F^S_{\tilde{p},a}X = \left\{v \in X^{\tilde{p}}_a \,\middle|\, \mathrm{L}_{\tilde{p},a}S^v \wedge \mathrm{U}_{\tilde{p},a}S^v\right\}. \tag{26}$$

$$F^H_{\tilde{p},a}X = \left\{v \in X^{\tilde{p}}_a \,\middle|\, \exists Y \subseteq H^v\big(\emptyset \neq Y \wedge \mathrm{L}_{\tilde{p},a}Y\big) \wedge \mathrm{U}_{\tilde{p},a}H^v\right\}. \tag{27}$$

Apart from checking whether *licensing* and *uniqueness* hold for the speaker and hearer's perspective, the operators check whether the joint communicative act represented by a possible world is an instantiation of a given extended joint project $\tilde{p}$ performed with a special act a.

With these operators, a fixed–point can be constructed as in 17, the only difference being that the construction is applied bottom up, level by level. We eschew the technical details and demonstrate their workings with some examples. Let us consider the graphs in **Table 9**. In graphs a), c), and d), the belief states of participants are subsets of the base level. In w_0 and w_1 the epistemic felicity conditions for assertions are satisfied in a) and b), and for definite references in c) and d). Graphically, it should be easy to check that the felicity conditions of *licensing* and *uniqueness* are satisfied for w_2 and w_3 in a), b), and d), and violated in c). Checking the formal definitions needs more effort. First, we note that for all graphs the fixed–point of the base level $\nabla_{\mathcal{P}}\{w_0, w_1\}$ is equal to the base level itself, and that in a) and b) assertions of φ are licensed in w_0, and of $\overline{\varphi}$ in w_1. We consider first world w_2 in a). The abbreviation $\overline{\varphi}, \varphi$ stands for the joint communicative act $\langle m, s, \varphi\rangle$ with a model m that supports $\overline{\varphi}$ and an assertion of a sentence s that expresses semantically that φ. Hence, asserting s in m violates the constituting rules of assertions. However, $\langle m, s, \varphi\rangle$ is an element of the extended joint project of asserting φ. We have to check the felicity conditions of *uniqueness* and *licensing* for w_2. As mentioned before, *uniqueness* is trivially satisfied for assertions, as we assumed that semantic meaning is not ambiguous. Only licensing has to be checked. This is identical to checking licensing for w_0 in graph (18), as the belief states of speaker and hearer in w_2 and w_0 are identical. As the felicity conditions are satisfied in w_0, it only remains to check the condition '$v \in X^{\tilde{p}}_s$' in Definition (26). As fixed-points are calculated level by level, X must be the restriction of $\overline{T}(w_2)$ to Level 1, i.e., $X = \{w_2\}$. As $\langle \overline{\varphi}, \varphi\rangle$ is an instance of the extended project of asserting φ, the condition is satisfied. Hence, applications of $F^S_{\tilde{p},s}$ and $F^H_{\tilde{p},s}$ to $\{w_2\}$ return again $\{w_2\}$. Clearly, further applications of these operators cannot change the result, so that $\{w_2\}$ must be a fixed-point of these operators. This shows that asserting s with interpretation φ satisfies the joint epistemic felicity conditions, and, hence, it is the case that both interlocutors agree on the interpretation of s, and that they both believe that they can mutually figure this out. The case of w_3 is symmetrical, where φ and $\overline{\varphi}$ change places. In sum, it follows that asserting a sentence s with meaning φ is epistemically felicitous in w_2, and asserting a sentence $\overline{s}$ with meaning $\overline{\varphi}$ in w_3 is likewise epistemically felicitious.

We next turn to b) in **Table 9**. From the hearer's perspective, the situation is identical to that of a) or that of Version 1 in (17) with graph (18). Hence, we only need to consider the speaker's perspective in w_2

TABLE 9 | Epistemic graphs: **(a)**, **(b)** for assertions in scenarios similar to (17); **(a)**: false belief case, **(b)**: lying; **(c)**, **(d)** for reference in scenarios similar to (13)–(16); **(c)**: failed joint reference on higher level, **(d)**: successful joint reference on higher level. Abbreviations: φ a proposition, $\overline{\varphi}$ its negation; *m*: *Monkey Business*, *d*: *A Day at the Races*.

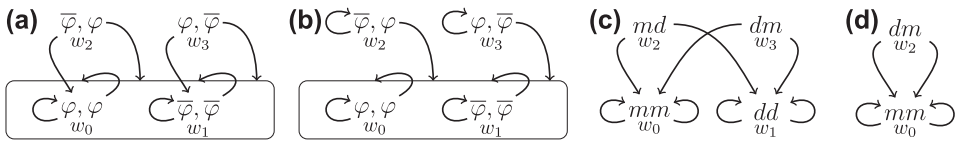
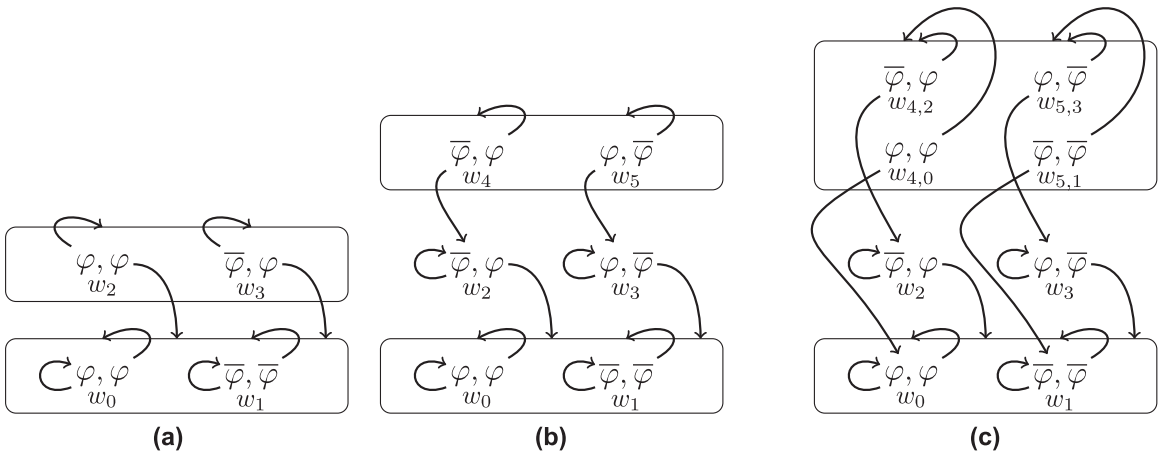


TABLE 10 | Epistemic graphs with uncertainty on higher levels. **(a)**, **(b)** for assertions in scenarios similar to (17); **(a)**: case of insufficient information, **(b)**: detected lying, **(c)**: hearer uncertainty: is speaker honest or lying. Abbreviations: φ a proposition, $\overline{\varphi}$ its negation.



and w_3. Clearly, the speaker thinks that in all her epistemically possible worlds an assertion of a sentence *s* with meaning φ is possible (as an extended joint act of asserting), and also thinks that it leads to a unique response. As w_2 is itself an instance of the extended project $\tilde{p}_{ass}$ of asserting φ, it follows that an application of $F^S_{\tilde{p},s}$ to $\{w_2\}$ just returns $\{w_2\}$. Further applications of $F^S_{\tilde{p},s}$ and $F^H_{\tilde{p},s}$ do not change the result, and, hence, w_2 is an element of the Level 1 fixed-point of extended project $\tilde{p}_{ass}$. Analogously, it follows that $\{w_3\}$ is a fixed-point of the extended project of asserting $\overline{\varphi}$.

With c) and d), we switch to referential uses of definite descriptions. Clearly, in c) the interpretation of *the* φ = '*The movie showing at the Roxy tonight*' cannot agree between speaker and hearer, neither in w_2, where *Monkey Business* is showing and a use of *the* φ has to result in a reference to *A Day at the Races*, nor in w_3, where *A Day at the Races* is showing and a use of *the* φ has to result in a reference to *Monkey Business*. In d), however, where *A Day at the Races* is showing but both interlocutors think that *Monkey Business* is showing, *the* φ will from both interlocutors' perspective felicitously refer to *Monkey Business*.

In all examples of **Table 9**, the belief states of interlocutors are subsets of the base level or singleton sets. We can also find natural situations with belief states with uncertainty at higher levels. Examples are shown in **Table 10**. In a) The speaker does not know whether φ or $\overline{\varphi}$ is true, but she is convinced that uttering *s* will lead in all her epistemic possibilities to joint interpretation φ. From the hearer's perspective, the situation is indistinguishable from the base situation. In contrast to b) in **Table 9**, a) is a case of an assertion with insufficient information, hence, a violation of Grice (1975) maxim of quality.

In b) of **Table 10**, a case is shown in which the hearer knows that the speaker is lying but does not know whether φ or $\overline{\varphi}$ is the case. Furthermore, the hearer knows that the speaker thinks him to be gullible. Graph c) seems at first overly complicated, but it represents a natural situation: in it the hearer does not know whether the speaker is honest and says the truth (w_0 and w_1), or is dishonest and lies (w_2 and w_3). Furthermore, the hearer does not know himself whether φ is true, or not. He again knows that the speaker knows the state of the world and that she thinks him to be unsuspecting. For all the worlds, our criterion predicts that the assertion of φ is mutually guaranteed to be successful in the worlds on the left side, and an assertion of $\overline{\varphi}$ in the worlds on the right side.

The examples that we have considered so far show strictly hierarchical belief states. This means, in every possible world that is not in the base level, there is one agent whose belief set has an order type that is smaller than the world's order type. Graphically, this means that the belief set of one agent is a subset of the levels that are below the actual world. More precisely, they are defined as follows:

28) A possible world is *strictly hierarchical*, if for all v in the transitive hull $\overline{T}(w)$ of w it holds that $\mathrm{otp}(v) > 0$ implies:

$$\mathrm{otp}(S^v) < \mathrm{otp}(v) \vee \mathrm{otp}(H^v) < \mathrm{otp}(v). \tag{28}$$

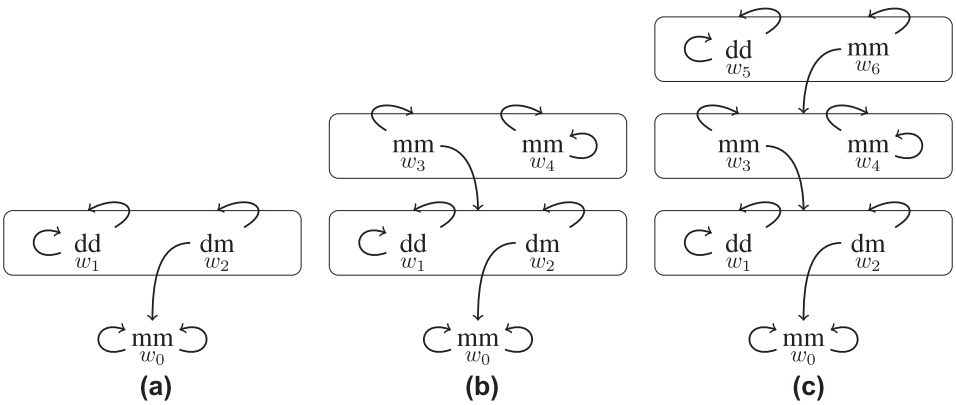
If belief states are not strictly hierarchical, they must show circular relations on higher levels. We consider some examples. **Table 11** shows three epistemic graphs with possible worlds that can be reached from each other via epistemic paths.

We consdier an example:

29) Ann and Bob attend a course on film studies. Together they listen as the lecturer tells the class that, this evening, the course will watch *Monkey Business* at the cinema. Later, in the library, Bob meets the lecturer as she talks to another film student. However, Bob cannot see who the student is. He thinks it is Clara, another student, or Ann. The lecturer notices him and says: "Oh, Bob! Good to see you. I made a mistake. The movie showing this evening is *A Day at the Races*, and not *Monkey Business*." Bob leaves without asking who the other student is. He knows that Ann cannot have learned about the correction if she was not in the library. Later, he receives a mail from Ann telling him that she doesn't like *the movie showing at the cinema tonight.*

What is Ann referring to? The situation is represented by Graph a) in **Table 11**. If Ann was not the other student in the

TABLE 11 | Graphs with circular belief states at higher levels.



library, then, clearly, she refers to *Monkey Business*. If she was there, then she knows that Bob knows that *A Day at the Races* is showing and that Bob knows that the other student knows it too. She also knows that he does not know that the other student was she herself. Hence, if she was the other student then she knows that Bob cannot know what *the φ = the movie showing at the cinema tonight* is referring to. There are two possibilities then: if she was not the other student, she thinks that *the φ* will successfully refer to *Monkey Business*, if she was the student, she should first tell Bob that she learned about the correction, and then refer to *A Day at the Races* with *the φ*. It follows that, if both of them assume that they are rational, that Bob can infer from an utterance of *the φ* that Ann was not the other student, and that she refers to *Monkey Business*. For Graph a) in **Table 11**, this means that the fixed–point construction on the first level should eliminate w_1 but not w_2. Unfortunately, this is not the case. If we first apply the operator checking the speaker's epistemic felicity conditions from (26), then both worlds survive. If we then apply the operator for the hearer's epistemic felicity conditions, then both worlds are eliminated as the *uniqueness* condition is violated for $H^{w_1} = H^{w_2}$.

At this point, we should recall that the iterative application of the felicity operators corresponds to the iterative reasoning about each other and the ensuing step by step elimination of epistemic possibilities that are not consistent with *uniqueness* and *licensing*. The problem with world w_2 is that the speaker's belief state is a subset of the base level, hence, she is oblivious to the reasoning that goes on on the first level. The hearer cannot eliminate w_2 with the argument that the speaker will not make an attempt at referring to *Monkey Business* because she can see that doing this would be inconsistent with the hearer's uniqueness condition. The elimination step in the construction of the fixed–point cannot be applied to worlds with belief states in the lower levels. We say that a world w is *speaker* or *hearer rooted* in the lower level with respect to an act a and a project p, if the speaker's belief state S^w, or the hearer's H^w, are subsets of the lower levels and satisfy the felicity constraints there.

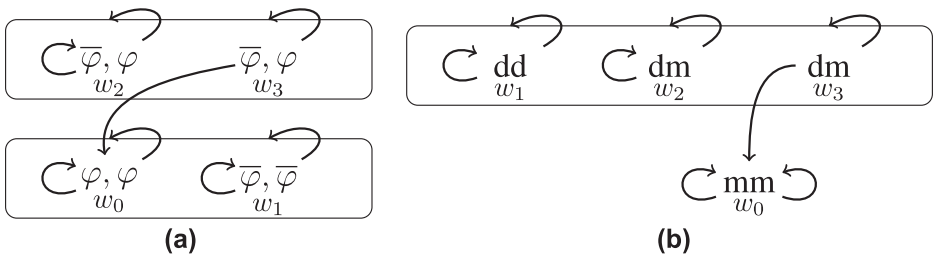
If a world w is rooted in the lower level with respect to an act a and a project p, and if $d^w \in \tilde{p}$ and $a^w = a$, then it should be re-introduced when it is eliminated by a felicity operator during fixed-point construction. For a) in **Table 11** this means that after the elimination of w_2 due to the violation of the hearer's *uniqueness* condition, w_2 has to be re–introduced into the graph. This results into the graph consisting of two worlds, w_0 and w_1, defined by the system of equations consisting of $w_0 = \langle\langle m, the\,\varphi, m\rangle, \{w_0\}, \{w_0\}\rangle$, $w_1 = \langle\langle d, the\,\varphi, m\rangle, \{w_0\}, \{w_1\}\rangle$. This graph also satisfies the two felicity constraints.

The next example is one that shows two levels with circular belief states. It uses the same type of communicative situation with uncertain bystander as Example (29).

30) Ann and Bob attend a course on film studies. Together they listen as the lecturer tells the class that, this evening, the course will watch *Monkey Business* at the cinema together. Later, in the library, Bob meets the lecturer as she talks to another film student. However, Bob cannot see who the student is. He thinks it is Clara, another student, or Ann. The lecturer notices him and says: "Oh, Bob! Good to see you. I made a mistake. The movie showing this evening is *A Day at the Races*, and not *Monkey Business*. Bob leaves without asking who the other student is. Still later, he meets the lecturer again in the cafeteria. She tells him that the program has changed again. Then Ann comes in. The lecturer tells her: "Hallo Ann, I have just told this student here that the program changed again. It is *Monkey Business* that is showing tonight." Bob noticed that Ann could not see him, that she must think that it could be him but that she could not be certain. He also knew that she must think that he could not learn about the change of program if he was not the student in the cafeteria. Bob also noticed that Ann must have been the other student in the library. Later, he receives a mail from Ann, telling him that she doesn't like *the movie showing at the cinema tonight*.

The situation is represented by b) in **Table 11**. It can be easily checked that the fixed–point on the second level is identical to the level consisting of w_3 and w_4. The fixed–point of the first level again consists of only w_2.

TABLE 12 | Graphs with circular belief states at higher levels.



In principle, we can add more and more levels with circular structure. Graph c) in **Table 11** shows an example with four levels. As world w_6 on Level 3 is rooted in level 2, it is not eliminated when the fixed–point on Level 3 is constructed. It is then predicted that a reference to *Monkey Business* in w_6 with *the movie showing at the cinema tonight* is felicitous, whereas a reference to *A Day at the Races* in w_5 is not felicitous.

As final example in this section, we present a situation that resembles (29) but is not about reference but about assertions.

31) Helga calls up her son Stephan and asks him whether he wants to visit her in Munich. Stephan tells her that he will watch the weather forecast this evening and call her in the morning. Helga knows the channel where Stephan watches the late news and learns that it is snowing in the Alps the next day. Next morning a mutual friend video calls her and mentions that the forecast has changed and that the streets are free of snow. In the background, Helga can see someone who resembles her son Stephan, but she cannot be sure. Shortly afterward, she receives a text message from Stephan saying that he cannot visit her because snow is forecasted and he doesn't want to drive then. He also wrote that he will not have his smartphone with him and cannot read text messages that day. She knows that Stephan has a new girl-friend and prefers to stay at home.

Is Stephan lying about snow in the Alps, or not? The situation is represented by the graph in **Table 12a**.

World w_3 is rooted in the base level, and anchored to a world in which the speaker is licensed to assert φ = '*It is snowing in the Alps.*' As it is itself an instance of the extended project of asserting φ, it will be in the fixed–point on Level 1. World w_2 is also an instance of the extended project, the speaker is licensed to assert φ in all epistemic possibilities, and the hearer's belief state also satisfies *licensing* of asserting φ. As mentioned before, *uniqueness* is trivially satisfied for assertions. Hence, w_2 will also be in the fixed–point. The prediction is then that Helga cannot tell whether Stephan lied or said what he believed to be true.

What is the difference between the graphs in **Tables 11a** and **12a**? The answer is that we chose a minimal representation of (29) in **Table 11a**. We saw in (18) and **Table 4** that the basic utterance situation for assertions can come in different varieties. The textual description of the utterance situation in (17) leaves the exact epistemic relations between speaker and hearer underspecified. The same underspecification is encountered with Example (29). An alternative to the graph in **Table 11a** is shown in **Table 12b**.[9] Here, world w_2 corresponds to w_2 in **Table 12a**. Both survive the tests for *licensing* and *uniqueness* conditions and the subsequent updates. In the case of assertions, there cannot be a possibility corresponding to world w_1 in **Tables 11a** and **12b**, as there is no ambiguity about semantic interpretation equivalent to ambiguity about choice of referent.

6 COMPARISON AND OUTLOOK

We developed a theory of epistemic felicity conditions and speech acts that followed a path charted by the works of J. Searle, H.P. Grice, and H.H. Clark. For both assertions and definite descriptions there is a large body of literature, so large that we can only hint at how our model fits into the general landscape of semantic and pragmatic theories. For both referential uses of definite descriptions and assertions we make minimal assumption about dialogue context. In our model, *familiarity* (Heim, 1982) and *uniqueness* (Russell, 1905, 1919) of referents are not semantic properties of definite descriptions but follow from pragmatic felicity conditions that hold in very basic epistemic graphs only. If the felicity conditions are not met, then the referent remains undefined (see Strawson, 1950). Our model also accounts for situation in which the decription of a definite does, or may not apply to the referent as in Donnellan's (1966) famous Martini-glass example (an example is shown in **Table 11**, Graph a)). For assertions, our constitutive rules only require that the asserted proposition is true (Weiner, 2005), from which the requirement that the speaker believes it (Williamson, 1996; Turri, 2016) follows as a felicity requirement of basic utterance situations, but it may be violated at higher order belief states. In particular, our model shows how the existence of non–cooperative language use and un-truthfulness can be reconciled with the constitutive requirement of truthfulness (see Pagin 2016 for an overview of the related philosophical discussion).

Our model is about epistemic felicity conditions of speech acts. Which speech acts can be performed in a dialogue situation is pragmatically dependent on the interlocutors' beliefs about the

[9]There are, in fact, an infinitude of alternatives. We leave the clarification of this issue to future research.

world and about each other. There are theories that try to predict possible speech acts without reference to private beliefs. Prominent examples are *commitment* theories, *discourse structural* approaches, or approaches based on the idea of common *scoreboards*. In a commitment approach, if a speaker asserts a sentence then s/he takes on the (social) obligation of defending its truth; s/he does, however, not necessarily express a belief in it.[10] Discourse structural approaches explain the possible sequences of speech acts by discourse relations that must hold between dialogue moves. Example are the *Segmented Discourse Representation Theory* (Asher and Lascarides, 2003) and the *Rhetorical Structure Theory* (Mann and Thompson, 1988). Relevant is here, for example, the account of strategic conversation in non–cooperative discourse by Asher and Lascarides (2013). The idea that information update in dialogue can be modeled with public scoreboards can be traced back to Lewis (1979). The scoreboard represents the public information of interlocutors. Each communicative act updates the scoreboard in specific ways. In ideal cases, the update only depends on the old scoreboard and the sentence uttered. Hence, the update after an honest assertion and a lie would be the same. A comparison of our model to any one of these approaches would go beyond the scope of this article. A common motivation for all of them are the problems that epistemic accounts of speech acts face when confronted with non-cooperative discourse or utterance situations with higher–order belief states. Our model shows how these problems can be overcome.

In the previous sections we have seen how the interlocutors' limited perspectives can give rise to extended uses of communicative acts. On the base level, where interlocutors follow constitutive rules and have only truthful beliefs, the joint communicative acts that mutually satisfy the epistemic felicity conditions of *licensing* and *uniqueness* can be found by a fixed-point construction. The fixed-point construction depends only on a given set of joint projects, hence, it generalizes to any type of communicative act, the constitutive uses of which can be represented by joint communicative acts of the form $\langle m, a, r\rangle$, i.e., as a set of triples consisting of a state of affairs m, an utterance act a, and a response r.

The elements of $\langle m, a, r\rangle$ are abstract representations of the state of affairs, acts and responses. For example, the state of affairs m can represent a concrete situation in the world, but it can also represent a more abstract dialogue scoreboard. As an example, we may consider Ginzburg (2012) KoS framework. In this framework a and r would each be the latest moves in a pair of dialogue states representing the precondition and effects of performing the respective speech act. If a and r belong to a joint project, then the effect state of a must be a sub-type of the preconditions of r. By identifying the pre-state of a with m, we can see how adjacency pairs in the KoS-format can be translated into joint communicative acts of the form $\langle m, a, r\rangle$, and, thereby, plugged into our epistemic model. In this way, our model could benefit from the additional fine-structure that KoS has to offer. It also shows how a scoreboard approach and an epistemic approach as the one proposed in this article can be reconciled. In contrast to chess, dialogue game boards are not physically given. They have to be maintained and coordinated by speaker and hearer, and so perspectives must have a role to play.[11]

We said before that constitutive rules define a form of social institution consisting in a conventionalized regularity of linguistic behavior. In the following, we tied this behavior to a class of well–behaved utterance situations at the base of the epistemic graphs that we have seen. From there, the behavior is extended to a wider class of hierarchical epistemic states. We have seen that, in extending the behavior, indistinguishability between utterance situations play a crucial role. The extended use of a communicative act can travel up the hierarchy along epistemic paths because the situation on the higher level is for one, or sometimes both interlocutors indistinguishable from one at a lower level.

If a communicative act is defined for a constitutive core, then our theory also predicts that extended uses that violate the constitutive rules exist. Hence, if honest, truthful assertions exist, then also assertions based on false beliefs must exist, as well as outright lies. This also means that the definitions of speech acts can be simplified considerably, as only constitutive rules for uses in the well–behaved core have to be considered.[12] A non-trivial observation is that extended uses can still be classified with the same name as the uses in the constitutive core. In the introduction we mentioned the following examples in (32) and (33):

32) Leo told me that it is snowing in the Alps, but I knew that the snowing had stopped. (false belief).
33) Leo told me that it is snowing in the Alps, but I knew that she is lying. (lying).

Assuming that *tell* reports an assertion event, then the examples show that classifying an utterance as *assertion* is consistent with false beliefs and lying. This raises a question about the semantics of *tell*. If the constitutive rules were part of the semantic meaning of assertions, then, given how we have defined the constitutive joint project of *asserting*, the sentence '*Leo told me that* φ' should mean that Leo uttered a sentence with meaning φ and she uttered it in a situation in which this sentence is true. Clearly, the examples in (3′) and (4′) are not consistent with such a semantic rule. **Table 13** shows two graphs for the examples.

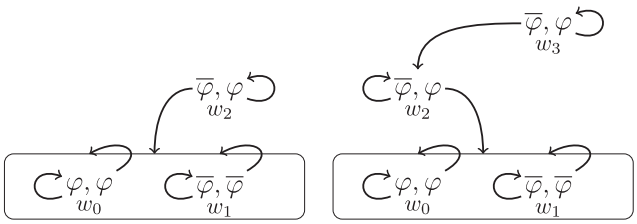
In our model, we distinguished between the project as defined by its constitutive rules and the extended project that is defined by the action–response pairs alone. This means, if p is a joint project, then the extended project $\tilde{p}$ is $\{\langle m, a, r\rangle \mid \exists m' \langle m', a, r\rangle \in p\}$. We make two assumptions: a) '*tell*' semantically applies to joint communicative acts in the constituting joint project of assertions p_{ass}, but it can be extended to joint communicative acts of the extended joint project $\tilde{p}_{ass}$; b) pragmatically an

[10]There are, however, various meanings that have been given to the term *commitment*. For an older overview, see (Brabanter and Dendale, 2008). For recent discussions, see e.g., (Krifka, 2012; Geurts, 2019; Krifka, 2019).

[11]In line with H.H. Clark's propositions 3 and 6, (1996, p. 23/24).

[12]However, we have to concede that extending the account to cover intricate problems that motivate, for example, dynamic syntactic theories like DS-TTR (see Gregoromichelaki et al., 2011) needs further work (e.g., on the problem of *split turn taking*; see Gregoromichelaki and Kempson, 2016 for an overview; I thank the reviewers for bringing this important phenomenon to my attention).

TABLE 13 | Epistemic graphs for examples of *assertions* with false beliefs and lying.



application of '*tell*' to a joint project is felicitous only if the denoted joint communicative acts is uttered in a world that belongs to a fixed–point of either the base, or one of the higher levels of an epistemic graph. These assumptions allow the felicitous use of '*tell*' to travel up the paths of an epistemic graph as indicated in the introduction. They also explain how the reports in (32) and (33) can be felicitous. In **Table 13**, the theory predicts a felicitous use of φ in world w_2 for (32) and in world w_3 for (33). The two assumptions entail that '*tell*' can be felicitously used for reporting the utterance events in these worlds.

This solution assumes that lexical meaning is flexible and allows for contextual adjustment taking the interlocutors perspectives and the resulting indistinguishability between utterance situations into account. There are other paths for seeking a solution that come to mind. For example, one could assume that the lexical meaning of *tell* has a meaning that is weak enough to be consistent with all epistemic graphs that we have considered in this article. *Commitment* approaches belong here. We must, however, leave the comparison and further pursuit of the semantic issues to future research.

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.

AUTHOR CONTRIBUTIONS

AB is the sole author of this article.

FUNDING

European Research Council, award number(s): ERC 787929 SPAGAD: Speech Acts in Grammar and Discourse. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

REFERENCES

Aczel, P. (1988). *Non-well-founded Sets*. Stanford, CA: Center for the Study of Language and Information.

Asher, N., and Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Asher, N., and Lascarides, A. (2013). Strategic Conversation. *Semantics and Pragmatics* 6, 1–62. doi:10.3765/sp.6.2

Baltag, A., and Renne, B. (2016). "Dynamic Epistemic Logic," in *The Stanford Encyclopedia of Philosophy*. Editor E. N. Zalta (Winter: Metaphysics Research Lab, Stanford University).

Baltag, A., van Ditmarsch, H. P., and Moss, L. S. (2008). "Epistemic Logic and Information Update," in *Handbook on the Philosophy of Information*. (Amsterdam: Elsevier Science Publishers), 361–455. doi:10.1016/b978-0-444-51726-5.50015-7

Barwise, J., and Etchemendy, J. (1989). *The Liar: An Essay on Truth and Circularity*. Oxford: Oxford University Press.

Barwise, J., and Moss, L. (1996). *Vicious Circles*. Stanford, CA: Center for the Study of Language and Information.

Barwise, J. (1989). "On the Model Theory of Common Knowledge," in *The Situation in Logic*. Editor J. Barwise (Stanford, CA: CSLI Lecture Notes), 201–220.

Benz, A. (2008). "Assertions in AFA Set Theory," in *New Approaches to Classes and Concepts*. Editor K. Robering (London: College Publications), 143–176.

Benz, A. (2012). On a Super Large Fixed-point of Common Information in Multi-Agent Signalling Games. *Logic J. IGPL* 20, 94–120. doi:10.1093/jigpal/jzr027

Clark, H. H., and Marshall, C. R. (1981). "Definite Reference and Mutual Knowledge," in *Elements of Discourse Understanding*. Editors A. K. Joshi, B. Webber, and I. Sag (Cambridge: Cambridge University Press), 10–63.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

De Brabanter, P., and Dendale, P. (2008). Commitment: the Term and the Notions. *Belgian J. Linguistics* 22, 1–14. doi:10.1075/bjl.22.01de

Donnellan, K. S. (1966). Reference and Definite Descriptions. *Phil. Rev.* 75, 281–304. doi:10.2307/2183143

Fagin, R., Halpern, J. Y., Moses, Y., and Moshe, Y. V. (1995). *Reasoning about Knowledge*. Cambridge MA: MIT Press.

Gerbrandy, J. (1998). *Bisimulations on Planet Kripke. Ph.D. Thesis, Institute for Logic, Language and Computation*. Amsterdam: Universiteit van Amsterdam.

Gerbrandy, J., and Groeneveld, W. (1997). Reasoning about Information Change. *J. logic, Lang. Inf.* 6, 147–169. doi:10.1023/a:1008222603071

Geurts, B. (2019). Communication as Commitment Sharing: Speech Acts, Implicatures, Common Ground. *Theor. Linguistics* 45, 1–30. doi:10.1515/tl-2019-0001

Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press.

Gregoromichelaki, E., and Kempson, R. (2016). "Joint Utterances and the (Split-) Turn Taking Puzzle," in *Interdisciplinary Studies in Pragmatics, Culture and Society. Perspectives in Pragmatics, Philosophy & Psychology*. Editors A. Capone and J. L. Mey (Switzerland: Springer International Publishing), 703–743. doi:10.1007/978-3-319-12616-6_28

Gregoromichelaki, E., Kempson, R., Purver, M., Mills, G. J., Cann, R., Meyer-Viol, W., et al. (2011). Incrementality and Intention-Recognition in Utterance Processing. *Dialogue and Discourse* 2, 199–233. doi:10.5087/dad.2011.109

Grice, H. P. (1975). "Logic and Conversation," in *Syntax and Semantics*. Editors P. Cole and J. L. Morgan (New York: Academic Press), 41–58. doi:10.1163/9789004368811_003

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Amherst: The University of Massachusetts. Ph.D. thesis.

Hintikka, J. (1962). *Knowledge and Belief*. Ithaca and London: Cornell University Press.

Korta, K., and Perry, J. (2020). "Pragmatics," in *The Stanford Encyclopedia of Philosophy*. Editor E. N. Zalta (Stanford: Metaphysics Research Lab, Stanford University).

Krifka, M. (2012). "Commitment Space Semantics: Declarative Questions, Negated Questions, and Question Tags," in *Proceedings of the 25th Semantics and Linguistic Theory Conference (SALT)*. Editors S. D'Antonio, M. Moroney, and C. R. Little.

Krifka, M. (2019). Commitments and beyond. *Theor. Linguistics* 45, 73–91. doi:10.1515/tl-2019-0006

Lewis, P. M. E. (1979). Editorial. *Org. Magn. Reson.* 12, 339. doi:10.1002/mrc.1270120602

Mann, W. C., and Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8, 243–281. doi:10.1515/text.1.1988.8.3.243

Morris, C. W. (1938). "Foundations of the Theory of Signs," in *International Encyclopedia of Unified Science*. Editors O. Neurath, R. Carnap, and C. W. Morris (Chicago: The University of Chicago Press), 1–59.

Pagin, P. (2016). "Assertion," in *The Stanford Encyclopedia of Philosophy*. Editor E. N. Zalta (Winter: Metaphysics Research Lab, Stanford University).

Russell, B. (1905). On Denoting. *Mind* XIV, 479–493. doi:10.1093/mind/xiv.4.479

Russell, B. (1919). *Introduction to Mathematical Philosophy*. London: George Allen and Unwin.

Schegloff, E. A., and Sacks, H. (1973). Opening up Closings. *Semiotica* 8, 289–327. doi:10.1515/semi.1973.8.4.289

Searle, J. R. (1969). *Speech Acts*. Cambridge: Cambridge University Press.

Strawson, P. F. (1950). On Referring. *Mind* LIX, 320–344. doi:10.1093/mind/lix.235.320

Turri, J. (2016). *Knowledge and the Norm of Assertion: An Essay in Philosophical Science*. Cambridge: Open Book Publishers. doi:10.11647/OBP.0083

van Benthem, J. (2011). *Logical Dynamics of Information and Interaction*. Cambridge: Cambridge University Press.

Weiner, M. (2005). Must We Know what We Say? *Phil. Rev.* 114, 227–251. doi:10.1215/00318108-114-2-227

Williamson, T. (1996). Knowing and Asserting. *Phil. Rev.* 105, 489–523. doi:10.2307/2998423

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 Benz. 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.

7 APPENDIX

7.1 How to read epistemic graphs

Table 14 shows how to read epistemic graphs. We assume that there is a speaker *A* and an addressee *B*. Arrows to the left of a world point to *A*'s information state, and arrows to the right to *B*'s information state. An information state is a set of possible worlds. If an arrow points to a single world *w*, then the respective information state is a set with *w* as single element. If an information state has more than one element, it is represented by a box encircling its elements. For convenience, basic building units of graphs are shown in the **Table 14**.

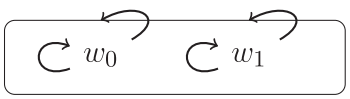
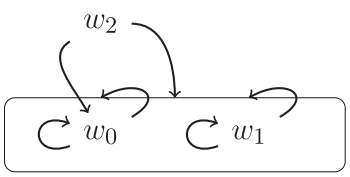
7.2 Epistemic Graphs and Kripke Frames

We address the question how epistemic graphs are related to *Kripke frames* for epistemic modal logic. Epistemic modal logic can be traced back to (Hintikka, 1962). For a newer introduction and an overview see e.g., (van Benthem, 2011) and (Baltag et al., 2008; Baltag and Renne, 2016). In epistemic modal logics, epistemic possibility is modeled by accessibility relations between worlds. For each agent *i* there is a relations R_i between possible worlds with the meaning that if $\langle v, w\rangle \in R_i$ then *i* believes in world *v* that *w* represents a possible state of affairs. If there are two agents, a speaker and a hearer, then the beliefs of each one is represented by his/her own accessibility relation R_S or R_H. Given a set of possible worlds $\mathcal{W}$ and accessibility relations R_S and R_H an equivalent epistemic graph is defined by the following system of equations:

$$v = \langle v, \{w \mid R_S(v, w)\}, \{w \mid R_H(v, w)\}\rangle, \quad v \in \mathcal{W}. \tag{29}$$

In reverse, if a system of equations is given that defines an epistemic graph, and $\mathcal{W}$ is the set of solutions, i.e., all $v \in \mathcal{W}$ are of the form $v = \langle d^v, S^v, H^v\rangle$, then $R_S := \{\langle v, w\rangle \mid v, w \in \mathcal{W} \wedge w \in S^v\}$ and $R_H := \{\langle v, w\rangle \mid v, w \in \mathcal{W} \wedge w \in H^v\}$ are the accessibility relations of the corresponding Kripke frame. If a modal logic with belief operators for speaker and hearer is interpreted in the Kripke frame and the epistemic graph, then the two constructions are equivalent in the sense that corresponding worlds make the same modal logic formulas true. As we are not concerned with modal logics but directly reason with epistemic graphs, there is nothing to be gained by using Kripke frames. For our purposes, Kripke frames have disadvantages. For example, the graph defined by $w = \langle m, \{w\}, \{w\}\rangle$ could be represented in infinitely many ways by equivalent Kripke frames. Hierarchical and circular structures and the order types of worlds are not immediately definable. Their definition would have required normalization with respect to maximal *bisimulations* (see Barwise and Moss, 1996; Gerbrandy and Groeneveld, 1997; Gerbrandy, 1998). AFA-set theory allows one to avoid this step.

TABLE 14 | Reading epistemic graphs. The comments to the right explain new features of the respective graphs.

	arrow to the left is pointing from w_0 back to w_0: A's information state in w_0 is $\{w_0\}$; arrow to the right is pointing from w_0 back to w_0: B's information state in w_0 is $\{w_0\}$;
	two worlds with arrows to the right that both point to the box encircling w_0 and w_1; they show that B's information states in w_0 and w_1 both equal $\{w_0, w_1\}$ (encircled worlds);
	arrow to the left of w_2 is pointing at w_0; it shows that A's information state in w_2 is $\{w_0\}$; arrow to the right of w_2 is pointing at the box encircling w_0 and w_1; it shows that B's information state in w_2 equals $\{w_0, w_1\}$ (encircled worlds);

frontiers in Communication

ORIGINAL RESEARCH
published: 07 September 2021
doi: 10.3389/fcomm.2021.624486

In a Manner of Speaking: How Reported Speech May Have Shaped Grammar

Stef Spronck and Daniela Casartelli*

University of Helsinki, Helsinki, Finland

We present a first, broad-scale typology of extended reported speech, examples of lexicalised or grammaticalised reported speech constructions without a regular quotation meaning. These typically include meanings that are conceptually close to reported speech, such as THINK or WANT, but also interpretations that do not appear to have an obvious conceptual relation with talking, such as CAUSE or BEGIN TO. Reported speech may therefore reflect both concepts of communication and inner worlds, and meanings reminiscent of 'core grammar', such as evidentiality, modality, aspect (relational) tense and clause linking. We contextualise our findings in the literature on fictive interaction and perspective and suggest that extended reported speech may lend insight into a fundamental aspect of grammar: the evolution of verbal categories. Based on the striking similarity between the meanings of extended reported speech and grammatical categories, we hypothesise that the phenomenon represents a plausible linguistic context in which grammar evolved.

Keywords: reported speech, quotation, perspective, TAME, fictive interaction, evolution of language, linguistic typology

OPEN ACCESS

Edited by:
Sonja Zeman,
Ludwig Maximilians Universität München, Germany

Reviewed by:
Sergeiy Sandler,
Independent researcher, Beer Sheva, Israel
Henrik Bergqvist,
Stockholm University, Sweden

****Correspondence:***
Stef Spronck
stef.spronck@helsinki.fi

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *31 October 2020*
Accepted: *12 July 2021*
Published: *07 September 2021*

Citation:
Spronck S and Casartelli D (2021) In a Manner of Speaking: How Reported Speech May Have Shaped Grammar. Front. Commun. 6:624486. doi: 10.3389/fcomm.2021.624486

1 INTRODUCTION: FICTIVE INTERACTION, REPORTED SPEECH AND GRAMMAR

The act of speaking is so fundamental to the human experience and our perception of other people that we routinely cast our interaction with the world as a dialogue (Bakhtin, 1981). We may represent intentions of others, even those attributed to non-humans, as 'speaking to us', as in (1a) or (1b). In these sentences a perception or, e.g., the expression of the face of an animal is described as a speech event.

(1) a. One look said it all.
b. Its wrinkled snout told me: "Don't come near!"

In a series of seminal accounts (Pascual, 2002; Pascual, 2007 and, especially, Pascual, 2014) characterises expressions as in (1) as 'fictive interaction', defined as the use of 'conversation as a frame to structure mental, discursive, and linguistic processes' (Pascual, 2014, 9). While this analysis is explicitly embedded in a cognitive linguistic account that sees conversation as a cognitive *Gestalt* that humans may use in order to make sense of the world, its empirical foundation is strong. Not only are examples of fictive interaction as in (1) common cross-linguistically (Pascual and Sandler, 2016a; McGregor, 2019), they affect a heterogeneous set of sentence types and linguistic structures (Pascual and Sandler, 2016b).

While in many Standard Average European languages fictive interaction appears to be a metaphor-driven type of creative language use (apart from possibly more idiomatic expressions like to 'speak to oneself'), a much more conventionalised form of fictive interaction occurs in many languages around the world, cf. (2).

(2) WAN (Mande; Ivory Coast)

yī ē gé bā kɔ́
water DEF say LOG boil

'The water was about to boil.' (lit., 'The water said: let me boil!') (Nikitina, 2020, 88)

Like in (1), (2) presents a 'speech event' that does not involve actual communication. The literal translation of (2) clearly demonstrates this: the water incites its own boiling, as if it was speaking. Crucially, however, this is not what (2) *means*. As the idiomatic gloss in (2) illustrates, the interpretation of this sentence is an inceptive meaning, i.e. 'to be about to do *p*'. Rather than *saying* "let me boil", as the lexicogrammatical elements in (2) suggest, the only plausible contextual meaning of the sentence is 'The water was about to boil'.

Whereas it is intuitively obvious to the reader that the interpretation of the 'fictive' example of direct speech in (1b) is metaphorical and is interpreted through a process of inference, the structural, semantic and pragmatic status of examples like (2) are much less clear. This is problematic, because this is not an isolated or anecdotal example. Examples that, judging by their lexico-grammatical components would seem to carry a meaning of direct speech (x said: "*p*"), may receive widely varying interpretations in the languages of the world. These interpretations include meanings that seem to have little in common with speech events, or the perspective-shifting function associated with reported speech.

In a first cross-linguistic typology of the phenomenon, specifically focusing on direct speech-like structures as in (2), Pascual (2014, 90) distinguishes meanings as varied as (1) mental states, (2) emotional and attitudinal states, (3) desires, (4) intentions, (5) attempts, (6) states of affairs[1], (7) causation, (8) reason, (9) purpose and (10) future tense. We may group these meanings into the four classes in (3).

(3) THOUGHT: mental states, desires, intentions, attempts;
ATTITUDE: emotional states;
CAUSE: causation, reason, purpose;
TIME: states of affairs, future tense.

Despite the extensive list of meanings Pascual's pioneering study uncovers, it raises several important questions. First of all: what is the status of the meanings in (3)? In order to answer this question we will need to understand, firstly, if the list in (3) is exhaustive with respect to the range of meanings attested in similar examples so far, secondly, if these meanings are random or show recurrent patterns in the languages of the world and, thirdly, what mechanisms give rise to the meanings as in (3)? These are questions the current article aims to address.

[1]Pascual (2014) uses this label to describe interpretations relating to the internal organisation of an event (particularly inchoative meanings; see **Section 2.2.2**). This class roughly corresponds to the function we will describe as 'aspect' below.

This objective immediately faces a methodological challenge: both in (1b) and (2), the lexico-grammatical make-up of the examples suggests a 'literal' direct speech interpretation, but the actual meaning in context varies, as well as the way in which this meaning arises pragmatically. As we will see below, the semantically based notion of 'direct speech' does not neatly apply to all relevant instances. This suggests that in order to examine the variation of the meanings involved, we need to start with a definition of a class of relevant examples based on their lexico-grammatical properties. For the sake of cross-linguistic comparison, this set of lexico-grammatical properties cannot be too restrictive, since it needs to be applicable to languages of distinct structural types. We cannot make *a priori* assumptions about the language-specific variation that might exist between these structures. On the other hand, it needs to capture a class of phenomena that are cross-linguistically comparable and can be identified based on the definition, so it cannot be too inclusive either (Haspelmath, 2010). We will return to the wider context of fictive interaction at the end of this article, but in order to maximally avoid the presumption of metaphoricity implicit in this label we will refer to the typological examples examined in this article as 'extended reported speech'. The identification of relevant reported speech examples will be based on the definition in (4a). Extended reported speech will be defined as in (4b).

(4) a. For the purposes of this study, a reported speech construction is a clausal or bi-clausal linguistic structure consisting of a Matrix unit (M) that has scope over a Report unit (R) that, at least under one of its interpretations, can be understood as a representation of an utterance. M minimally consists of or contains an element that can be translated as 'say'.
b. For the purposes of this study, extended reported speech is a reported speech construction that is *not* interpreted as representing a speech act.

We begin in **Section 2** with an extensive illustration of the definitions, or 'comparative concepts' (Haspelmath, 2010), in (4a) and (4b), showing how they can be applied across languages and what type of examples they unveil. These illustrations may also clarify some of the specific formulations in the definitions above, so we will address further motivations for the comparative concepts in (4a) and (4b) in **Section 2**. The section begins with a brief contextualisation of the typological and descriptive literature on which we will draw (**Section 2.1**), before exemplifying some of the main attested types of extended reported speech in **Section 2.2**. These observations both support and expand the initial classification in (3), and **Section 2.3** presents an updated list of extended reported speech meanings.

As we will show, two meanings that seem particularly well documented so far in extended reported speech constructions are those with an intention reading (i.e. a 'want' meaning) and those with a complementiser meaning. These are two types we will explore in more detail in **Section 3**, based on a cross-linguistic sample of 100 genetically diverse languages. These are the first results of a sample study aiming to develop a broad-scale typology of extended reported speech. The sample and methodology of the study are introduced in **Section 3.1** and **Section 3.2** presents and illustrates the results. The observations and initial analyses from **Sections 2** and **3** are summarised and integrated in **Section 3.3**.

With this empirical foundation in place, in **Section 4** we suggest some implications of our observations for the understanding of the semantics of reported speech, perspective constructions, particularly the subtypes of reported speech and, speculatively, the evolution of grammar. In **Section 4.1** we relate the extensions of reported speech to the semantics of reported speech constructions and identify three pathways towards extended reported speech, specifically, recasting, rescaling and semantic bleaching. The implications of extended reported speech for our understanding of the diachrony of perspective expressions in language and theories of classifications of direct and indirect speech are explored in **Section 4.2**. We outline our main motivation behind this project in **Section 4.3**, where we claim that the observations about extended reported speech demonstrate so many similarities with the meaning of common grammatical categories that the phenomenon holds fundamental implications for the evolution of grammar.

Finally, **Section 5** presents a brief conclusion.

2 STUDIES ON EXTENDED REPORTED SPEECH: A SURVEY

2.1 Background

Extended reported speech has been relatively well documented in the descriptive and typological literature. One of the earliest in-depth studies of the phenomenon appears to be Larson (1978), who discusses reported speech in the South-American language Aguaruna, demonstrating that it can be used to express meanings far beyond speech representation, cf. (5).

(5) AGUARUNA (Chicham; Perú)

a. *Wetatjai, "Pujusti," tau san.*
I-will-go that-he-stay one-who-said I-being
Meaning in context: 'I will go because I want him to stay.' (Larson, 1978, 93)
Mock English gloss: '{I will go} "that he stays," I am the one who says'

b. *Shutuktut tawa nunak ayujut taku tawai.*
shutuktut it-says that-obj-topic give-to-eat saying it:says
Meaning in context: ' 'Shutuktut' means to give someone something to eat.' (Larson, 1978, 113)
Mock English gloss: "Shutuktut" it says that "give to eat" saying it says

Following the definitions in (4), the examples in (5) illustrate extended reported speech since they both contain Report and Matrix units that, as the glosses illustrate, can be interpreted as representing reported utterances and clauses of saying, respectively (cf. 4a). Yet, as the idiomatic glosses (i.e., the third line of the examples) illustrate, the contextual interpretation of these examples does not involve a speech event (cf. 4b). This is how we will apply the definitions throughout this study: the comparative concept of a reported speech construction (4a) is evaluated against the morphemic gloss (i.e., the second lines of the examples), that of extended reported speech (4b) against the idiomatic gloss (i.e., the third lines of the examples).

In order to increase readability, we will also add a fourth line to each example, as in (5). This line is a mock English gloss that represents what the example could be expected to mean based on its lexico-grammatical content, i.e. it is a prose interpretation of the morphemic glosses[2]. Crucially, however, the Mock English gloss should not be taken to indicate the actual meaning of the full example; it is a presentational device in order to make the morphemic gloss more accessible[3]. In order to highlight this interpretative status, the fourth line also appears in a different font, below the translation given in the source. Elements placed between curly brackets in the Mock English glosses (as in 5a) are not part of the extended reported speech construction.

Apart from in Aguaruna, extended reported speech has been attested in languages across South America (van der Voort, 2002; Everett, 2008; Birchall, 2018). Several studies have described it as a regional phenomenon, occurring in languages in the Tibetan area (Saxena, 1988), in Africa (Güldemann, 2008), among Sinitic languages (Chappell, 2012) and across Siberia (Matić and Pakendorf, 2013). Furthermore, numerous studies of extended reported speech in Australia (Rumsey, 1990; McGregor, 2014, cf.), Austronesia and Papunesia (Deibler, 1971; Reesink, 1993; Klamer, 2000, cf.) and Central Asia Baranova (cf. 2015) have established it as a common phenomenon in languages of these areas as well. **Figure 1** shows the location the languages cited in this section, illustrating that descriptions of extended reported speech are not restricted to any particular geographical area or language family[4].

Given how widespread the phenomenon of extended reported speech appears to be across the languages of the world, it is not surprising that it can carry many diverse meanings. What calls for an explanation, however, is the observation that these meanings, while wide-ranging, seem far from random. **Figure 2** summarises the most frequently occurring meanings described in the studies on extended reported speech that we will survey in this section.

The labels in **Figure 2**, which we will refer to as 'functions' or 'interpretations', represent a short, standardised summary of the meaning description given in each of the sources. The descriptions and classifications for the individual languages are listed in **Appendix Table A1**. Following this standardised list, (5a) carries a WANT function and (5b) a NAME function. Throughout this section we will introduce and illustrate each of the functions in **Figure 2**, as well as a few less commonly described ones.

[2]Since the segmentation and labelling in the morphemic gloss (i.e., the second line in all examples) reflect careful analytical choices on the part of the individual authors cited, most glosses in the examples introduced here have been preserved from the original source reference. Where these include abbreviations that do not follow the standard of the Leipzig Glossing Rules (Bickel et al., 2008) these are listed in the glossary at the end of this article. The only exception to this practice has been glosses that conflict with those in the Leipzig Glossing Rules, as in (13a), which uses 'S' for 'singular', whereas it indicates an intransitive subject in Bickel et al. (2008). This example also contains the gloss 'DEC' for 'declarative', which is minimally distinct from the standardised gloss 'DECL'. In such cases we have revised the glosses in accordance with the Leipzig Glossing Rules and have explicitly indicated this in the reference by adding 'gloss updated'.

[3]We will avoid the misleading term 'literal meaning' to refer to this line, since it suggests that the meaning indicated in the translation line (i.e., the third lines of the examples) is a metaphorical interpretation of the Mock English translation, which we do not assume to be the case for these examples.

[4]The larger dots on the map represent areal studies, which include multiple languages near the indicated location.



FIGURE 1 | Map of locations of languages cited in this section.

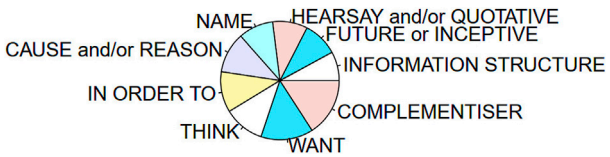


FIGURE 2 | Common functions in descriptions of extended reported speech (based on the studies listed in **Appendix Table A1**).

In order to provide an initial grouping into separate types, we will distinguish between examples of extended reported speech that, although conventionalised, are still clearly identifiable as reported speech constructions (**Section 2.2.1**) and those that show more signs of grammaticalisation (**Section 2.2.2**). Since we will focus on the functions of these examples and languages may show varying degrees of grammaticalisation in their forms of extended reported speech, these groupings are not entirely clear-cut, nor mutually exclusive, but they allow us to most clearly connect with existing descriptions in the literature.

2.2 Examples of Extended Reported Speech

2.2.1 Lexicalised and Conventionalised Examples

Perhaps the most common type of extended reported speech is examples with a 'think' interpretation. On a trivial level, this use of reported speech may seem familiar to Standard Average European expressions like 'I would say *p*', which signals 'I think *p*', and, naturally, saying *p* implies thinking *p*. However, the extended reported speech version of this interpretation arises in languages in which the distinction between reported speech and reported thought is principally underspecified, cf. (6).

(6) KAVALAN (Austronesian; Taiwan)

"sessen=ti sunis unay," zin-ku; bula-an-ku tu qudus-ku.
AF.cold=PFV child that say-1SG.GEN give-LF-1SG.GEN OBL clothes-1SG.GEN

' "That child was cold," I thought; I gave my clothes to him.' (Hsieh, 2012, 467)

ME: "That child was cold," I said. {I gave my clothes to it}

In (6), the verb that constitutes the Matrix unit, while glossed as 'say', could equally mean 'say' or 'think'. Hsieh (2012), 467 writes about this example: 'when no obvious addressee can be found in the clause, this may pose some difficulties in deciding whether the term in question denotes an act of speaking or an act of thinking. The correct interpretation depends heavily on pragmatic inferences'. In the languages for which reported thought has been described as a function of extended reported speech, the absence of an explicit reported addressee appears to be a common prompt for a thought interpretation (also cf. Spronck, 2015, 1–2). Particularly in languages in which the verb used in the Matrix unit does not indicate a strict lexical distinction between 'say', 'think' and, e.g., a generic action, as is the case in several Australian (Rumsey, 1990; McGregor, 2014) and South American languages (van der Voort, 2002), reported speech and reported thought are often virtually indistinguishable. This is also the case for many of the examples in the African languages Güldemann

(2008) describes under the label of 'quotative indexes'. These are Matrix units, often consisting of a single morpheme, that typically (diachronically) derive from a lexeme meaning 'say', but that, synchronically, have a much broader meaning[5].

As Hsieh (2012) suggests, the interpretation process involved in extended reported speech with a THINK function is often one based on inference, but there is a crucial difference with SAE examples like 'to say to oneself': the THINK examples often do not strictly codify the distinction between reported speech and reported thought: the example in (6) strictly expresses neither reported speech or thought; it can equally express both. No language has been reported to have a dedicated reported thought construction extending to a speech meaning. That is, for all languages, the speech interpretation appears to be the most common and versatile. However, the lexico-grammatical structure does not unambiguously specify this. Hence, the inferential interpretation *narrows* its meaning to THINK, rather than metaphorically extends it from a specific speech interpretation to a thought interpretation. The absence of a (clear) second referent indicating a person spoken to in such cases, suggests an undirected monologue, which leads to the interpretation that the subject referent of the Matrix unit is thinking the content of the Report unit, rather than saying it.

THINK-type extended reported speech appears common in the literature and some authors even assume that it underlies other subsequent meaning extensions illustrated below. For example, Reesink (1993) suggests that all extended reported speech could be seen as a form of 'inner speech', a term first coined by Vygotsky (1987), to reflect the idea that verbalised (but non-spoken) thought is like speaking in one's mind. While this connection highlights the universal human cognitive principles behind the phenomenon, the metaphorical extension from SAY to THINK in, e.g., Standard Average European languages should not be confused with extended reported speech as intended here. For languages that do display the phenomenon as defined in (4) and exemplified in this section, THINK could be seen as the first stage crossing the Rubicon from 'regular' reported speech to extended reported speech.

A type slightly further removed from this stage is formed by the 'intention' interpretation of extended reported speech, which often can be translated with a lexeme meaning WANT. An example of this type is shown in (7) and in (5a) above.

(7) WARRWA (Nyulnyulan; Australia)

miliya ka-nda *nga-n-di-ny-jarri* *marlu nga-la-nda-na*
today 1MIN.NOM.FUT-go 1MIN.NOM-CM-say-IMP-SEQ not 1MIN.NOM-IRR-go-PST

'I wanted to go today, but didn't' (McGregor, 2014, 310)

ME: "I will go today," I said, {but I did not go}

[5]Both the observation that the main element in the Matrix unit can have a broader lexical meaning than 'say' and does not even need to be a fully inflecting verb, as is the case for many of Güldemann's 'quotative indexes' motivate the inclusive formulation in our comparative concept in 1 that 'M minimally consists of or contains an element that can be translated as 'say' '.

The way in which the Warrwa and Aguaruna strategies in (5a) and (7) are interpreted may rely on a similar inferential process as described for THINK: both examples are semantically underspecified. In the absence of an explicit reported addressee, like with THINK, (5a) and (7) suggest a monologic or internal process. Furthermore, in both instances the Report unit describes a future event with a first person subject, which seems appropriate for an intentional interpretation. Note again, however, that as with all examples of extended reported speech, the meanings of (5a) and (7) are those of the idiomatic glosses: even though we may be able to understand some of the compositional elements that give rise to the 'want' interpretation, these constructions are either the only, or a common way to express WANT complement constructions in the respective languages. (For similar observations about the grammatical status of 'intentional' reported speech constructions, see Rumsey (1990), Everett (2008) and Konnerth (2020), among others.) We return to this type in more detail in **Section 3.2**.

A final type of extended reported speech in which the apparent reported speaker is engaged in a mental rather than a speech activity is a broad class of attitudinal meanings that several authors discuss. Two relevant examples occur in (8).

(8) USAN (Trans–New Guinea; Papua New Guinea)

a. *me ya-g-ab eng yag umo-nd is-orei qamb ye-toat is-umirei*
not me-see-SS the river down-at:the descend-3s:FP say:SS me-follow:SS descend-3s:FP
'...and when they didn't see me they followed me thinking I had gone down to the river' (Reesink, 1993, 222)
ME: {they did not see me down at the river and descended,} they said they followed me

b. *Mi qei-qei mani umer-iner qamb gitab ig-oun*
thing some-RED yam wilt-3s:UF say:SS abstain:SS be-1p:PRS
'We abstain from various things lest the yam wilts' (Reesink, 1993, 222, gloss updated)
ME: '"the yam might wilt"' we say, {and abstain}

Example (8a) could be interpreted as an example of reported thought, but Reesink (1993) suggests that while the sentence attributes the thought that the current speaker had descended towards the river to the subject of *qamb* 'they say' in (8a), the example primarily conveys that this thought was mistaken, not that it was held (or uttered). In this survey, we will not explore this type beyond these observations, but attitudinal meanings are more commonly described in the literature on extended reported speech and the irrealis interpretation reported for Sinitic (Chappell, 2012) may be related to this as well.

The attitudinal meaning is perhaps even more explicit in (8b), since the idiomatic translation does not even include a cognitive or utterance verb. This meaning seems related to the interpretation of 'warning', which van der Voort (2002) lists for Kwaza and 'deontic modality', which Güldemann (2008) describes for his African sample.

We will return to the more general principles behind the interpretation of each of the types of extended reported speech introduced in this section, but a common element in these examples appears to be that they all cast the reported speaker in a different role: as a thinker, as someone who holds an intention as described in the Report, or as a referent with specific attitudinal qualities.

Such cognitive activity appears to be completely absent in the next subtype, constituted by aspectual/temporal examples of

extended reported speech as in (2) above. Two further examples of this type are shown in (9).

(9) a. KALMYK (Mongolic; Russia)
mörən ük-čə odə-n gi-ʒä-nä
horse die-CVB.IPFV go.away-CVB.MOD say-PROG-PRS
'The horse is close to death.' (Baranova, 2015, 71)
ME: The horse is saying: "I go away, I will die"

b. WAN (Mande; Ivory Coast)
ké wà tí gé ɓā ɓó mɔ̄
if time say LOG arrive PRT
'When the time comes...' (or 'When the time is about to arrive.') (Nikitina, 2020, 88)
ME: If the time says: "I arrive"

Both examples in (9), like (2), have non-human subject referents in the Matrix unit, so it is clear that they do not involve actual speakers, but, more importantly, the Report unit describes an inceptive event that does seem to reflect any other perspective than that of the current speaker uttering these sentences. The converb constructions in (9a) occur in regular reported speech constructions and allow for a direct speech translation (Baranova, 2015, 64), but the example is not a statement about the mental state of the horse. Similarly, in (9b) the communicative relevance of the example is not some dramatic re-enactment of visions of time; it is about the inceptive or inchoative aspectual status of the content of the Report unit[6].

Once more, it should be stressed that the inchoative interpretation in these examples is not a poetic invention by the speakers of these sentences. Rather, the examples represent a common way to express aspectual meanings in these languages. Birchall (2018) describes similar examples for languages of the Chapacuran family as expressions of incipient action or future tense. Güldemann (2008) also demonstrates the future tense meaning for other African languages and van der Voort (2002) reports it for Kwaza (see **Appendix Table A1**).

Another example in which the Report unit does clearly not signal an utterance or mental state is the NAME type, as in (10).

(10) NGARINYIN (Worrorran; Australia)
malyan-nga junba jandu jirri ngarr-w$_2$*a*$_2$*-ma-ø-nangka*
for.nothing-ONLY dance designer M.ANAPH 1PL.INCL.S-IRR-do-PRS-3SG.IO
'We don't just call him 'corroboree designer' for nothing' (Coate, 1966, 106, lines 39-40, cited by Spronck, 2015, 98, gloss updated)
ME: We do not say "corroboree designer" about him for nothing

The term *junba jandu jirri* 'the dance designer' in (10) does not refer to a specific speech act, but is a general description of the oblique referent, which can be translated into English with the lexical verb 'call' or 'name'. Among the examples of extended reported speech illustrated here, this type is slightly different in the sense that the 'name Report' is commonly assumed to be spoken, but the status of the Report does not correspond to an utterance, which qualifies this as an extended meaning. Among the literature surveyed for this section, similar examples are attested in Ainu (Bugaeva, 2008) and in African (Güldemann, 2008), Tibeto-Burman (Saxena, 1988) and Siberian languages (Matić and Pakendorf, 2013).

A final type that we would like to introduce in this section is extended reported speech used for the purpose of information structuring, specifically topic marking, as in (11).

(11) BEIJING MANDARIN (Sino-Tibetan; China)
wǒ jiāglái jiù shì shénme ya, shuō kǎo dàxué, ráguǒ kǎo-bu-shàng, wǒ jiù lián kǎo liǎng cì.
1SG future then be what PRT say$_{TOP}$ sit:exam university if exam-NEG-DIR 1SG then successive sit:exam two CLF
'So what will I do in the future?; as for taking university entrance exams, if I'm not admitted, I'll just repeat the exams.' (Chappell, 2012, 88, gloss updated)
ME: {So what will I do in the future?} Say "take university entrance exams" {if I'm not admitted, I'll just repeat the exams}

In examples like (11) the Report unit describes information that, presumably, has already been raised in the conversation and is subsequently commented on. Interestingly, information structuring examples of extended reported speech are described as signalling both that the content of the Report unit is a 'topic' and that the content is 'highlighted', which would rather suggest a focus function. Matić and Pakendorf (2013) also attest reported speech with a topic interpretation in their Siberian sample, and, more generally, discourse functions are attested in Aguaruna and African languages, as indicated in **Appendix Table A1**.

Güldemann (2008), 510 and Reesink (1993), 223 furthermore report that extended reported speech may have a 'listing' interpretation (e.g., 'say *x*, say *y*, say *z*'), which could be seen as an instance in which the Reports are presented as a series of discourse topics.

2.2.2 Grammaticalised Extended Reported Speech

The interpretations of extended reported speech described in the previous section mostly corresponded to common reported speech constructions in the respective languages. They also shared the feature that the Matrix unit often corresponded to a lexical (matrix) verb in English, that is, 'think', 'want' or 'call', although the translations were more diverse for the attitudinal, aspectual/temporal and information structuring types of extended reported speech. All authors cited specifically introduce these examples because they represent common, conventional ways to express the meanings described and, hence, they involve a degree of constructionalisation. However, in most cases, the elements involved in the constructions do not appear to have developed into grammatical formatives.

This is different for the types that we will discuss in this section, for which a more straightforward argument can be made that the constructions have conventionalised to a degree that, at least in some languages, they have fully grammaticalised. A useful starting point for classifying these types is the overview in Kuteva et al. (2019), who list no fewer than eleven morphological categories into

[6]The observation that the Matrix unit in (9b) is subordinated under the conditional/temporal adverb *ké* 'if' is potentially relevant for the interpretation of this example, but not a requirement, as demonstrated by (2). We will explore potential connections between morphosyntactic structure and interpretation in **Section 4**.

which the lexeme SAY may grammaticalise[7]. These meanings/categories are shown in (12).

(12) SAY (Kuteva et al., 2019, 375-388)
a. CAUSE
b. COMPLEMENTIZER
c. CONDITIONAL
d. DISCOURSE MARKER
e. EVIDENTIAL, QUOTATIVE
f. EVIDENTIAL, REPORTED
g. PURPOSE
h. QUOTATIVE
i. SIMILATIVE
j. SUBORDINATOR
k. TOPIC

Since the classification developed by Kuteva et al. (2019) constitutes a 'lexicon of grammaticalisation', and grammaticalisation is defined as a diachronic process in which a lexeme becomes a grammatical element, presenting the types in (12) as deriving from SAY is a useful shorthand. Note, however, that as with the examples of extended reported speech presented before, the types of extended reported speech illustrated in this section commonly include a recognisable Matrix and Report unit. What characterises these types, though, is that, more frequently than in the previous examples, these units are integrated into other morphosyntactic structures. For this reason, 'grammaticalised' extended reported speech is often slightly distinct from other reported speech in the respective languages. The examples introduced here, therefore, often carry slightly more structural cues than those presented in the previous sections as to their 'extended' interpretation.

Taking the list in (12) as a guide, we will briefly illustrate the various types below. The CAUSE function (12a), exemplified in (13), appears to be particularly common.

(13) a. KWAZA (Isolate; Brazil)
'ja kui-'nĩ-da-ki
already drink-CAUS-1SG-DECL
'I already let (him) drink (van der Voort, 2002, 320, gloss updated)
ME: I already said: 'let him drink!'
b. ORO WIN (Chapacuran; Brazil)
maw era n-on
go.SG 2SG.FUT 3.NFUT-3M
'He sent him away' (Birchall, 2018, 21)
ME: "You will leave" he (said) to him
c. NGARINYIN (Worrorran; Australia)
wurrumiy-angga bud-ma-ra-ngarrugu
w-class:ob.3pl.sub-fut.-'recognize' 3pl.sub.-ma-past-1pl.incl.obl.
'They wanted us to recognize it' (Rumsey, 1990, 349, gloss updated)
ME: "They will recognize it" they said with regard to us

As the Mock English translations in (13) illustrate, each of these examples can still be interpreted as reported speech, so in this sense the function is less clearly grammaticalised than some of the other ones discussed below. However, the examples in (13) all involve an interpretation of (indirect) causation that partially requires a structural re-analysis of the reported speech construction involved: in (13b) and (13c) the entity who is coerced into performing the act described in the 'Report' is introduced as an oblique object in the Matrix unit. This involves a change in semantic roles: in both examples the subject of the Matrix becomes the 'causer' argument and in (13b) the indirect object, i.e. the 'addressee' is interpreted as a causee, as is the oblique object, i.e. the 'object talked about', in (13c). In (13a), the causal interpretation appears to arise slightly differently because of the presence of a morpheme glossed as causative in combination with the reported speech construction. In this example, the causee is left implicit.

These three examples already show that even though extended meanings in reported speech may be similar across languages, it should not necessarily be assumed that these meanings arise through exactly the same (diachronic) pathways.

Some typical examples of the complementiser function of reported speech (12b) are shown in (14).

(14) a. SANZHI (Northeast Caucasian; Russia)
T'ahir-ri-j han b-ič-ib [du q'an Ø-iɣ-ub Ø-ik'-ul]
Tahir-OBL-DAT seem n-be.PFV-PRET 1sg late m-be.PFV-PRET m-say.IPFV-ICVB
'Tahir thought that I (=current speaker) was late.' (Forker, 2019, 185)
ME: It seemed to Tahir saying I was late
b. AINU (Ainu; Japan)
nep Ø=uhuy hum an sekor ihunar=an
something 3.S=burn NONVIS.EV be.SG QUOT search=LOG.S
'We were searching (here: smelling around) whether something was burning.' (Bugaeva, 2008, 62)
ME: We search saying something appears to be burning

Judging by the Mock English translations of (14a) and (14b), these contain a redundant verb of saying that serves the main function of connecting a main clause describing some cognitive activity with a complement clause specifying this cognitive activity. Both the complementiser (12b) and the more general subordinator use (12j) of reported speech constructions are introduced more fully in **Section 3.2.3**, but the examples in (14) already reveal two important qualities of this subtype of extended reported speech. On the one hand, it is less obvious that these examples involve a Matrix and Report, since the 'complementiser' interpretation only emerges in the context of another bi-clausal structure. Therefore, two equally plausible analyses present themselves: either the Matrix and Report units fully overlap with these two clauses (e.g., the clause between square brackets in (14a) both derives from a Report *and* is a complement clause of the preceding clause *T'ahir-ri-j han b-ič-ib* 'it seemed to Tahir') or the verb SAY grammaticalises as a complementiser without bringing its associated Matrix and Report structure. We will briefly discuss this problem in **Section 3.2.3**, but refer to each of the examples cited here as extended reported speech. Second, even though we refer to the function in (14) as a complementiser, both examples represent cognitive actions, which raises the question of to what extent the 'SAY complementiser' interpretation can be generalised beyond predicates expressing meanings closely related to speech and thought. Matić and Pakendorf (2013), in particular, do show a variation of complement types with which a SAY-derived complementiser may occur: in some languages such an element may only combine with speech or cognition complements, in others it extends further, e.g. to verbs of perception and (eventually) any

[7]This makes the lexeme SAY the most productive source for grammaticalised elements in Kuteva et al.'s lexicon, with only the entry 'locative' listing more functions.

complement/subordinate clause. The examples of reported speech 'conjunctions', listed in **Appendix Table A1**, may fall on various parts of this spectrum, and we will discuss these varying degrees of grammaticalisation in **Section 3.2.3** as well.

Another clause linking function, that seems related to the attitudinal senses illustrated in (8), is the conditional function (12c), as in example (15).

(15) GA (Niger–Congo; Ghana)

máha o nɪ́yenû ké oba.
give:1:SG:FUT you food (say) you:come

'I'll give you some food if/when you come.' (Heine and Kuteva, 2002, 265, cited from Lord)

ME: I will give you food say you come

A generalisation that could be made over this subtype is that in (8) and (15) the 'Report unit' indicates a hypothetical or otherwise qualified event or action. As we will discuss in **Section 4**, this meaning can be derived quite simply from the full meaning of a reported speech construction, which, as we will argue, is also the case for the following three functions on the list: discourse markers (12d), as Chappell (2012) illustrates (and which might also include the 'listing function' referred to above) and the evidentials 'quotative' (12e) and reported (12f). The distinction between these evidential categories is variously defined in the literature: Aikhenvald (2004) suggests that quotative evidentials introduce a specific source referent (i.e. the reported speaker is explicitly mentioned), whereas reported/repor(ta)tive evidentials, otherwise labelled 'hearsay' or 'reported evidence', do not. However for Boye (2012) the relevant distinction lies in the semantic status of the Report unit: a reportative embeds a proposition, while a quotative embeds a speech act (also cf. Wiemer, 2018). Kuteva et al. (2019, 381) note the close diachronic relation between the two evidential categories.

The next function Kuteva et al. (2019) list is that of purpose (12g), cf. (16).

(16) BESLENEY KABARDIAN (Northwest Caucasian; North Caucasus)

[[q̇ə-z-de-ʔapɜq̇^{w}ə-n] ç̌'ale-m ž'-je-rjə] q̇e-k̆^{w}-a
DIR-1SG.IO-COM-help-POT youth-OBL mouth-3SG.A+say-ADD came

'The boy came to help me.' (Ershova, 2012, 14)

ME: "Maybe help me (= the current speaker)" the boy saying, {he came}

The purpose interpretation appears on the one hand related to the WANT or intention interpretation as illustrated in **Section 2.2.1**, but the translation 'in order to' also reflects a more grammatical interpretation, which involves elements that may be used to introduce additional syntactic constituents. Like in the 'complementiser' examples, the Matrix unit in (16) occurs in subordination (the additive marker marks the Matrix as a converb Ershova, 2012, 76)[8].

Example (16) is notable for another reason: the striking indexical features of the embedded first person pronoun, which refers to the current speaker, and not to the subject of the matrix clause. Such indexical patterns are in part a typical genetic property of languages like Besleney Kabardian, but also hold implications for the relation between common categories of reported speech, such as direct and indirect speech in relation to extended reported speech. Unlike the impression sometimes given in the literature[9], extended reported speech is not restricted to typical direct speech structures (as can also be seen from logophoric examples as in (2) and apparently indirect constructions, such as 15 and 16). For further discussion, see **Section 4.2**.

Purpose interpretations are common among the languages listed in **Appendix Table A1**, but an interesting further extension occurs in Tibeto-Burman (Saxena, 1988); the interpretation 'to do intentionally, deliberately', i.e., *on purpose*. Cf. (17).

(17) NEWARI (Sino-Tibetan; Nepal)

rām-ɔ̃ gilās kurke dha-k-a-a kurk-lɔ
Ram-ERG glass break-NPC say-CAUS-PTCP-NF break-PD

'Ram broke the glass deliberately.' (Saxena, 1988, 381, gloss modified)

ME: {Ram broke the glass} saying making it "break"

The 'on purpose' meaning of (17) clearly constitutes a slightly separate type from the more common 'purpose' interpretation which Kuteva et al. (2019) distinguish, but like many of the other more grammaticalised examples of extended reported speech it too involves a subordinating structure, specifically a Matrix consisting of a participle predicate.

In addition to 'evidential quotative', Kuteva et al. (2019) also list a separate category of 'quotative', which refers to what Güldemann (2008) calls a 'quotative index': a Report unit that consists of a single morphological element that (often) diachronically derives from a lexical verb SAY. Although such Report units may develop extended meanings, they do not necessarily count as examples of extended reported speech under our definition in (4b).

We discuss the subordinator function (12j) together with complementation (12b) in **Section 3.2.3** and we have illustrated the information structuring subtype of 'topic' (12k) in 11 above, which leaves only one final class from Kuteva et al.'s list; that of similative (12i), as illustrated in (18).

(18) TAMIL (Dravidian; India/Sri Lanka)

kumaar puli en̲-a paay-nt-aan̲.
Kumar tiger say-INF jump-PAST-3:M:SG

'Kumar jumped like a tiger.' (Heine and Kuteva, 2002, 269, cited from Lehmann)

ME: {Kumar to say tiger he jumped}

Comparison/similarity meanings are attested rather widely in the literature (cf. Güldemann, 2008; Matić and Pakendorf, 2013) and, like attitudinal meanings, can be derived from a common semantic component of reported speech constructions, as we argue in **Section 4.1**. Note that, as in many of the examples of

[8]The Matrix verb in (16) contains the incorporated noun 'mouth', which could suggest that the intention is actually spoken, but this construction is also used for the expression of thought (cf. Ershova, 2012, 78), so it does not seem a necessary interpretation for this example.

[9]For example, Pascual (2014), 83 presents her pioneering study as a 'cross-linguistic study of *direct speech* for non-quotation' (emphasis added), despite citing examples that do not represent direct speech in the chapter and allowing for a more inclusive description of the phenomenon elsewhere.

grammaticalised extended reported speech in this section (but not the causative subtype), the Matrix predicate *puli* 'to say' in (18) appears in a non-finite form.

2.3 Interpretations of Extended Reported Speech: An Inventory

Despite the wide variety of interpretations illustrated above, what stands out in the literature is how regular the meaning extensions in reported speech appear to be across unrelated languages. None of the subtypes illustrated in the previous sections appears only once in the literature summarised in **Appendix Table A1** and the very few additional functions that are attested can be related to more regularly described ones. For example, Saxena (1988) distinguishes 'expletive' and onomatopoeic functions in Tibeto-Burman, which indeed do not constitute typical Reported units, but may be categorised as a form of speech and/or sound emission.

One possible further subtype is mentioned by multiple sources but not included in Kuteva et al.'s (2019) list of grammaticalised functions. This is the category of 'auxiliary' and/or 'light verb', which Güldemann (2008) and Matić and Pakendorf (2013) find in their African and Siberian samples, respectively. This type reflects the observation that the verb SAY (or, more accurately, a predicate diachronically related to the meaning SAY) can bleach semantically over the course of grammaticalisation to the extent that it no longer has any distinguishable lexical meaning. As such, it often combines with types illustrated above, like the aspectual interpretations in (9) or the causative ones in (13). In such examples, the (historical) verb SAY does not contribute any lexical meaning to the construction, but merely connects elements in the sentence, or hosts cross-referential or temporal affixes, like a light verb (cf. Matić and Pakendorf, 2013, 385).

With respect to our present analysis, two aspects of this observation are relevant: on the one hand, first, it constitutes a rather different level of generalisation to the one adopted for most of the examples introduced above, that is, it focuses on the predicate SAY, rather than a full reported speech construction and, second, cross-linguistically, the development from speech verb into light verb can be seen to occur in the opposite direction in some languages. Particularly, for a number of Australian languages it has been observed that instead of having a specialised speech predicate, reported speech constructions in languages such as Ngarinyin (Rumsey, 1990) and languages of the Nyulnyulan family (McGregor, 2014) contain a generic action verb, often glossed as 'do' (cf. example 10). In the grammatical context of a reported speech construction this predicate assumes the lexical meaning 'say'.

While assuming that the interpretations illustrated in the preceding sections arise out of grammaticalised (or re-lexicalised) uses of the lexeme SAY is a possible analysis for some languages, it is less appropriate for others. It is also variably applicable to the subtypes of extended reported speech so far introduced. For example, the complementising/linking function may be inviting focus on the word unit of SAY itself, but it equally involves a link between two clauses, not unlike the Matrix and Report units already involved in a reported speech construction. If our analysis of meaning extension starts from a lexeme SAY, it is problematic to argue that the verbs used in (extended) reported speech may either entirely lose their speech interpretation, or that non-speech verbs can be recruited as matrix verbs in reported speech. This is not the case if we take reported speech constructions, i.e. Matrix and Report units with or without a lexical speech verb as the (diachronic) source for the extensions reported here.

This analysis also provides a consistent solution for the possible problem van der Voort (2002) diagnoses, that meaning extensions of the type illustrated in the preceding sections occur *regardless* of the lexico-grammatical status of the Matrix. Even affixes or particles like quotatives, or highly abstract constructions like the reported speech construction formed by the declarative marker in Kwaza (13a), may give rise to such interpretations as 'want' or 'cause to do'. This creates the theoretical problem that under the SAY grammaticalisation analysis we would have a lexical meaning emerging from a grammatical construction (i.e., degrammaticalisation)[10]. Furthermore, simply focusing on the lexeme SAY removes from sight the similarities with meaning extensions arising from other types of Matrix units.

Before exploring the consequences of this integrated approach to extended reported speech further, let us take stock. The observations in **Section 2** expand the initial inventory of extended functions of reported speech based on Pascual (2014) in (3) to the set of functions in (19). Although the distinction between lexical and 'grammatical' functions is not clear-cut, we may further divide these functions into a more lexical group summarised in (19a) and a group that bears a resemblance with morphosyntactic categories, or functional elements in the sentence, listed in (19b).

(19) a. *lexical extensions*
NAMING: call;
THOUGHT: think, want.
b. *morphosyntactic extensions*
ATTITUDE: modality, irrealis;
EVIDENTIALITY: quotative, reportative;
TIME: inchoative aspect, future tense;
VALENCY CHANGING: causation, reason, purpose;
CLAUSE LINKING: complementiser, connective;
INFORMATION STRUCTURE: topic, highlighting.

Before placing the functions in (19) in a broader context in **Section 4**, we will first try to delve slightly deeper into the distribution and origin of some of these functions, by presenting a typological study of two specific subtypes of extended reported speech in **Section 3**. As we will show, there are many difficulties inherent in studying extended reported speech as a typological topic, but in order to contextualise the

[10]Depending on an author's theoretical stance, this situation may or may not jeopardise their account, but in any case it complicates it if one has to unify the observation that a similar meaning extension arises from two different sources (a lexical and a non-lexical one), which is an additional step not required for the analysis that the Matrix unit is the relevant source element.

observations above it will be useful to gain an impression of how widespread the phenomenon is in the languages of the world. In order to develop an understanding of how extended meanings arise out of the structural features of reported speech constructions, we will also present brief case studies of two such meanings, that is, the WANT and complementiser/linker subtypes, which can be identified relatively reliably in descriptive grammars.

3 A SAMPLE STUDY

3.1 Methodology and Distributions

In this section we present the first results of a broad typological study on extended reported speech based on a cross-linguistic, genetically balanced sample of 100 languages. We study the distribution of the phenomenon, aiming to show that it is not restricted to certain areas or language groups but can be found around the world (**Section 3.2.1**) and present case studies of extended reported speech with a WANT interpretation (discussed in **Section 3.2.2**) and with a complementising/clause linking function (see **Section 3.2.3**). The purpose of these case studies is to examine structural similarities between examples of extended reported speech with comparable interpretations in unrelated languages, which should lend insight into how these interpretations arise. The two subtypes chosen are particularly useful for such an exploratory analysis, since we will be able to draw on some clear hypotheses for such structural features based on previous literature, which we will be able to test on the basis of our sample.

Before presenting these results, however, we introduce our sample and sampling procedure in **Section 3.1.1** and briefly reflect on our methodology and its possibilities and limitations in **Section 3.1.2**.

3.1.1 Sample

Linguistic typology is a branch of linguistics that seeks to classify and understand the range of variation found in the world's estimated 7,000 languages. It does so by conducting sample studies of features that are explicitly pre-defined on the basis of semantic and/or abstract formal properties (Haspelmath, 2010), mostly using descriptive grammars, i.e., maximally comprehensive descriptions of individual languages organised in a way that allows for cross-linguistic comparison.

The selection of languages in a typological sample, Rijkhoff et al. (1993) suggest, qualifies these samples as one of two kinds: probability and variety samples. Probability samples are intended as a maximally representative selection of the world's languages, aimed at answering statistical questions about the frequency with which a feature occurs. To this end, larger language families are better represented in probability samples than smaller language families and the primary focus is on diffused categories[11]. Variety samples, on the other hand, aim to capture a maximum amount of genealogically and topographically distinct languages. To this end, larger language families are not prioritised over smaller ones in the sample, which means that typologically 'rare' languages are included in the same ratio as more familiar ones. A variety sample allows us to address the qualitative question whether a linguistic feature is restricted to a particular area or language group and within what range the observed values fall.

For our purposes of demonstrating that extended reported speech (as defined in 1) occurs globally and to understand the variability of the phenomenon, our case study involves a variety sample, constructed following the method proposed by Miestamo et al. (2016). This method is based on the distribution of languages across six macro-areas and according to a classification in genera, defined by Dryer (1989) as a set of closely related languages with a common time-depth of no more than 3,500 to 4,000 years. Such a classification is inherently subject to ongoing academic debate, with occasional reclassification of individual genera as new diachronic evidence emerges, but for our sample we follow the list of genera distinguished in Dryer and Haspelmath (2013). The notion of genus also allows us to take into account the diachronic influence of language contact in areas where genetically diverse languages have long been in close proximity, which could indicate patterns of borrowing.

In constructing our sample, we have randomly selected 100 genera, following the areal distributions proposed by Miestamo et al. (2016), but have favoured languages with larger descriptive grammars over languages with fewer available resources in order to maximise the chance of finding relevant descriptions of extended reported speech. The full sample of languages, including the respective genera and sources used is described in **Appendix Table B1**.

3.1.2 Methodological Limitations: What This Study *Can* and *Cannot* Tell Us

A typological study as attempted in this section faces the obvious challenge that negative evidence does not demonstrate non-existence and positive evidence is not necessarily exhaustive. Put differently, if a descriptive grammar does not present examples of extended reported speech in accordance with our definition this cannot be taken as evidence that the phenomenon is absent in the respective language and if a descriptive grammar does include examples of extended reported speech, these do necessarily illustrate the full range of functions that the phenomenon can have. Unlike the specialised studies surveyed in **Section 2**, the descriptive grammars examined here do not aim to provide a full and detailed account of extended reported speech and may be based on corpora that lack the phenomenon, even though it exists in the language concerned. For each of the languages in our sample, we fully rely on the judgements by the author of the grammar, who, no matter how thorough and comprehensive the description, inevitably presents a 'doculect' (Cysouw and Good, 2013), a language-as-described based on a limited amount of contexts of use and selected, glossed and analysed by an author. Therefore, distributions may under-represent occurrences of extended reported speech if the corpora on which a description is based did not include them, even though extended reported speech does occur in the language. On the other hand, accounts of extended reported speech may be relatively over-represented in languages that belong to an area in which extended reported speech posited is as an areal feature (e.g., Cohen et al., 2002) so that it is on the radar of the respective grammar writer.

[11]An increased awareness of the importance of language contact and Sprachbund phenomena in the spread of linguistic features casts doubt on the assumption that genealogical affiliation can be taken as a primary selection criterion in probability samples, but this issue should not concern us here.

Despite these limitations, using the definition of extended reported speech in (1) we should be able to identify relevant examples in the sample. We should not expect the phenomenon to be limited to any specific area and to only involve a specific number of meanings. We would also not expect the phenomenon to be limited to certain structural types of reported speech, or involve any particular grammatical features. However any patterns we do find will lend further insight into the nature of extended reported speech.

In this section we only explore a few such patterns with respect to two subtypes of extended reported speech, but for a fuller analysis of the sample see Casartelli (fc). We begin with a more general question: where can examples of the phenomenon be found?

3.2 Results

3.2.1 Distribution

The map in **Figure 1**, based on the specific studies surveyed in **Section 2**, suggested that extended reported speech is not an isolated phenomenon only attested in some parts of the world, but occurs independent of language families or contact areas. The 100-language sample affirms this impression, indicating that we find relevant examples on all major continents.

Figure 3 shows the distribution of such examples: for the locations of languages indicated in orange we find evidence for the occurrence of extended reported speech in accordance with our definition in (1), for the ones indicated in blue the respective descriptive grammars do not include such examples[12]. As discussed, these observations cannot be taken as definite proof that extended reported speech is absent from the respective language, just that in the most comprehensive description of this language it has not been raised as an example or theme.

Figure 3 does not specify the types of meaning extensions found in the sample. For a fuller analysis the reader is referred to Casartelli (fc). However, the distribution confirms the wide spread of extended reported speech across areas and language families, with about half of the languages in the sample displaying the phenomenon (see **Appendix Table B1** for a list of included languages).

The discontinuities in the distributions in **Figure 3** are somewhat more instructive than the continuous groups of blue or orange dots, since our main goal is to demonstrate the occurrence of extended reported speech independent from geographical regions. Nevertheless, two areas slightly stand out: the sample does not include instances of extended reported speech in the languages of Western Europe, whereas in South-East Asia sources quite commonly describe it. Although such patterns should be interpreted with care given the considerations discussed above, they highlight the distinction between our more restricted notion of extended reported speech, as opposed to the common phenomenon of the creative, metaphorical use of conversation to express non-speech meanings in fictive interaction (Pascual, 2014). While the latter forms of use are common in (spoken) Standard Average European languages, extended reported speech is not[13]. This is particularly clear in the case of Catalan, which figures prominently in the literature on fictive interaction with examples such as (3).

(20) CATALAN (Indo-European; Spain/Andorra/France/Italy)

Ara digue'm tu a qui se li acudiria fer una cosa així.
now tell-me you to who it occur-would to do something like that

'You tell me who would do something like that' (Pascual, 2002, 181, gloss and translation as in the source)

Example (3) counts as fictive interaction since the addressee of this utterance is not actually expected to tell anything about the person 'who would do something like that', but it is not an example of extended reported speech within the definition provided in (1). This is not to say that such examples definitely do not exist in Catalan or any of the other SAE languages in our sample[14]: as indicated in **Section 3.1.2**, it simply means that using the selection criteria we have set for our study we have not identified such examples in the descriptive grammars.

Although the more general cognitive principles that Pascual (2014) describes are likely to be relevant for both synchronically metaphorical uses of fictive interaction and lexicalised and grammaticalised forms of extended reported speech, our approach visualises the latter phenomenon and shows that it can be demonstrated to occur relatively frequently around the world.

3.2.2 WANT

In this section and **Section 3.2.3** we will illustrate two different subtypes of extended reported speech in our sample: examples with an intention/WANT interpretation and those with a complementiser interpretation. Our aim with these case studies is to examine an aspect of the phenomenon that has so far received little attention, but that has important implications for our understanding of extended reported speech in relation to perspective expressions more widely and other types of reported speech in particular. This concerns the (diachronic) structural means through which the relevant meaning extensions arise.

Our reason for focusing on these two subtypes, the 'lexicalised' interpretation WANT and the 'grammaticalised' complementiser subtype, is that for these two classes of examples the literature presents sufficient evidence to form hypotheses about cross-

[12]Like the map in **Figures 1**, **3**, was produced using the R-package lingtypology (Moroz, 2017).

[13]It is likely that European sign languages show more evidence of extended reported speech, given other observations about grammaticalised forms of fictive interaction found by, e.g., Jarque and Pascual (2015) and Jarque (2016). Unfortunately, our sample only includes oral languages but the increasing availability of descriptive grammars will hopefully allow us to discuss examples from sign languages in future work.

[14]And this English sentence is, in fact, an indication that fictive interaction is a much broader phenomenon than extended reported speech.

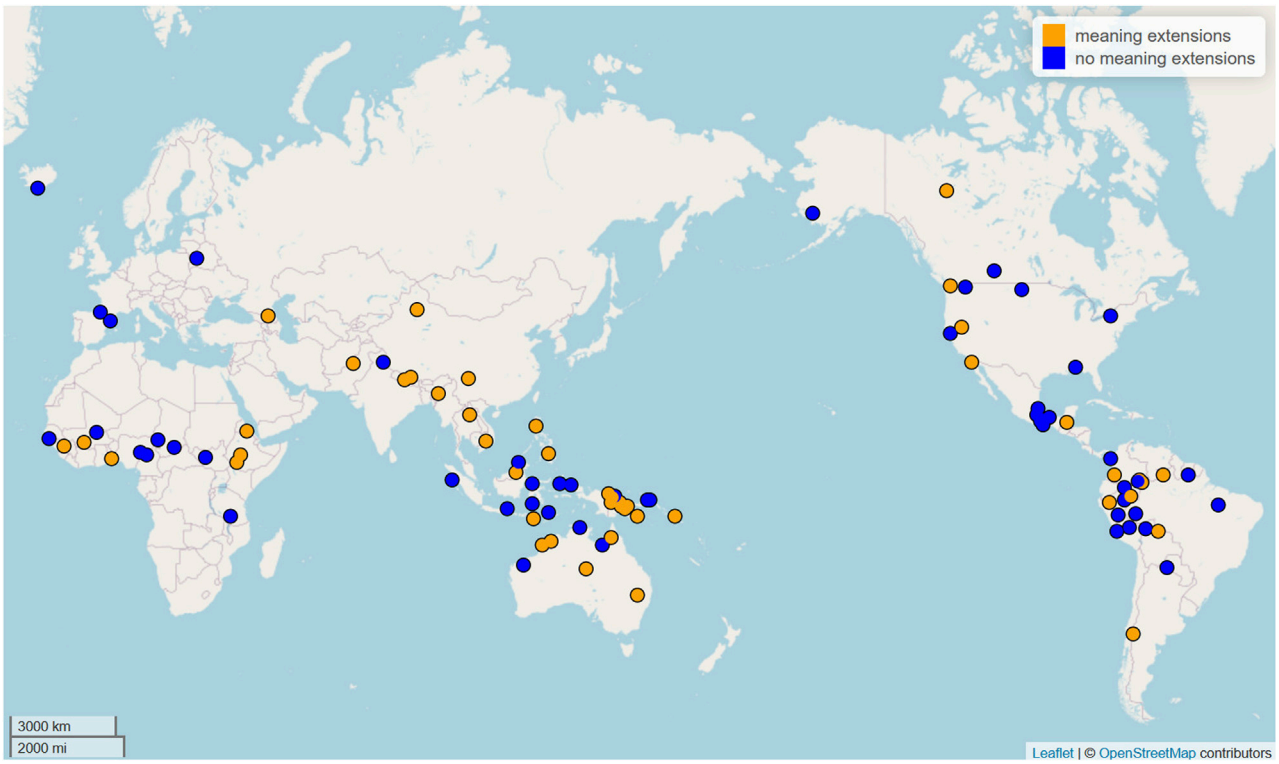


FIGURE 3 | Extended meanings of reported speech in a 100-language sample.

linguistic regularities in their structural composition[15]. The WANT interpretation of extended reported speech has variously been described as an 'intentional' (cf. Everett, 2008; Konnerth, 2020) or 'desiderative' (cf. McGregor, 2007) construction but its cross-linguistic structural realisation appears to be rather consistent: as first described by Rumsey (1982) for Ngarinyin, it often includes an embedded first person and a non-present/non-actual tense in the Report. The schematic representation in (21), adapted from Spronck (2015, 100), illustrates these features.

(21) [[... S_{1sg} + FUT]$_{Report}$ SAY]$_{Matrix}$

Throughout this section we introduce various schematic representations of extended reported speech as in (21). Here and below, the order of the Matrix and Report elements is non-iconic: the representation in (21) may reflect a structure in which the Matrix either follows or precedes the Report. The order of the morphemes and lexeme SAY is variable as well. What is relevant, in this instance, are the person and number features of the subject and the future tense in the Report. Examples closely resembling the representation in (21) indeed occur relatively frequently in the sample in extended reported speech with a WANT interpretation, as illustrated in (22).

(22) a. DARAI (Indo-European, Eurasia)[16]
her nə məi punu səti dze-m kəhə-lə tə pəran nihidzə dza-inei
look-IMP PART I also immolation go-HOR say-PST PART life NEG go-MIR
'Look, I want to immolate but I did not die!' (Dhakal, 2012, 79)
ME: {Look} "Let I also go immolation" I said {but life did not go}

b. KALAM (Trans-New-Guinea, Papunesia)
"Laplap d-in!" ag nŋ-i, ktg ow-a-k
clothes take-1SG.HORT say think-SS.PRIOR leave come-3SG-PAST
'She wanted to take (steal) clothes (but) left (them) behind.'(Pawley and Bulmer, 2011, 64)
ME: '"Let me take clothes!" say, having thought, {leaving she came}.'

c. GUMER (Afro-Asiatic; Africa)
asa ə-t'əβt' bar-əc-im mərəβ t'əβət'-əc-im
fish 1sS-take.IPFV say.PFV-3sfS-CV.M net take.PFV-3sfS-M
'Wanting to catch fish she took a net.' (Völlmin, 2017, 171)
ME: {She took a net} saying: "I catch fish"

d. WORRORRA (Worrorran; Australia)
nga-nya=ya kuN-nga=yi-eerri
1-OPT=go VCOMP-1=do-PROG
'I want to go.' (Clendon, 2014, 351)
ME: "Let me go' that I am saying/doing[17]

In addition to singular first person subjects in the Report, all examples in (22) are combined with a non-present tense or non-realis mood. Future tense occurs in several examples

[15]See Casartelli (fc) for more detailed analyses and accounts of other subtypes of extended reported speech.

[16]For the remaining examples in this section we list the macro-area in the sample, rather than countries in which the respective language is spoken.

[17]Like in other Worrorran languages (Rumsey, 1990) and Nyulnyulan languages (McGregor, 2014), the Matrix predicate yi-in Worrorra can both be translated as 'say', 'think' or 'do'. Clendon (2014) opts for the gloss 'do', but the description in the grammar demonstrates that 'say' is one of the available translations, qualifying this example as extended reported speech in accordance with our definition in (4b).

below (cf. 24b and 24c), but in these examples we find hortative or optative mood (22a, 22b, 22d), or imperfective aspect (22c). On the basis of these observations we may conclude that the future tense in the Report is slightly too specific: although it occurs in the sample, the common feature between all tenses and moods in the Reports of extended reported speech illustrated so far appears to be that they place the event described in the Report in some time other than the here-and-now. We will label this observation IRRealis, as in (23).

(23) [[... S_{1sg} + IRR]$_{Report}$ SAY]$_{Matrix}$

In addition to first person singular, we also find other person and number values in the Reports of WANT extended reported speech, such as non-singular forms. In the Yeri example in (24a), both the subject of the Matrix and Report are first person plural. In (24b) the Matrix subject is coreferential with a first person dual in the Report. In contrast, (24a) has a third person subject in the Report and also in the Matrix. In accordance with (23), the tense/mood values in the Report units in (24) are all non-present/non-realis.

(24) a. YERI (Torricelli; Papunesia)
Ø-nobia h-ero.
1pl-talk.REALIS 1pl-go.IRR
'We wanted to go.' (Wilson, 2017, 380, gloss updated)
ME: We talked "we might go"

b. MA MANDA (Trans-New-Guinea; Papunesia)
sisa gaamiyong=lit laai ku-ntaa-mot taa-ka ku-gû-mot
±2days PN=COM PN go-FUT-1DU say-SS go-RP-1DU
'The day before yesterday I wanted to go to Lae with Gamiyong, so we went.' (Pennington, 2018, 183)
ME: The day before yesterday I said "We will go to Lae with Gamiyong", {so we went}

c. OKSAPMIN (Oksapmin; Papunesia)
niu jia max bɔten x-t-pel=o li-m xe-ja
new(English) year(English) RECG pray(Tok Pisin) DO-PFV-IF.PL=QUOT say-SEQ
be-PRS.PL
'They wanted to pray for, you know, that New Year.' (Loughnane, 2009, 124)
ME: They were saying like they will pray for New Year{, you know}

These examples indicate that rather than taking the specific person and number values first person singular as a typical feature of WANT extended reported speech, a better generalisation is to highlight what it signals: a first person subject in the Report necessarily indicates co-referentiality with the subject of the Matrix. In addition to first person singular marking in the Report, co-referentiality may also be indicated by having the same person/number values in both the Report and Matrix units, viz. first person plural in (24a) and in (24b) (also combined with same subject marking in the Matrix) and co-referential third person plural marking in (24c).

In accordance with these observations, we may update the schematic representation of WANT extended reported speech as in (25), in which the coreferential relations between the subject S in the Report and in the Matrix are indicated by the subscript index i.

(25) [[... S_i + IRR]$_{Report}$ S_i SAY]$_{Matrix}$

The remaining examples of WANT interpretations in the sample show minor variations on the pattern illustrated above. Kambera in example (26a) has a first person subject in the report, but no apparent tense/mood marking on the auxiliary verb 'try' (but note the ME based on the author's alternative translation with 'let's'). A similar observation can be made for the Paiute example in (26b), which has a generic tense (TNS) form. This form is due, however, to a morphosyntactic restriction in the language, which disallows the combination of any other tense forms with applicative marking (Thornes, 2003, 398).

(26) a. KAMBERA (Austronesian, Papunesia)
"Ka ta-kama wà-nya$_j$ [i Ama]$_j$" wà-mu nú
CNJ lpl.NOM-try use-3sg.DAT ART father say-2sg.GEN DEI
'You want(ed) to try out father' (Klamer, 1998, 294, gloss updated)
ME: "Let's try out father" you said'[18]

b. PAIUTE (Uto-Aztecan; North America)
oo=kaina mi=tiya?i-pi miu ta na-ni-naka-ki-ti
so=MOD PL=die-PFV QUOT 1.DU MM-IP/speech-hear-APPL-TNS
'Perhaps those who have passed on want us to hear them.' (Thornes, 2003, 340, gloss updated)
ME: Perhaps those who have died say about us (like) they themselves speak-hear with respect to us

Even though both examples in (26) could be seen as slight variations of the representation in (25), it appears to capture most of the examples of the WANT subtype of extended reported speech in the sample, and the previous literature (again, note that the word order in (25) is non-iconic).

This leads us to three preliminary conclusions: first, the relative similarity of WANT extended reported speech across unrelated languages and areas is unlikely to be coincidental. This suggests a more fundamental common factor underlying these examples. Second, the similarities between the occurrences of extended reported speech are not only semantic, the examples in this section also appear to share a structural basis. This observation is not new, for example Reesink (1993), 223 notes that all examples of extended reported speech in Usan involve a same subject marker, indicating co-referentiality between the subject of the Report and that of the Matrix clause, whereas 'regular' reported speech in the language does not require this. While we would not predict that all extended reported speech across languages can be qualified in terms of a restricted set of formal features, the relative frequency and correspondence of structures involved in extended reported speech deserves more attention than it has received in the literature so far. Third, and perhaps most significantly, the features as represented in (25) cross-cut common subtypes of reported speech, such as the binary opposition between direct and indirect speech. This has implications for our understanding of the boundaries between perspective constructions and non-perspectival constructions, as we argue in **Section 4**.

3.2.3 Complementiser/Clause Linking

Examples of extended reported speech displaying a complementiser/clause linking function are slightly less numerous in our cross-linguistic sample, but nonetheless occur five times across five language families and two linguistic macroareas[19].

[18]This alternative translation is provided in the original source.

[19]Again, note that no conclusions can be drawn about the absolute or relative occurrence of this subtype of extended reported speech on the basis of these frequencies, since the sources in the sample do not necessarily provide a fully comprehensive overview of the phenomenon in the respective language.

Typical examples of this strategy are shown in (27), where in Gumer the construction 'consists of a quoted sentence concluded by a converbal form of *bar* ['say'] followed by the matrix verb' (Völlmin, 2017, 168) and in Stieng, spoken in Cambodia and Vietnam, the clause linking function is expressed with a conjunctive form of the speech verb.

(27) a. GUMER (Afro-Asiatic; Africa)

cənə-x^{w}-im ba-x^{w}-im od-x^{w}-in-im.
come.PFV-1sgS-'M' say.PFV-1sgS-CVB.'M' tell.PFV-1sgS-3sgmO-'M'

'I told him that I came.' (Völlmin, 2017, 168, gloss updated)

ME: I told him saying "I came"

b. STIENG (Austro-Asiatic; Eurasia)

ɟun praŋ paŋ gɔt [laːh ɟəmraːh] paŋ ʔaluː ʔatɔp hej
deer running 3 know [say:CONJV cliff] 3 stop fall 1SG

'The deer runs, he knows there's a cliff, he stops and makes me fall.'[20] (Bon, 2014, 488, gloss updated and translated)

ME: The deer knows saying "a cliff" {it stops and makes me fall}

In both of the languages in (27), the complementising use of speech verbs is restricted to the semantic domains of speech and cognition (Völlmin, 2017, 168; Bon, 2014, 487): in (27a) it precedes a form of *od-* 'tell', in (27b) of *gɔt* 'know'.

A variation on the subordinated forms in (27) is shown in Darai (28), where the complementising speech verb receives a sequential marker.

(28) DARAI (Indo-European; Eurasia)

lə dzɛrʰi dul-hi-r-ke rə kancʰi dul-hi-r-ke ijə kəh-ikun kəhə-lə
PART elder wife-2SGPOSS-DAT and young wife-2POSS-DAT this say-SEQ say-PST

'Well, (the hermit) said that that was for his elder wife and another (that) for his younger wife.' (Dhakal, 2012, 144)

ME: He said saying this for your young wife and for your elder wife

Alternatively, SAY-derived 'complementisers' may also remain uninflected, as in (29).

(29) a. FONGBE (Niger-Congo; Africa)

Sìká / Kɔ̀kú ɖɔ ɖɔ yé tùn ɖɔ émì ɲin Àsíbá sín gbɔ́
Sika / Koku say say 3pl know say LOG steal Asiba GEN goat

a. 'Sika$_i$ said that they know that she$_i$ stole Asiba's goat. '

b. 'Koku$_i$ said that they know that he$_i$ stole Asiba's goat.'

c. 'Sika/Koku said that they$_i$ know that they$_i$, stole Asiba's goat.' (Lefebvre and Brousseau, 2002, 79)

ME: Koku say say they know say s/he/they stole Asiba's goat

b. KAKABE (Mande; Africa)

fóó káà kɔn kó kàrá pɛsɔtɔ máá dí
UNIV PFV.TR 3SG know say studies-ART get PFV.NEG be.easy

Everybody knows that studying is not easy. (Vydrina, 2017, 151)

ME: Everybody knows say the studies are not easy

The Fongbe example (29a) has three occurrences of *ɖɔ* 'say', the latter two of which act as a linking element between the main and complement clauses, and to which we therefore refer as a complementiser. A similar structure is attested in (29b).

The examples above share several features that we might represent schematically as in (30): all include two clausal units linked by a non-finite form of SAY. As in the representation in (25), the order of the elements in this representation varies depending on the respective language.

(30) [[clause] SAY:non-finite [clause]]

The variable 'SAY:non-finite' may either constitute a non-inflecting form or a subordinate form of SAY in the examples given here, which seem mostly representative for the type of examples commonly presented for the complementiser type of extended reported speech adduced in the literature (cf. Klamer, 2000; Heine and Kuteva, 2002; Güldemann, 2008; Matić and Pakendorf, 2013; Kuteva et al., 2019). The examples all contain a clause on each side of the SAY verb, which is consistent with Matrix and Report units. However, they also contain an additional main verb, providing lexical meaning to the sentence/respective clause. For this reason, the units represented in (30) have received the more abstract label 'clause', although they could mostly be interpreted as (originating from) Matrix and Report units as well.

We would like to address three observations about the examples of the 'complementiser' subtype illustrated in this section and in **Section 2.2.2**. A first observation that stands out, particularly given the broad grammatical label 'complementiser' that we have given to this subclass, is the very small lexical range of main verbs with which it appears to combine: the lexical matrix verbs used in the examples above are either speech verbs as well (27a, 28 and 29a), or cognition verbs: more specifically, verbs of knowing (27a and 29b). In **Section 2.2.2**, examples (14a) and (14b) also involved cognition verbs, viz. 'think' and 'search', respectively. Consequently, calling the SAY:non-finite form in (30) a 'complementiser', 'linker' or 'subordinator' is perhaps slightly deceptive: in many languages, the application of this form is limited to only a small class of complement-taking verbs, closely related to the semantic domain of speech and thought.

This impression is affirmed by the complement types Heine and Kuteva (2002), 261–265/Kuteva et al. (2019), 375–379 and Matić and Pakendorf (2013), 372–375 illustrate, which mainly involve main clauses of speech, thought and knowing, as well as perception and fear. However, the gradual dissemination of the structure represented in (30) with various types of main verbs is instructive. On the one hand, it illustrates a common process in grammaticalisation, in which the shift from a lexical to a grammatical element is not a matter of all-or-nothing, but spreads from one or a few lexical combinations and constructions to ever more lexical contexts (De Smet, 2012). It also neatly suggests a path through which structures as in (30) become established, from occurring with more speech-like Matrix/main clauses, to increasingly less speech-related ones.

This suggest that qualifying the status of SAY:non-finite in (30) in strict categorial terms, i.e. as either a lexical element or a complementiser/'linking element', may not always be possible, because this status varies between occurrences. The types of non-finite forms found in the examples above reflect this as well: dependent inflections as in (27) may signal varying degrees of conventionalisation.

Studying the behaviour of complementiser uses of SAY in Austronesian languages, Klamer (2000) presents a similar conclusion about the syntactic status of these elements and

[20]Original translation: 'Le cerf court, il sait qu'il y a une falaise, il s'arr'te et me fait tomber'.

proposes that the interpretation falls out from the defective inflection patterns of SAY:non-finite forms which (30) displays. Specifically, the 'SAY complementiser' in languages that have it, commonly shows no or non-matching person features to co-index arguments in the main clause, and this 'bleached' argument structure coerces a 'bleached' semantic interpretation. Klamer's analysis is consistent with our findings and leads us to a second observation about complementiser extended reported speech: although the details differ, the WANT and complementiser subtypes of extended reported speech both have consistent structural features, as we have schematically represented in (25) and (30) that correlate with their respective interpretation. In both instances, these involve, among other features, the use or lack of certain person referential features and/or tense and mood forms[21].

A third observation we would like to address here goes back to the complement types found in (30), which, in the sample, divide into speech complements and knowledge complements. The exact syntactic status of these complement types requires closer investigation for each individual language and would be weighed differently by various syntactic models, so we will refrain from detailed generalisations about the syntactic structures involved. However, there is widespread agreement among both formal and functional approaches to syntax that the scope properties of speech complements and knowledge complements (i.e., a clausal structure that expresses what-is-said as opposed to a clausal structure that expresses what-is-known) are distinct (Hooper and Thompson, 1973; Boye, 2012; Gentens, 2020). Specifically, the syntactic integration of knowledge complements is assumed to be 'tighter' than that of speech complements, which has direct consequences for their interpretation: the content of the former is asserted by the speaker, whereas that of the latter is not, cf. (31)[22].

(31) a. Alex said that [you knew that the batteries were flat.]$_{illocution}$
b. Alex knew that [you said that the batteries were flat.]$_{proposition}$

As (31) illustrates, both orders of complement taking predicates are possible, but in (31a) the interpretation of the unit between square brackets is an illocution, some utterance attributed to Alex, which the current speaker does not state as fact. In (32b), the complement clause marked by the square brackets is asserted by the speaker: the suggestion that the second person referent actually made the statement about the batteries is an integral part of the speaker's message. This effect cannot simply be attributed to the difference between the verbal predicates 'say' and 'know' (see Gentens, 2020). It relates to more general observations about scope relations in language in which elements to illocutionary meaning have a wider scope and are less tightly integrated in clauses than, e.g., elements that relate to epistemic meanings, which, in turn, have a wider scope than, e.g., temporal elements, cf. (32).

(32) a. Alex, frankly, probably did not read the instructions quickly.
b. ?Alex quickly, probably, did not read the instructions frankly.

As has been observed by both functionalist and formalist grammarians (cf. Dik, 1997a; Dik, 1997b; Cinque, 1999), adverbs targeting various parts of a sentence can be used to diagnose boundaries and scope relations between them. In (32a), the adverb 'quickly' (a temporal adverb) only has scope over the activity 'read the instructions', 'probably' (an epistemic adverb) over 'did not read the instructions quickly' and 'frankly' (an illocutionary adverb) the entire sentence. Re-ordering the adverbs in (32a) with the effect that, e.g., temporal and epistemic adverbs have scope over an illocutionary adverb results in an unintelligible sentence (32b).

Readers will weigh observations like those about the English sentences in (31) and (32) differently and, depending on other assumptions about the nature of language, explanations vary. However, the idea that sentence units have distinct scope properties that correlate with their meaning and can be classified into units that are more and less deeply syntactically embedded, is both pervasive and robust (Hengeveld, 1989; Boye, 2012; Cinque, 2013).

With respect to the distinction between the complements found with the complementiser subtype of extended reported speech, we suggest that they seem to either constitute illocutions or propositions, which suggests varying degrees of syntactic integration (as in 31).

3.3 Summary

In **Section 3** we reported on the first results of a sample study into extended reported speech. All observations introduced here will be discussed further in Casartelli (fc), but the initial analysis revealed several properties of extended reported speech that provide further insight into the phenomenon.

We found that both of the subtypes examined that they display considerable structural similarities within each respective type. We also identified three more general processes in the grammaticalisation and conventionalisation of these subtypes, which we would like to summarise as in (33).

(33) a. Recasting: the subject of the Matrix unit takes on a different role, e.g., from reported speaker to a reported thinker or a person with some intention;
b. Rescaling: units within extended reported speech become associated with different scope properties, e.g., the difference between speech and knowledge complements;
c. Bleaching: a more general process in grammaticalisation in which elements gain a broader semantic interpretation, which may be signalled by a reduction of structural complexity (cf. Klamer, 2000)

In the next section we will relate the three processes described in (33) to properties of extended reported speech more widely.

4 DISCUSSION: REPORTED SPEECH AND THE EVOLUTION OF GRAMMAR

In this section we place the empirical observations from the preceding sections in a broader perspective and suggest some implications for our understanding of reported speech as a linguistic structure and its relation to grammatical and lexical

[21]For similar observations regarding the Biblical Hebrew complementiser *lemor* and further analysis in the context of fictive interaction, see Sandler and Pascual (2019).

[22]We thank a reviewer for emphasising the relevance of assertion for the interpretation of extended reported speech and apply it in our notion of 'rescaling' below.

meaning. First, in **Section 4.1**, we return to the three processes summarised in (33) and examine their role in the grammaticalisation of extended reported speech. Particularly, we relate these processes to the meaning of reported speech as a source construction for all the various meanings and structures observed in the preceding sections. In **Section 4.2** we briefly contemplate the variety of structures involved in extended reported speech and compare these to standard, commonly recognised subtypes of reported speech, particularly direct and indirect speech constructions and quotative/reportative evidentiality. We suggest that the observation that extended reported speech cross-cuts such classical categorisations of reported speech indicates that there is more continuity within the domain of reported speech than is sometimes assumed. Finally, in **Section 4.3** we return to the research programme of fictive interaction and propose an interpretation of extended reported speech that not only places metaphors of communication centrally in the way in which humans think and speak about the world, but that acknowledges meta-linguistic reflection and reported speech as shaping forces in the emergence and evolution of grammatical categories. This is, admittedly, a speculative story, but for us it is also a significant motivation for the importance of understanding the nature and variation of extended reported speech.

4.1 Back to the Source (Construction): Recasting, Rescaling and Semantic Bleaching in Extended Reported Speech

The analysis that (at least some of) the meanings attested in extended reported speech fall out from a diachronic process of semantic bleaching suggests that it should be possible to relate them to meaning components in the original source construction. Spronck and Nikitina (2019) propose three such meaning components for reported speech, as summarised in (4.1):

(34) a. *'A semiotic status of 'demonstratedness'*
This component qualifies the reported message as a representation of an utterance, as demonstrated (or 'depicted') discourse;
b. *Evidentiality*
This component reflects a deictic relation between two events: the alleged original situation of discourse production (the 'source of information') and the current speech moment;
c. *Modality*
This component reflects an epistemic evaluation of the represented utterance, allowing the reported speaker to qualify the reported message, either by strengthening or weakening truth commitment' (Spronck and Nikitina, 2019, 25).

Despite the great variety of forms of reported speech in the languages of the world, the definition in (34) suggests that a reported speech construction should at least indicate three meanings: first, it should signal, as per (34a), that the Report unit is 'demonstrated' or 'depicted' rather than stated (Clark and Gerrig, 1990; Davidson, 2015; Clark, 2016). This property sets R apart from immediately surrounding clauses. As per (34b), reported speech also introduces an opposition between a perception event and the current speech event, which is the definition of an evidential meaning Jakobson (1957) coins. Third, as per (34c), reported speech explicitly or implicitly allows for (inferences about) the attitude of the current speaker towards the content of the attributed utterance (cf. 'distancing' in terms of Güldemann, 2008)[23].

If the definition in (34) is on the right track, the process of 'semantic bleaching' in extended reported speech should draw on one or more of these meaning components. That is, over the course of grammaticalisation some of these semantic features become irrelevant or develop a broader interpretation.

For each of the extended meanings illustrated in **Section 2** we may indeed hypothesise that this is the case: 'demonstratedness' may serve as a source meaning for (grammatical) functions relating to prominence (cf. discourse functions), comparativity (cf. similative) and unithood (cf. complement clause marking). Interestingly, these meanings are very close to the kinds of meaning extensions of fictive interaction which Jarque (2016), 175–181 finds in sign languages[24]. Under our approach to evidentiality this semantic component of reported speech could extend to other functions that introduce a contrast between two events, such as temporal meanings (cf. Zeman, 2019), as well as evidential extensions themselves. The modal meaning of reported speech may further account for the multitude of attitudinal meaning extensions.

Table 1 summarises the hypotheses briefly stated above. Specifically, for the meanings listed in (19) the table suggests to which meaning components (or combinations thereof) they may be related; the Evid(ential) meaning (34b), Mod(al)meaning (34c) or Dem(onstrated) status (34a). For meaning extensions for which the respective component appears to have been completely backgrounded, the label is struck out in **Table 1**. If an Evidential, Modal or Demonstrated meaning could be interpreted as having served as input for the specific meaning extension, that is it may explain part of the extended meaning but does not fully correspond to the extended meaning itself (as in the hypotheses posited in the preceding paragraph), it has been italicised and underlined. We will not discuss these possible grammaticalisation paths in further detail; our main aim in proposing them is to suggest that despite that great variety of subtypes of extended reported speech, they may be given explanations based on a limited number of variables: the semantics of reported speech constructions and a combination of three processes, viz. semantic bleaching, recasting and rescaling.

The processes of recasting and rescaling were introduced in (33) and roughly correspond to those types of extended reported speech in which the reported speaker appears to have acquired a non-locutionary role, for example that of a 'thinker' or 'intention holder', and those in which the Report is not interpreted as a reported utterance. These two processes obviously mutually imply each other, but could still be seen as distinct diachronic pathways. **Table 1** suggests the relevance of these processes for each of the subtypes of extended reported speech.

A full analysis of the structural diachronic changes and dynamic variation in the sample languages lies beyond the scope of this

[23]This property explains why elements that in other grammatical contexts do not carry any specific attitudinal properties, such as pronouns or tenses, can gain modal meanings in the context of reported speech (cf. Zemp, 2020).

[24]While Jarque (2016) discusses 'fictive questions', not extended reported speech, the grammaticalised meanings of fictive questions correspond quite closely to the ones we attribute to the 'demonstrated' status of reported speech. We thank a reviewer for pointing out this connection.

article (but see Casartelli, 2019), but **Table 1** suggests that rescaling takes several forms: the Report unit is typically an illocutionary unit, but may be reinterpreted as a clausal unit of various kinds (in the case of, e.g., 'highlighting' and 'topic'). In, for example, modal extended reported speech, R is interpreted as a proposition[25], and even a smaller scope unit for, for example, the aspectual subtype, which we have labelled 'event' in **Table 1**. In addition, the subject of the Matrix unit, typically the reported speaker, may be recast as, for example, a thinker, an intention holder, or 'aspectual viewpoint taker', often in combination with a rescaled R.

With this brief semantic discussion we have aimed to show that rather than constituting a scattered range of unrelated meanings, the functions attested in extended reported speech can be captured using a rather restricted set of variables that are directly related to the semantics of reported speech.

4.2 Extended Reported Speech and the Study of Perspective

As indicated in our case studies in **Section 3** and as suggested by several observations in **Section 2**, meaning extensions in reported speech often seem restricted to specific structural contexts. For example, Reesink (1993) notes with respect to Usan extended reported speech:

> 'It is clear that all seven [extended] "functions" exhibit only one form of the verb 'to say', the medial [Same Subject] form [. . .] I would suggest, then, that Usan has only two functions for *qamb* 'to say'. The first is the general function to refer to the act of speaking or telling. This allows all possible forms of the verb paradigm. The second function is what we could call a grammaticalized one, which allows only the medial Same Subject form *qamb*. This one covers all instances that refer to "inner speech"' (Reesink, 1993, 223).

The relative flexibility of the 'regular' reported speech construction compared to reported speech structures with extended meanings in Usan is mirrored by multiple accounts. Also, decreased variation in the choice of indexical values of pronouns may covary with the extended meanings of reported speech more generally. This is the case in the example of Sanzi reported thought in (14a), which shows conflicting referential values (in itself a more common property in Caucasian languages). In (14a), while the bound pronouns in the Report have a third person referent, the free pronouns have a first person value, yet both index the same referent, viz. the person uttering the example at the current speech moment (Forker, 2019). In Sanzi this appears to be a strategy to identify specific referents both in reported speech and extended interpretations, but in the Daghestanian language Tabasaran such referential conflicts between bound pronouns and pronominal clitics appear to be restricted to reported speech, and not allowed in (otherwise similar) forms of reported thought (Yaroshevich, 2020).

As Nikitina (2020) discovers, logophoric pronouns, which typically signal coreferentiality between a referent of the Matrix (often the subject) and the subject of the Report, are also required for extended meanings such as the inchoative interpretation in Wan (2). As we found in **Section 3.2.2**, the observation by Rumsey (1990) that the WANT interpretation of reported speech in Ngarinyin is restricted to Reports with first person subjects, a finding replicated in other Australian languages (McGregor, 2007, 2014) and elsewhere (cf. Everett, 2008, 389), also occurred in our cross-linguistic sample.

Chappell (2012), 81 explicitly proposes the following constructional frames in Sinitic which correspond to specific subtypes of extended reported speech:

(35) a. Complementizer: $Verb_1$ $Verb_{SAY}$
b. Topic marker: SAY + Clause
c. (Composite) conjunctions: Conjunction - SAY
d. Hearsay evidentiality: $Verb_{SAY}$ $Verb_2$
e. Irrealis marker: (if) SAY + Clause
f. Clause-final discourse marker: Clause + SAY

The construction frames in (35) are distinguished by word order (i.e., the position of SAY) and the specific combination of elements. An interesting example of such a combination is the conditional embedding 'if SAY' in (35e), which results in an irrealis reading.

It remains to be seen to what extent the subtypes of extended reported speech correlate with consistent, cross-linguistically recurring structural features. What these observations do suggest, however, is that in the languages surveyed in this paper, a number of structural elements, like those summarised in (36), can be recruited to signal a range of extended meanings.

(36) a. Pronouns and other indexical elements (including tense and mood)
b. Integrity of Matrix and Report units
c. Order of Matrix and Report units
d. Modification
i. of Matrix
ii. of Report

These strategies are by no means a complete list of possible structural prompts for meaning extensions (e.g., prosodic distinctions are likely to occur more widely as well; also cf. Spronck, 2016), but they hold an important implication: each of the properties in (36) is associated with other aspects of the classification of reported speech constructions. For example, the indexical properties of reported speech are commonly associated with the opposition between direct speech and indirect speech (as in 37 and 38, respectively). The integration of the Matrix and Report units corresponds to a distinction between having two syntactically separate (or loosely connected) clauses as in direct speech, two more integrated clauses, as in the complementation structure of (English) indirect speech and, e.g., even further structurally integrated expressions of Matrix units, as in adverbial (or morphological) expressions of reportative evidentiality (as in 39). Finally, we have also observed that over the course of grammaticalisation, Matrix clauses may become less clearly marked, a distinction commonly associated with the opposition between types of reported speech with a

[25]Following Boye (2012), 204 we also classify the difference between quotative and reportative evidentiality in terms of the type of embedded unit: a locution vs a proposition, respectively.

TABLE 1 | Suggested processes of semantic bleaching, rescaling and recasting in extended reported speech. The table lists for each of the subtypes of extended reported speech, which of the three semantic components of reported speech, viz. evid(entiality), mod(ality) and dem(onstratedness), are bleached, indicated by being ~~struck out~~ or extended, in which case the relevant semantic component is *underlined and italicised*. For rescaling 'R >' indicates the type of semantic unit into which the Report is reanalysed (the precise labels 'name', 'event' etc. are indicative and should be more narrowly defined in future research). The roles indicated after 'recasting' suggest the semantic interpretation of the referent who is marked as the reported speaker in the extended reported speech construction.

MEANING CLASS	SUBTYPE	SUGGESTED PROCESSES
naming	'Call'	Bleaching: Evid/Mod/Dem; rescaling: R > name (noun phrase); recasting: M_S > generic 'caller'
thought	'Think'	Bleaching: Evid/Mod/Dem; rescaling: R > thought (locution/inner utterance); recasting: M_S > 'thinker'
thought	'Want'	Bleaching: Evid/Mod/Dem; rescaling: R > wish/intention (proposition); recasting: M_S > 'intention holder'
attitude	'Modality'	Bleaching: ~~Evid~~/Mod/~~Dem~~; rescaling: R > proposition; recasting: M_S > *various*
attitude	'irrealis'	Bleaching:~~Evid~~/Mod/~~Dem~~; rescaling: R > proposition; recasting: *various*
evidentiality	'Quotative'	Bleaching: Evid/~~Mod~~/~~Dem~~; rescaling: R > -; recasting:
evidentiality	'Reportative'	Bleaching: Evid/~~Mod~~/~~Dem~~; rescaling: R > proposition; recasting:
time	'Inchoative aspect'	Bleaching: *Evid*/~~Mod~~/~~Dem~~; rescaling: R > proposition; recasting: 'Temporal viewpoint taker'
time	'future tense'	Bleaching: *Evid*/~~Mod~~/~~Dem~~; rescaling: R > event; recasting: 'Aspectual viewpoint taker'
valency changing	'Causation'	Bleaching: *Evid*/~~Mod~~/~~Dem~~; rescaling: R > event; recasting: Referent (causer)
valency changing	'Reason'	Bleaching: *Evid*/*Mod*/~~Dem~~; rescaling: R > event; recasting: Referent (causer)
valency changing	'Purpose'	Bleaching: *Evid*/*Mod*/~~Dem~~/; rescaling: R > name; recasting: Referent (intention holder)
clause linking	'Complementiser'	Bleaching:/~~Evid~~/~~Mod~~/*Dem*; rescaling: R > clause; recasting:
clause linking	'Connective'	Bleaching:/~~Evid~~/~~Mod~~/*Dem*; rescaling: R > clause; recasting:
information structure	'Topic'	Bleaching:/~~Evid~~/~~Mod~~/*Dem*; rescaling: R > (sub)clause; recasting:
information structure	'Highlighting'	Bleaching:/~~Evid~~/~~Mod~~/*Dem*; rescaling: R > (sub)clause; recasting:

clearly indicated source and types in which this in not the case, as in free indirect speech (as in 40), where only the Report unit is explicitly expressed.

(37) Sasha said: "I will go."

(38) They said that they would attend.

(39) Allegedly, you were there.

(40) He looked at her. Would she really leave? Her gaze froze.

Extended reported speech intersects the four types of reported speech illustrated in (37–40), but also defies this classification, with some examples not clearly belonging to any of these four classes. For the study of perspective this has the implication that in extended reported speech we see non-perspective expressions emerge, both semantically (Gentens et al., 2019, 159) and structurally, out of perspectival constructions. Reported speech typically signals that the content of the Report is grounded in a perspective other than that of the current speaker at the speech moment. For most examples of extended reported speech the perspective associated with the Matrix and the Report is the same for both unit, however, that is, that of the current speaker. Where the construction involved is still structurally clearly identifiable as reported speech it constitutes a form-function mismatch in which the typical meaning of this construction would indicate a change in perspective, but its interpretation is 'perspective persistent' in terms of Gentens et al. (2019) and Spronck et al. (2020). The loss of perspective meaning may also be iconically signalled in the linguistic structure through the various marking variations found in extended reported speech[26].

[26]Note that this phenomenon complements a reverse diachronic direction that elements *within* reported speech can display with respect to perspectival interpretations: word classes and categories that do not necessarily signal perspective meanings may gain such a meaning in the context of reported speech. A particularly prominent example of such a development is formed by pronouns, which may develop evidential meanings (cf. Zemp, 2020) or take on referential meanings specific to the reported speech context (cf. Nikitina, 2012).

The examples illustrated in this study appear to suggest that extended reported speech often also operates in the categorial twilight area between direct speech and non-direct speech. Even though most authors in our survey in **Section 2** consider reported speech expressions other than direct speech marked or even exceptional in the respective language, very few of the examples of extended reported speech cited are common direct speech structures. Pascual (2014), 49 makes a similar observation about her data sample: 'On the one hand, the cases discussed in this section share all the formal characteristics of direct speech. On the other hand, their possible appearance after complementizer 'that', their multifunctionality, and their type rather than token interpretation constitute features traditionally associated with indirect speech'. We would add that also *structurally*, extended reported speech often displays 'indirect-like' features.

4.3 A Speculative Story: Reported Speech as the Origin of Grammar

After having noted that extended reported speech constitutes a wide range of subtypes, that are nevertheless quite regular and can be related to a common semantic origin and (more impressionistically) share certain structural features, we would like to return to the research programme that we started out with at the beginning of this paper: the study of fictive interaction. The implication that extended reported speech has for this research programme is admittedly speculative, but to us it also seems to be the most exciting one: in extended reported speech a connection appears to emerge between the representation of other people's utterances and grammar. This allows us to propose a fundamental hypothesis about how these grammatical meanings may ultimately have arisen in the evolution of language.

Pascual (2014) convincingly demonstrates that metaphors of conversation are a frequent strategy for speakers to explain complex concepts and may affect language at any grammatical level. Furthermore, our ability to reason, according to Mercier

and Sperber (2017), arose out of a discursive need to evaluate the effectiveness of our arguments in conversation. In human evolution, this ability must have been predated by the capacity for being able to recognise the world view and knowledge of others as different from our own, popularly referred to as 'theory of mind' (Tomasello, 2014). Like most evolved capacities, this is not a uniquely human trait (de Waal, 2016), but it is a necessary step for the use of symbolic communication (Dor, 2017).

Built on these cognitive foundations, the assumption that language started out as situation-specific calls, developing into non-situation specific symbolic conventions for communication of ever increasing complexity (cf. Dor, 2015, ch. 8) seems relatively uncontroversial. But this scenario also assigns a central role to linguistic reflexivity in language evolution: it requires speakers to reflect on the form and meaning of what-is-said, the ability to 'turn language on itself' (Lucy, 1993). The type of linguistic structures specifically dedicated to this task are reported speech. If linguistic reflexivity, that is, thinking and talking about language, is at the heart of the complexification of grammar, reported speech is at the heart of language evolution, which would at once explain its universality in the languages of the world and its relation to grammatical categories, as indicated by the range of functions summarised in **Section 2**.

We do not wish to suggest that any of the languages cited in this paper represent an evolutionary early stage of grammatical development. Given the importance of metaphors of conversation in language (Pascual, 2014), grammaticalisation and semantic extension of reported speech structures may be cyclical or run parallel to other diachronic developments. We also do not suggest that in deep history all markers of, e.g., aspect or causation must have emerged out of reported speech. Rather, we would propose that the semantic components of reported speech provide a model for the lexical and grammatical meanings listed in (19). Once the communicative utility of this meaning is adopted by the speech community, it may have been marked through a special form of a reported speech construction, or a newly emerged form dedicated to this specific meaning. In this scenario, reported speech constructions may either have acted as a formal source for grammatical categories associated with the functions in (19) or a semantic model.

In order to test this hypothesis we need to further examine the semantic commonalities between reported speech and the respective grammatical categories involved in the extensions, as well as the semantic oppositions that exist between extended reported speech and morphological categories in the languages that both have, e.g., tense meanings based on reported speech forms and a separate morphological tense form.

Nonetheless, the regularity of the large range of semantic extensions of reported speech, as well as their apparent similarity to the meanings of some of the most basic grammatical categories in the languages of the world, is unlikely to be coincidental. Although the evolutionary story sketched here is inevitably speculative, we believe that it is also a plausible story about the development of grammatical complexity and constitution of grammatical categories. Above all, it motivates the importance of gaining a deeper understanding of the diversity of structures and meanings associated with extended reported speech and their relation to perspective expressions and grammar more generally.

5 CONCLUSION

In this article we have aimed to develop a typological approach to extended reported speech, highlighting both the wide-ranging forms and functions of the phenomenon and its apparent regularity. Ultimately, this leads us to suggest that extended reported speech constitutes a fertile birth environment for core grammatical meanings: the list of subtypes summarised in (19) includes lexical extensions alongside some of the most common verbal categories found in the languages of the world: evidentiality, modality, aspect/tense, valency change, among others.

Much work remains to be done in order to gain a fuller picture of both the semantic patterns found in extended reported speech around the world, and of the structural patterns employed to express these meanings. These typological questions should be answered in dialogue with theoretical discussions about how quotation shifts perspective and what the semantic status is of the content of a Report; as well as what aspects of reported speech are conventional and which are pragmatic.

This may ultimately lead us to an understanding of why grammar is the way it is.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

Both authors discussed the content and structure of the article. DC constructed the typological sample, produced the map and collected the examples cited in **Section 4**. SS designed the study and wrote the article.

ACKNOWLEDGMENTS

We are grateful to the reviewers for their helpful and constructive comments, to Sonja Zeman for exceptional editorial support and comments, to Angela Terrill for comments and copy-editing, as well as to the participants of our bi-weekly online data workshop on reported speech (October 2020 - June 2021), particularly (in alphabetical order), Anna Bugaeva, Sonja Gipper, Guillaume Guitang, Izabela Jordanoska, Tatiana Nikitina, Rebecca Paterson, Lacina Salué, Elena Skribnik, Adam Tallman and Denys Teptiuk for comments on an earlier draft of this paper and stimulating discussion. We thank the University of Helsinki for financial support through the project 'Language emerging from human sociality: the case of speech representation.'

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.624486/full#supplementary-material

REFERENCES

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford etc: Oxford University Press.

Bakhtin, M. (1981). *The Dialogic Imagination: Four Essays by M. M. Bakhtin*. Austin: University of Texas. Caryl Emerson and Michael Holquist.

Baranova, V. (2015). Grammaticalization Paths of the Verb *Gi-* 'say' in Kalmyk. *Mongolica Pragensia. Linguistics, Ethnolinguistics, Religion Cult.* 8 (2), 57–76.

Bickel, B., Comrie, B., and Haspelmath, M. (2008). The Leipzig Glossing Rules: Conventions for Interlinear Morpheme-By-Morpheme Glosses. Most recent version available at: https://www.eva.mpg.de/lingua/resources/glossing-rules.php.

Birchall, J. (2018). Historical Change in Reported Speech Constructions in the Chapacuran Family. *Jhl* 8 (1), 7–30. doi:10.1075/jhl.00003.bir

Bon, N. (2014). *Une grammaire de la langue stieng, langue en danger du Cambodge et du Vietnam*. Ph. D. thesis. *Université Lumière Lyon* 2.

Boye, K. (2012). *Epistemic Meaning: A Crosslinguistic and Functional-Cognitive Study*. Berlin/Boston: De Gruyter Mouton.

Bugaeva, A. (2008). Reported Discourse and Logophoricity in Southern Hokkaido Dialects of Ainu. *Gengo Kenkyu* 133, 31–75.

Casartelli, D. (2019). *Beyond Speech Reporting: A Crosslinguistic Analysis of Non-locutionary Meanings in Reported Speech*. Ph. D. thesis, Helsinki: University of Helsinki In preparation.

Chappell, H. (2012). "*Say* in Sinitic: From *Verba Dicendi* to Attitudinal Discourse Markers," in *Grammaticalization and (Inter-)subjectification*. Editors J. van der Auwera and J. Nuyts (Wetteren: Universa Press), 81–110.

Cinque, G. (1999). *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford etc: Oxford University Press.

Cinque, G. (2013). Cognition, Universal Grammar, and Typological Generalizations. *Lingua* 130, 50–65. doi:10.1016/j.lingua.2012.10.007

Clark, H. H. (2016). Depicting as a Method of Communication. *Psychol. Rev.* 123 (3), 324–347. doi:10.1037/rev0000026

Clark, H. H., and Gerrig, R. J. (1990). Quotations as Demonstrations. *Language* 66, 764–805. doi:10.2307/414729

Clendon, M. (2014). *Worrorra: A Language of the north-west Kimberley Coast*. Adelaide: University of Adelaide Press.

Coate, H. H. J. (1966). The Rai and the Third Eye North-west Australian Beliefs. *Oceania* 37 (2), 93–123. doi:10.1002/j.1834-4461.1966.tb01790.x

Cohen, D., Simeone-Senelle, M.-C., and Vanhove, M. (2002). "11. The Grammaticalization of 'say' and 'Do'," in *Reported Discourse. A Meeting Ground for Different Linguisitc Domains*. Editors T. Güldeman and M. von Rocador (Amsterdam/Philadelphia: John Benjamins), 227–251. doi:10.1075/tsl.52.17coh

Cysouw, M., and Good, J. (2013). Languoid, Doculect, and Glossonym: Formalizing the Notion 'language'. *Lang. Documentation Conservation* 7, 331–360.

Davidson, K. (2015). Quotation, Demonstration, and Iconicity. *Linguist Philos.* 38, 477–520. doi:10.1007/s10988-015-9180-1

De Smet, H. (2012). The Course of Actualization. *Language* 88 (3), 601–633. doi:10.1353/lan.2012.0056

de Waal, F. (2016). *Are We Smart Enough to Know How Smart Animals Are?* New York/London: W. W. Norton.

Deibler, E. (1971). Uses of the Verb 'to Say' in Gahuku. *Kivung* 4 (2), 101–110.

Dhakal, D. N. (2012). *Darai Grammar, Volume 489 of Languages of the World/Materials*: München: Lincom.

Dik, S. C. (1997a). *The Theory of Functional Grammar: Part 1 the Structure of the Clause*. Berlin/New York: De Gruyter Mouton.

Dik, S. C. (1997b). *The Theory of Functional Grammar: Part 2 Complex and Derived Constructions*. Berlin/New York: De Gruyter Mouton.

Dor, D. (2015). *The Instruction of Imagination: Language as a Social Communication Technology*. Oxford etc: Oxford University Press.

Dor, D. (2017). The Role of the Lie in the Evolution of Human Language. *Lang. Sci.* 63, 44–59. doi:10.1016/j.langsci.2017.01.001

Dryer, M. S. (1989). Large Linguistic Areas and Language Sampling. *Stud. Lang.* 13 (2), 257–292. doi:10.1075/sl.13.2.03dry

Dryer, M. S., and Haspelmath, M. (2013). in *The World Atlas of Language Structures Online* (Leipzig: Max Planck Institute for Evolutionary Anthropology).

Ershova, K. (2012). Reported Speech and Reportative Grammaticalization in Besleney Kabardian. In Proceedings of the Second Central European Conference in Linguistics for Postgraduate Students, 71–87.

Everett, D. L. (2008). "Wari' Intentional State Constructions," in *Investigations of the Syntax–Semantics–Pragmatics Interface*. Editor R. D. Van Valin Jr. (Amsterdam/Philadelphia: John Benjamins), 381–409. doi:10.1075/slcs.105.27eve

Forker, D. (2019). Reported Speech Constructions in Sanzhi Dargwa and Their Extension to Other Areas of Grammar. *Sprachwissenschaft* 44 (2).

Gentens, C., Sansiñena, M. S., Spronck, S., and Van linden, A. (2019). Irregular Perspective Shifts and Perspective Persistence, Discourse-Oriented and Theoretical Approaches. *Prag.* 29 (2), 155–169. Special issue on perspective-indexing constructions: discourse and theory. doi:10.1075/prag.18050.gen

Gentens, C. (2020). *The Factive-Reported Distinction in English*. New York: De Gruyter.

Güldemann, T. (2008). *Quotative Indexes in African Languages: A Synchronic and Diachronic Survey*. Berlin: Mouton de Gruyter.

Heine, B., and Kuteva, T. (2002). *World Lexicon of Grammaticalization*. Cambridge etc: Cambridge University Press.

Hengeveld, K. (1989). Layers and Operators in Functional Grammar. *J. Ling.* 25 (1), 127–157. doi:10.1017/s0022226700012123

Hooper, J. B., and Thompson, S. A. (1973). On the Applicability of Root Transformations. *Linguistic Inq.* 4 (4), 465–497.

Hsieh, F. (2012). On the Grammaticalization of the Kavalan SAY Verb Zin. *Oceanic Linguistics* 51 (2), 464–489. doi:10.1353/ol.2012.0025

Jakobson, R. (1957). *Shifters, Verbal Categories and the Russian Verb*. Cambridge, MA: Department of Slavic Languages and Literatures Harvard University.

J. A. Lucy (1993). in *Reflexive Language: Reported Speech and Metapragmatics* (Cambridge etc: Cambridge University Press).

Jarque, M. J., and Pascual, E. (2015). Conversational Structures Marking Evidential Values in Catalan Sign Language. *eHumanista/IVITRA* 8, 421–445.

Jarque Moyano, M. J. (2016). "What About? Fictive Question-Answer Pairs for Non-Information-Seeking Functions Across Signed Languages," in *The Conversation Frame: Forms and Functions of Fictive Interaction*. Editors Pascual and Sandler. (Amsterdam/Philadelphia: John Benjamins), 171–192. doi:10.1075/hcp.55.09jar

Klamer, M. (1998). *A Grammar of Kambera*. Berlin, New York: Berlin: Mouton Grammar LibraryMouton de Gruyter.

Klamer, M. (2000). How Report Verbs Become Quote Markers and Complementisers. *Lingua* 110, 69–98. doi:10.1016/s0024-3841(99)00032-7

Konnerth, L. (2020). Recycling through Perspective Persistence in Monsang (Trans-himalayan). *Fol* 27 (1), 55–77. doi:10.1075/fol.20002.kon

Kuteva, T., Heine, B., Hong, B., Long, H., Narrog, H., and Rhee, S. (2019). *World Lexicon of Grammaticalization*. Cambridge University Press.

Larson, M. L. (1978). *The Functions of Reported Speech in Discourse*. Dallas: Summer Institute of Linguistics.

Lefebvre, C., and Brousseau, A.-M. (2002). *A Grammar of Fongbe of Mouton Grammar Library*. Berlin, New York: Berlin: Mouton de Gruyter.

Loughnane, R. (2009). *A Grammar of Oksapmin*. Ph. D. thesis, Melbourne: University of Melbourne.

Martin Haspelmath, M. (2010). Comparative Concepts and Descriptive Categories in Crosslinguistic Studies. *Language* 86 (3), 663–687. doi:10.1353/lan.2010.0021

Matić, D., and Pakendorf, B. (2013). Non-canonical SAY in Siberia. *Stud. Lang.* 37 (2), 356–412.

McGregor, W. B. (2007). "2. A Desiderative Complement Construction in Warrwa," in *Language Description, History and Development*. Editors J. Siegel, J. Lynch, and D. Eades (Amsterdam/Philadelphia: John Benjamins), 27–40. doi:10.1075/cll.30.08mcg

McGregor, W. B. (2019). Reported Speech as a Dedicated Grammatical Domain - and Why Defenestration Should Not Be Thrown Out the Window. *Linguistic Typology* 23 (1), 207–219. doi:10.1515/lingty-2019-0010

McGregor, W. B. (2014). "The 'Say, Do' Verb in Nyulnyul, Warrwa, and Other Nyulnyulan Languages Is Monosemic," in *Events, Arguments and Aspects: Topics in the Semantics of Verbs*. Editor K. Robering (Amsterdam/Philadelphia: John Benjamins), 301–328. doi:10.1075/slcs.152.08gre

Mercier, H., and Sperber, D. (2017). *The Enigma of Reason*. Cambridge, MA: Harvard University Press.

Miestamo, M., Bakker, D., and Arppe, A. (2016). Sampling for Variety. *Linguistic Typology* 20 (2). doi:10.1515/lingty-2016-0006

Moroz, G. (2017). *Lingtypology: Linguistic Typology and Mapping*.

Nikitina, T. (2020). Logophoricity and Shifts of Perspective. *Fol.* 27 (1), 78–99. doi:10.1075/fol.20001.nik

Nikitina, T. (2012). Personal Deixis and Reported Discourse: Towards a Typology of Person Alignment. *Linguistic Typology* 16 (2), 233–263.

Pascual, E. (2007). "Fictive Interaction Blends in Everyday Life and Courtroom Settings," in *Mental Spaces in Discourse and Interaction*. Editors T. Oakley and A. Hougaard (Amsterdam/Philadelphia: John Benjamins), 79–107.

Pascual, E. (2014). *Fictive Interaction: The Conversation Frame in Thought, Language, and Discourse*. Amsterdam/Philadelphia: John Benjamins.

Pascual, E. (2002). *Imaginary Trialogues: Conceptual Blending and Fictive Interaction in Criminal Courts*. Utrecht: LOT.

Pascual, E., and Sandler, S. (2016a). *In The Conversation Frame: Forms and Functions of Fictive Interaction* (Amsterdam/Philadelphia: John Benjamins).

Pascual, E., and Sandler, S. (2016b). "Fictive Interaction and the Conversation Frame," in *The Conversation Frame: Forms and Functions of Fictive Interaction*. Editors E. Pascual and S. Sandler (Amsterdam/Philadelphia: John Benjamins), 3–22. doi:10.1075/hcp.55.01pas

Pawley, A., and Bulmer, R. (2011). *Research School of Pacific and Asian Studies*. Canberra: Australian National University. *A Dictionary of Kalam With Ethnographic Notes*, Volume 630 of *Pacific Linguistics*.

Pennington, R. (2018). *A Grammar of Ma Manda: A Papuan Language of Morobe Province, Papua New Guinea, Volume 18 of Outstanding Grammars from Australia*. München: LINCOM.

Reesink, G. P. (1993). 'Inner Speech' in Papuan Languages. *Lang. Linguistics Melanesia* 24, 217–225.

Rijkhoff, J., Bakker, D., Hengeveld, K., and Kahrel, P. (1993). A Method of Language Sampling. *Sl* 17 (1), 169–203. doi:10.1075/sl.17.1.07rij

Rumsey, A. (1982). *An Intra-Sentence Grammar of Ungarinjin*, North-Western Australia. Canberra: Pacific Linguistics 92 (2), 157–166. doi:10.1525/aa.1990.92.2.02a00060

Rumsey, A. (1990). Wording, Meaning, and Linguistic Ideology. *Am. Anthropologist* 92 (2), 346–361. doi:10.1525/aa.1990.92.2.02a00060

Sandler, S., and Pascual, E. (2019). In the Beginning There Was Conversation: Fictive Direct Speech in the Hebrew Bible. *Prag.* 29 (2), 250–276. doi:10.1075/prag.18047.san

Saxena, A. (1988). On Syntactic Convergence: The Case of the Verb 'say' in Tibeto-Burman. In Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society, 375–388. doi:10.3765/bls.v14i0.1781

Spronck, S. (2015). *Reported Speech in Ungarinyin: Grammar and Social Cognition in a Language of the Kimberley Region*. Western Australia: Ph. D. thesis The Australian National University.

Spronck, S. (2016). "Evidential Fictive Interaction (In Ungarinyin and Russian)," in *The Conversation Frame: Forms and Functions of Fictive Interaction*. Editors E. Pascual and S. Sandler (Amsterdam/Philadelphia: John Benjamins), 255–275. doi:10.1075/hcp.55.13spr

Spronck, S., and Nikitina, T. (2019). Reported Speech Forms a Dedicated Syntactic Domain. *Linguistic Typology* 23 (1), 119–159. doi:10.1515/lingty-2019-0005

Spronck, S., Van linden, A., Gentens, C., and Sansiñena, M. S. (2020). Perspective Persistence and Irregular Perspective Shift. *Fol.* 27 (1), 1–6. doi:10.1075/fol.20005.spr

Thornes, T. J. (2003). *A Northern Paiute Grammar and Texts*. EugeneAnn Arbor: Ph. D. thesis University of Oregon.

Tomasello, M. (2014). *A Natural History of Human Thinking*. Cambridge/London: Harvard University Press.

van der Voort, H. (2002). "The Quotative Construction in Kwaza and its (De-) Grammaticalisation," in *Current Studies on South American Languages Indigenous Languages of Latin America*. Editors M. Crevels, S. van de Kerke, S. Meira, and H. van der Voort, 307–328.

Völlmin, S. (2017). *Towards a Grammar of Gumer: Phonology and Morphology of a Western Gurage Variety*. Ph. D. thesis, Zürich: University of Zürich.

Vydrina, A. (2017). *A Corpus-Based Description of Kakabe, a Western Mande Language: Prosody in Grammar*. Paris: Ph. D. thesis INALCO.

Vygotsky, L. S. (1987). *Thinking and Speech. The Collected Works of Lev Vygotsky*. New York, NY: Plenum Press.

Wiemer, B. (2018). *Catching the Elusive: Lexical Evidentiality Markers in Slavic Languages (A Questionnaire Study and its Background)*. Berlin: Peter Lang.

Wilson, J. (2017). *A Grammar of Yeri: A Torricelli Language of Papua New Guinea*. Ph. D. thesis, Buffalo: State University of New York at Buffalo.

Yaroshevich, S. (2020). Is *Thinking* like *Saying*? the Case of Tabasaran. Paper read at the 53rd Annual Meeting of the *Societas Linguistica Europaea* in the workshop 'The grammar of thinking: Comparing reported thought andreported speech across languages'.

Zeman, S. (2019). The Emergence of Viewpoints in Multiple Perspective Constructions. *Prag.* 29 (2), 226–249. doi:10.1075/prag.18051.zem

Zemp, M. (2020). Evidentials and Their Pivot in Tibetic and Neighboring Himalayan Languages. *Fol.* 27 (1), 29–54. doi:10.1075/fol.20003.zem

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 Spronck and Casartelli. 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.

GLOSSARY

{} elements in the ME gloss that are not part of the extended reported speech construction

ADD additive (Besleney Kabardian)

AF agent focus (Kavalan)

ANAPH anaphoric pronoun

ART article (Kakabe)

CM conjugation marker (Warrwa)

CONJV conjunctive (Stieng)

CVB.MOD modal converb or converb of manner (Kalmyk)

DIR directional (Besleney Kabardian, Mandarin)

FP far past (Usan)

HOR/HORT hortative (Darai, Kalam)

ICVB

imperfective converb (Sanzhi)

IF immediate future (Oksapmin)

IP instrumental prefix (Northern Paiute)

LF locative focus (Kavalan)

LOG logophoric pronoun

'M' verbal '-m' suffix (Gumer) (Völlmin, 2017, 152)

MM middle marker (reflexive, reciprocal, passive, middle; Northern Paiute)

MIN minimal number (Warrwa)

MIR mirative (Darai)

NF non-finite (Newari)

NFUT non-future

NONVIS.EV nonvisual evidential (Ainu)

NPC non-past conjunct (Newari)

OBL oblique stem marker (Sanzhi)

PART particle (Darai)

PD past disjunct (Newari)

PN personal name (Ma Manda)

POT potential mood

PRET preterite (Sanzhi)

PRIOR priorite (Kalam)

PRT particle (Wan, Mandarin)

RECG recognitional (Oksapmin)

RED reduplication (Usan)

SEQ sequential (Darai)

SS same subject (Usan)

UF uncertain future (Usan)

W-CLASS second ('Wu-') neuter gender class (Ngarinyin)

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 September 2021
doi: 10.3389/fpsyg.2021.698986

Reading Minds, Reading Stories: Social-Cognitive Abilities Affect the Linguistic Processing of Narrative Viewpoint

Lynn S. Eekhof[1]*, *Kobie van Krieken*[1], *José Sanders*[1] and *Roel M. Willems*[1,2,3]

[1] Centre for Language Studies, Radboud University, Nijmegen, Netherlands, [2] Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, Netherlands, [3] Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

Although various studies have shown that narrative reading draws on social-cognitive abilities, not much is known about the precise aspects of narrative processing that engage these abilities. We hypothesized that the linguistic processing of narrative viewpoint—expressed by elements that provide access to the inner world of characters—might play an important role in engaging social-cognitive abilities. Using eye tracking, we studied the effect of lexical markers of perceptual, cognitive, and emotional viewpoint on eye movements during reading of a 5,000-word narrative. Next, we investigated how this relationship was modulated by individual differences in social-cognitive abilities. Our results show diverging patterns of eye movements for perceptual viewpoint markers on the one hand, and cognitive and emotional viewpoint markers on the other. Whereas the former are processed relatively fast compared to non-viewpoint markers, the latter are processed relatively slow. Moreover, we found that social-cognitive abilities impacted the processing of words in general, and of perceptual and cognitive viewpoint markers in particular, such that both perspective-taking abilities and self-reported perspective-taking traits facilitated the processing of these markers. All in all, our study extends earlier findings that social cognition is of importance for story reading, showing that individual differences in social-cognitive abilities are related to the linguistic processing of narrative viewpoint.

Keywords: social cognition, narrative, viewpoint, perspective, eye tracking

OPEN ACCESS

Edited by:
Jorrig Vogels,
University of Groningen, Netherlands

Reviewed by:
Lena Wimmer,
University of Freiburg, Germany
Chia-lin Lee,
National Taiwan University, Taiwan

***Correspondence:**
Lynn S. Eekhof
lynn.eekhof@ru.nl

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

Received: *22 April 2021*
Accepted: *27 August 2021*
Published: *28 September 2021*

Citation:
Eekhof LS, van Krieken K, Sanders J and Willems RM (2021) Reading Minds, Reading Stories: Social-Cognitive Abilities Affect the Linguistic Processing of Narrative Viewpoint. Front. Psychol. 12:698986. doi: 10.3389/fpsyg.2021.698986

INTRODUCTION

Although reading might seem a rather solitary activity compared to engaging in social interaction, many scholars have argued that social-cognitive processes play an important role during story reading. That is, the abilities we use in our daily lives to make sense of the emotions, beliefs, intentions, and behavior of others, such as empathy, emotion recognition, and theory of mind, are also engaged when we read about fictional others in stories (Oatley, 1999; Zunshine, 2006; Mar and Oatley, 2008). However, despite research underlining the importance of social-cognitive abilities for story reading, it is not clear exactly what aspects of narrative processing require readers to put these abilities to work. In other words, there is relatively little research on the relationship between social-cognitive abilities and the processing of specific narrative characteristics.

In this study we therefore investigated how individual differences in readers' social-cognitive abilities are related to a crucial aspect of narrative processing, namely the linguistic processing of narrative viewpoint. In what follows, we will first discuss the role of social cognition during narrative reading. After introducing the multidimensional concept of narrative viewpoint, we will discuss why the linguistic processing of narrative viewpoint is likely related to readers' social-cognitive abilities.

Social Cognition and Narrative Reading

The contention that narratives engage social-cognitive abilities follows from two views on what constitutes a narrative. Firstly, narratives are often loosely defined as the representation of a sequence of events that are related in time (e.g., Toolan, 2001; for an overview, see Ryan, 2007; Abbott, 2008). In line with these plot-focused definitions, Zunshine (2003, 2006) has argued that much like displays of behavior in real life, textual descriptions of narrative events can invite readers to use their theory of mind abilities to assign mental states to the characters performing these events. For example, descriptions of the actions and/or body language of characters might leave the reader wondering why a character behaves in a certain way, or guessing how the character feels, living through these events. Hence, on this account, social cognition might play an important role in making sense of the plot of narratives.

In addition, scholars have stressed the subjective aspect of narratives (e.g., Bruner, 1986; Bal, 2009). For example, Bal (2009, p. 10) gives the following definition: "[…] a series of connected events caused or *experienced by actors* presented in a specific manner" (emphasis added). On such accounts, narratives are unique because the events always presuppose the presence of someone who experiences these events. As a result, authors can choose to directly represent the internal states of their protagonist through the use of, for example, mental verbs (*to think, to believe*) or other perspectivization techniques that grant the reader direct access to the mind of story characters (van Krieken et al., 2017; Eekhof et al., 2020). These mental representations might also elicit a form of perspective taking in readers (van Krieken et al., 2017).

Comprising the above approaches, narratives can be seen as a sequence of textual cues, guiding the reader to form a cognitive, social, and emotional simulation of what is described in the narrative (Oatley, 1999; Mar and Oatley, 2008). Crucially, such a simulation also requires readers to employ social-cognitive abilities to "reconstruct the minds" of the narrative characters (Ryan, 2007, p. 28). In a similar vein, Koopman and Hakemulder (2015, p. 91) argue that an important aspect of being absorbed in a story world is "empathic imagination," a process whereby the reader uses empathic abilities to imagine "how it would be to be in the shoes of a particular character."

Several studies provide (indirect) evidence for the involvement of social-cognitive abilities during narrative reading. For example, a range of fMRI studies has shown that brain regions that are part of the mentalizing network (e.g., mPFC, bilateral pSTS/TPJ) are also activated during narrative comprehension (e.g., Mason and Just, 2009; for a meta-analysis see Mar, 2011). Furthermore, theory of mind development in children parallels developments in the processing of narratives. For example, recall of socially relevant details of a story has been found to increase between adolescence and adulthood, potentially mirroring a development in social-cognitive abilities in the same period (Pavias et al., 2016). Similarly, in a story retelling task, both age and theory of mind abilities were found to positively predict the ability to coordinate story characters' actions and mental states in preschoolers (Pelletier and Wilde Astington, 2004; see also Fernández, 2013). Finally, in a longitudinal study, children's theory of mind scores at age four predicted narrative comprehension and recall two and a half years later (Atkinson et al., 2017). Taken together, these studies tentatively suggest that social-cognitive abilities play a role in narrative comprehension, both in adults and in children.

However, many of the previous studies have taken a rather broad look at narrative processing, looking at the relationship between social cognition and story reading in general (fMRI studies), or narrative comprehension and recall after reading (developmental studies). As a result, not much in known yet about the specific aspects of narrative processing that engage readers' social-cognitive abilities. Two fMRI studies, however, did find that processing stories rich in descriptions of characters' mental states (Tamir et al., 2016) and stories with negative valence (i.e., dealing with negative events such as crimes and disasters; Altmann et al., 2012) elicited more activation in brain regions related to theory of mind (e.g., dmPFC subnetwork) compared to stories with less socially relevant content and stories with positive valence, suggesting that, broadly speaking, processing of social and negatively valenced narrative content draws on social-cognitive abilities. Nevertheless, more research is needed to elucidate what exactly it is about narratives that requires readers to use their social-cognitive abilities. That is, future studies should provide a more detailed account of the facets of narrative processing that engage social cognition.

Narrative Viewpoint

An aspect of narrative processing that might play a role in the engagement of social-cognitive abilities during reading is the linguistic processing of narrative viewpoint. As explained above, narratives presuppose the presence of an "experiencing subject" (Sanders and Redeker, 1996). Typically, the events in narratives are always grounded in and related through the subjective viewpoints (or perspectives) of these experiencing characters and/or narrators (Sanders, 1994). During reading, readers align themselves with the events and dynamically take the perspective of one or more of the characters and/or narrators, both in terms of their spatio-temporal viewpoint in the narrative world and in terms of their inner viewpoint or consciousness (Vandelanotte, 2017). In their Linguistic Cues Framework, van Krieken et al. (2017) distinguish between multiple dimensions of viewpoint and argue that each dimension is regulated by different linguistic cues. For example, perceptual

Abbreviations: ART, Author Recognition Test; IRI, Interpersonal Reactivity Index; MET, Multifaceted Empathy Test; STOMP, Spontaneous Theory of Mind Protocol; VPT, Visual Perspective-taking Task.

viewpoint, referring to the narrative representation of characters' perceptions and sensations, can be expressed through the use of perceptual verbs (*to watch, to hear*), emotional viewpoint, referring to the narrative representation of characters' emotions, can be expressed through the use of emotion adjectives (*angry, delighted*), and cognitive viewpoint, referring to the narrative representation of characters' mental states, can be expressed through epistemic markers (*probably, definitely*). Crucially, these linguistic viewpoint markers are hypothesized to invite the reader to identify with a particular subject in the narrative (van Krieken et al., 2017). In other words, linguistic markers of viewpoint can be seen as a signal to the reader to engage in perspective taking. As such, viewpoint markers might play an important role in eliciting social-cognitive processes during narrative reading, given that perspective taking is an important aspect of social cognition (Frith and Frith, 2006; Goldman and de Vignemont, 2009; Healey and Grossman, 2018).

Interestingly, literature on the development of language and theory of mind provides evidence that social cognition plays a role in the linguistic processing of viewpoint markers such as verbs of cognition and emotion, although this has not always been tested in narrative contexts (for a general overview on the relationship between language acquisition and theory of mind acquisition see Milligan et al., 2007). For example, comprehension of verbs of cognition in short stories has been found to be related to performance on first-order false belief tasks, and to a lesser degree to second-order false belief tasks in children aged between 4 and 8 years (Antonietti et al., 2006), and to second-order false belief tasks in children aged between 8 and 11 years (Grazzani and Ornaghi, 2012). Similarly, in a task that required children to make sense of spoken instructions to find an object, comprehension of modal verbs and adjectives, which can be considered markers of cognitive viewpoint (van Krieken et al., 2017; Eekhof et al., 2020), was significantly related to performance on first-order false belief tasks in four-year-olds (Moore et al., 1990). Furthermore, comprehension of verbs of emotion on a short-story task was significantly correlated to emotion understanding in a study with seven- to ten-year-olds (Ornaghi and Grazzani, 2013). These results indicate that individual differences in social-cognitive abilities are related to the linguistic processing of at least emotional and cognitive viewpoint markers in children, suggesting that social cognition and the processing of narrative viewpoint somehow go hand in hand.

All in all, viewpoint markers are likely to play a role in engaging social-cognitive processes during the reading of narratives, as at least in childhood the processing of viewpoint markers has been found to be related to individual differences in social-cognitive abilities. Hence, we wanted to further investigate the relationship between the linguistic processing of viewpoint markers in narratives and social-cognitive abilities in adult readers. Our rationale was that if individual differences in social-cognitive abilities affect the linguistic processing of narrative viewpoint, this highly suggests that markers of narrative viewpoint engage these social-cognitive abilities.

The Current Study

We set out to study how the linguistic processing of narrative viewpoint markers is affected by individual differences in social-cognitive abilities, using eye-tracking. Hence, as a first step we aimed to find out how perceptual, cognitive, and emotional viewpoint markers affect reading behavior. More importantly, we then aimed to study how these effects are modulated by social-cognitive abilities. In sum, the current study aimed to answer the following research question:

> *What is the effect of perceptual, cognitive, and emotional viewpoint markers in narratives on reading behavior, and how is this effect modulated by individual differences in social-cognitive abilities?*

Based on a study by Mak and Willems (2018), who found that narrative passages describing characters' perceptions, thoughts, and emotions increased reading times, we hypothesized viewpoint markers to be processed slower than non-viewpoint markers. We also hypothesized that, in general, social-cognitive abilities would modulate the effect of viewpoint markers on reading behavior. More specifically, and based on the research that shows that theory of mind abilities positively predict narrative comprehension in general (e.g., Atkinson et al., 2017), and the acquisition of epistemic markers, verbs of cognition, and verbs of emotion specifically (Moore et al., 1990; Antonietti et al., 2006; Grazzani and Ornaghi, 2012; Ornaghi and Grazzani, 2013), we tentatively hypothesized that social-cognitive abilities lead to faster processing of viewpoint markers (i.e., more skipping, shorter gaze durations, less rereading). We did not have specific hypotheses about the modulating effect of social-cognitive abilities for each specific viewpoint marker category separately.

METHODS

An eye-tracking study was designed to study the linguistic processing of viewpoint markers, by looking at the effect of these markers on skip rate, gaze duration, and re-reading rate (for the justification of these eye-tracking methods, see Pre-Processing of Eye-Tracking Data). We chose to focus on markers of perceptual, cognitive, and emotional viewpoint as we expected the processing of these viewpoint dimensions to be most relevant to the domain of social cognition. We opted for eye tracking as an appropriate method for several reasons. Firstly, contrary to, for example, self-paced reading, eye tracking provides a relatively ecologically valid way to study reading, as participants can be presented with large pieces of texts without any additional task. Furthermore, eye tracking has proven to be a useful method to study individual differences in narrative processing, as evidenced by recent studies on individual differences in mental simulation (Mak and Willems, 2018), sensitivity to literary style (van den Hoven et al., 2016), sensitivity to lexical characteristics and absorption (Eekhof et al., 2021), metaphor processing (Vries et al., 2018), and reading style (Faber et al., 2020) during story reading. Contrary to previous studies, we used a non-fictional rather than a fictional narrative, published in a well-established journalistic weekly magazine. A main function of narrative journalism is to increase the general audience's understanding of

society in all its complexities and to enhance the audience's sense of being part of that society (van Krieken and Sanders, 2021). In this genre, narrative perspective taking is typically stimulated by multiple linguistic viewpoints that readers are invited to share (van Krieken et al., 2015). As viewpoint techniques are typical of narratives regardless of their fictionality we believe research on the relationship between social cognition and narratives should be expanded to include non-fictional narratives as well (see also Koopman and Hakemulder, 2015).

Participants

Based on a power simulation (see **Supplementary Materials**) we aimed for a sample of 90 participants. Taking into account the high rate of data rejection in eye-tracking studies with long texts, we recruited 114 native speakers of Dutch with normal or corrected to normal vision and no history of reading disorders from the participant pool of Radboud University to take part in the experiment in exchange for money (€15) or course credit. Three participants did not finish the experiment because of time constraints or technical failure. Of the remaining participants, 21 had to be excluded due to poor quality of eye-tracking data (see Pre-Processing of Eye-Tracking Data). After data rejection, the final sample contained data from 90 participants, aged between 18 and 48 ($M = 23.30$, $SD = 5.49$, 67 females, 23 males). The experiment was conducted in accordance with the declaration of Helsinki and was approved by the institutional ethics assessment committee (Approval Number 2018-3568).

Materials

Narrative

A Dutch non-fictional narrative (i.e., journalistic narrative; see van Krieken, 2019; van Krieken and Sanders, 2021) published in a weekly Dutch news magazine, *Vrij Nederland*, was presented to all participants[1]. The story describes a real-life missing person case and is told from the perspective of the missing man's brother, who struggles to find peace during the 16 years that his younger brother is missing. At the end of the story, the missing man's remains are found in a river and it is revealed that he has passed away as the result of a car crash. All paratextual elements (e.g., pictures and pull quotes) except the title were removed for the experiment, resulting in a 5,077-word text[2].

The ViewPoint Identification Procedure (VPIP; Eekhof et al., 2020) was applied to identify all markers of perceptual, cognitive, and emotional viewpoint in the narrative. This procedure defines perceptual viewpoint markers as content words that express the perceptual viewpoint, i.e., the perceptions and bodily sensations, of characters and/or narrators, and operationalizes these as verbs of perception (e.g., *to see, to hear, to smell*), verbs of bodily sensation (e.g., *to itch, to sting*), and other content words morphologically related to these verb types (e.g., *sight-to see, itchy-to itch*). Cognitive viewpoint markers are defined as content words that express the cognitive viewpoint, i.e., the thoughts, beliefs, intentions and/or desires, of characters and/or narrators. These markers are operationalized as verbs of cognition (e.g., *to think, to believe*), including modal epistemic verbs (e.g., *should, might*), modal epistemic adverbs (e.g., *possibly, definitely*), and morphologically related content words (e.g., *thought-to think, possible-possibly*). Finally, the VPIP defines emotional viewpoint markers as content words that express the emotional viewpoint, i.e., the emotions, of characters and/or narrators, and operationalizes these as verbs of emotion (e.g., *to disdain, to love*), adjectives of emotion (e.g., *angry, bewildered*), and morphologically related content words (e.g., *disdain-to disdain, anger-angry*).

The narrative was coded by the first author according to the steps of the VPIP (Eekhof et al., 2020). That is, first the text was read, then the narrative was divided into 5,032 lexical units, with complex phrasal verbs (e.g., *uitkijken, hij kijkt uit* "to look out, he looks out") being treated as a single lexical unit. Function words were then disregarded, and for the remaining content words it was determined whether the lexical unit in its narrative context was related to one of the three viewpoint dimensions, and whether the lexical unit could be considered a viewpoint marker for that dimension. To assess the reliability of the procedure, 20% of the content words of the narrative were then also independently coded by the second author. As inter-rater reliability for both the binary decision (viewpoint marker vs. non-viewpoint marker; 96.81%, $\kappa = 0.84$), and categorical decision (perceptual vs. cognitive vs. emotional vs. non-viewpoint markers; 96.31%, $\kappa = 0.82$) were almost perfect, the ratings of the first author were used for the analyses. Two hundred and ninty-two lexical units (300 words) were scored as viewpoint markers: 86 lexical units (93 words) were marked as perceptual viewpoint markers, 146 lexical units (148 words) were marked as cognitive viewpoint markers, and 59 lexical units (59 words) were marked as emotional viewpoint markers. An example from the coded narrative is given in **Table 1**. More examples can be found in **Supplementary Table 1**. All words that were not coded as perceptual, cognitive, or emotional viewpoint markers, were marked as "non-viewpoint marker." As the viewpoint markers were all content words, we decided to also disregard function words from the non-viewpoint marker category. Hence, besides the 300 viewpoint markers, 2,510 non-viewpoint marking content words were used as a baseline in the analyses (see also Pre-Processing of Eye-Tracking Data). For information on the distribution of word classes in the different viewpoint marker categories, see **Supplementary Table 2**.

Measures of Social-Cognitive Abilities

As previous research is unclear about the specific aspects of social-cognitive abilities that could play a role in the processing of narrative viewpoint, we decided to use a combination of self-report and performance-based measures that tap into a broad spectrum of social-cognitive abilities. As much as possible, we included measures that were not susceptible to ceiling effects in a neurotypical population. Moreover, we included both linguistically-mediated tasks (e.g., Spontaneous Theory of Mind Protocol; Rice and Redcay, 2015) and measures that, at least at

[1]Source: Teunissen, P. (2015). *Zestien jaar vermist, zestien jaar zoeken. Waar was Joske gebleven?* Vrij Nederland. Available online at: https://www.vn.nl/zestien-jaar-vermist-zestien-jaar-zoeken-waar-was-joske-gebleven/.

[2]Due to experimenter error, 134 words of the original narrative were not presented to the participants. These words belonged to the introduction and were not crucial to the coherence or comprehensibility of the narrative plot structure.

TABLE 1 | A coded excerpt from the stimulus narrative.

Dutch original	English translation
"Jos is niet thuisgekomen." Pa was die morgen op zijn kamer gaan **kijken**. Zijn bed was onbeslapen. Pa had niets **gehoord**. […] Een mengeling van **ergernis** en **ongerustheid** welde in Gerard op. […] De volgende dag kwamen twee politiemensen bij de familie Mahler op bezoek. Ze **hoorden** het verhaal van Gerard en zijn zussen **aan**. Dat Jos wel van feestvieren **hield**, maar nooit zomaar weg **zou** blijven. De beambten **suggereerden** dat het viertal ergens anders was gaan doorfeesten.	"Jos hasn't come home." Dad had gone to **look** in his room that morning. His bed was untouched. Dad had **heard** nothing. [...] A mixture of **annoyance** and **anxiety** welled up in Gerard. [...] The next day, two policemen visited the Mahler family. They **listened** to the story of Gerard and his sisters. That Jos did **like** to party, but **would** never just stay away. The officers **suggested** that the foursome had continued partying somewhere else.

Viewpoint Markers are Printed in Bold, with Perceptual Viewpoint Markers Marked in Blue, Cognitive Viewpoint Markers Marked in Green, and Emotional Viewpoint Markers Marked in Yellow.

face value, are not linguistically-mediated (e.g., the emotional trials of the Multifaceted Empathy Test; Dziobek et al., 2008). We reasoned that if social-cognitive abilities, as measured with non-linguistic tasks, affect the processing of narrative viewpoint, this is extra strong evidence that there is a relationship between social cognition and narrative processing that goes beyond any potentially confounding effects of language ability.

Interpersonal Reactivity Index

As a first measure, we included the validated Interpersonal Reactivity Index (IRI; Davis, 1983), which is a multi-dimensional, self-report measure of trait empathy that taps into participants' tendency to feel concern for others (Empathic Concern, e.g., *I often have tender, concerned feelings for people less fortunate than me*), take the perspective of others (Perspective Taking, e.g., *I try to look at everybody's side of a disagreement before I make a decision*), feel anxious in emotional situations (Personal Distress, e.g., *I sometimes feel helpless when I am in the middle of a very emotional situation*), and emotionally engage with fictional others (Fantasy, e.g., *I really get involved with the feelings of the characters in a novel*). The 28 items of the IRI (Davis, 1983) were presented with 7-point scales (1 = *disagree*, 7 = *agree*). A Dutch translation partially based on De Corte et al. (2007) and Mak and Willems (2018) was used.

Multifaceted Empathy Test

Although previous research on the relationship between reading narratives and social-cognitive performance has often used the Reading the Mind in the Eyes Task (RMET; Baron-Cohen et al., 2001), this measure has recently received criticism for its poor internal consistency, homogeneity, and content validity (e.g., Olderbak et al., 2015; Turner and Felisberti, 2017; Black, 2019). Hence, as an alternative for the RMET we chose to include the Multifaceted Empathy Test (MET; Dziobek et al., 2008), which is a validated measure that uses participants' responses to ecologically valid pictures (i.e., full-body pictures of people in various daily situations experiencing a wide range of emotions) to assess both emotion recognition[3] and emotional empathy. A potential downside of the MET is that it has been developed mainly for use in non-neurotypical populations (e.g., patients with an autism spectrum disorder, Dziobek et al., 2008; patients with narcissistic personality disorder, Ritter et al., 2011; patients with borderline personality disorder, Dziobek et al., 2011), and as a result may be susceptible to ceiling effects in a neurotypical population (Turner and Felisberti, 2017).

For the MET (Dziobek et al., 2008) participants viewed 40 pictures of people in various situations and were asked to select an emotion word from a list of four options that matched the emotion the person in each photo was experiencing as closely as possible (emotion recognition), and to rate the degree to which they "felt along" with the person in the picture by indicating the degree to which they experienced the same emotion as the person in the picture on a 9-point scale (1 = *not at all*, 9 = *a lot*; emotional empathy). Emotion recognition and emotional empathy were assessed in alternating blocks. Hence, each picture occurred twice: once in an emotion recognition block, and once in an emotional empathy block. Each block consisted of 10 pictures, resulting in a total of 8 blocks (4 emotion recognition blocks and 4 emotional empathy blocks). To avoid a confounding effect of vocabulary knowledge, a glossary of synonyms and example sentences for each emotion word that was used in the emotion recognition trials was provided.

The 109 German emotion words of the emotion recognition trials were translated into Dutch, using a similar method as Foell et al. (2018), who translated the test from German to English. The first author translated the words from German to Dutch using the online version of the dictionary *Van Dale Groot woordenboek der Nederlandse taal* (Den Boon and Geeraerts, 2005). Then, a backtranslation was performed by an independent German scholar. For 76 words, the backtranslation matched the original German word, in which case the Dutch translation was finalized. The procedure was repeated for the remaining 33 cases for which the backtranslation did not match the original German word. After the second round, 21 unclear cases remained. These were resolved by discussion between the first author and the German translator. The translation resulted in a list of 107 unique Dutch emotion words. In two cases, a single Dutch word was chosen as a translation for two distinct German words (*träumerisch* and *verträumt* were both translated as *dromerig*, "dreamy"; *beglückt* and *erfreut* were both translated as *verheugd*, "joyful").

Visual Perspective-Taking Task

We also included the Visual Perspective-taking Task (VPT; Samson et al., 2010), which measures participants' ability to alternate between their own perspective and the perspective of an avatar. Although strictly speaking the VPT is a measure

[3]Although Dziobek et al. (2008) argue that the MET measures cognitive empathy, Oakley et al. (2016) have argued that social-cognitive tasks that measure participants' ability to assign mental states or emotions to pictures of faces or eyes reflect emotion recognition rather than theory of mind or cognitive empathy.

of visual perspective taking, we included it as a measure of social cognition, as the capacity to switch between egocentric and altercentric perspectives has been described as one of the fundamentals of social cognition (Fuchs, 2015). Moreover, aspects of trait empathy have been related to reduced altercentric intrusion, i.e., reduced interference from the perspective of the avatar (e.g., Nielsen et al., 2015; Mattan et al., 2016), supporting the use of the VPT as a measure of social cognition.

In the VPT (Samson et al., 2010), participants viewed 96 pictures of a room with an avatar in it and were asked to verify the number of circles visible on the side walls from either their own or the avatar's perspective. Before each picture was shown, a fixation cross appeared for 750 ms. After 500 ms, a cue appeared for 750 ms signaling participants to either verify their own perspective (YOU) or the perspective of the avatar (HE/SHE). 500 ms later, a number cue between 0 and 3 would appear for 750 ms. Lastly, the picture appeared on the screen. The participant's task was to verify whether the number cue matched the number of circles on the wall as visible from the perspective that was cued, i.e., their own perspective (YOU) or the perspective of the avatar (HE/SHE). Crucially, on half of the trials the number of circles visible from the participant's perspective was identical to the number of circles visible from the avatar's perspective (CONGRUENT), but on the other half of the trials a different number of circles would be visible from the different perspectives (INCONGRUENT). Participants used the mouse to indicate whether the number cue matched the number of circles seen from the cued perspective (MATCH; index finger) or not (MISMATCH; middle finger). If no answer was given within 2,000 ms, the next trial would start. Feedback was given after every trial. The pictures were presented in two blocks. Perspective, congruence, and correct response were counterbalanced. Six practice trials were presented at the start of the task, which could be repeated until the participant felt comfortable with the procedure.

Spontaneous Theory of Mind Protocol

Finally, the Spontaneous Theory of Mind Protocol (STOMP; Rice and Redcay, 2015) was included as a promising new measure that may be sensitive to individual variation among healthy adults (Rice and Redcay, 2015; Warnell and Redcay, 2019). Scores on this measure reflect a spontaneous tendency to mentalize when describing the events in two naturalistic videos and have been found to correlate with individual variability in cortical thickness of brain areas related to theory of mind in a neurotypical population (Rice and Redcay, 2015).

For the STOMP task (Rice and Redcay, 2015) participants viewed two silent videoclips taken from existing movies that are centered around complicated social interactions, and were then asked to describe what they had just seen in seven to ten sentences. One videoclip was a 2-min excerpt from the movie *John Tucker Must Die*, in which a girl comes back from a date with a boy whom she has to distract by pretending to flirt with him, so that her friend, who has been secretly following their date by hiding in his car, can escape. The other videoclip was a 3-min excerpt from the movie *Rear Window*, in which a woman is looking for something in an apartment, while being watched by the neighbors across the street, when the owner of the house comes home. Participants saw both videoclips in a random order.

Measures of Reading-Related Individual Differences

As we wanted to control for a possible confounding effect of print exposure, a Dutch version (Koopman, 2015) of the Author Recognition Test (ART; Stanovich and West, 1989) was used as an implicit measure of print exposure: participants were presented with a list of 30 real author names and 12 foils, and were asked to select the names of authors they knew.

Shallow narrative comprehension was measured using three multiple choice questions with four response options each (see **Supplementary Materials**) to check whether participants paid enough attention during reading. All participants scored above chance on these questions, hence, no data was excluded based on the comprehension questions.

Data Recording and Stimulus Presentation

During reading, eye movements were recorded with a desktop-mounted EyeLink 1,000 plus eye tracker, recording at 1000 Hz. A head and chinrest were used to reduce head movements. For most participants the dominant eye was tracked, unless this lead to noisy signal, in which case the other eye was tracked (∼15% of participants).

The narrative was presented using SR Research Experiment Builder. The narrative was divided into 56 sections that fit on the screen and contained between 42 and 151 words ($M = 90.66$, $SD = 25.02$). Most sections contained exactly one paragraph of the 64 paragraphs that made up the narrative, but in some cases the sections contained more than one paragraph, and/or a section break had to be inserted between sentences belonging to the same paragraph. The text was presented in black letters, set in 16 points Times New Roman, on a white page with 120 pixel margins on all sides and double line spacing on a BenQ XL 24020T 24″ LED screen (resolution: 1,024 × 768, 32 bits per pixel). Participants were seated 108 centimeters (42.52 inches) from the screen. The eye tracker was calibrated and validated on a 9-point grid until the largest difference between any target point and computed fixation position was $<1^{\circ}$. A drift check and, if necessary, drift correction took place after every seven slides. At the start of each section a fixation cross marked the position of the first word for 1,000 ms. Participants used the space bar to go to the next section of the text. It was not possible to go back to a previous section.

All questionnaire-based measures (i.e., IRI, ART, and shallow comprehension) and the STOMP were administered digitally in Qualtrics (Provo, UT). The Multifaceted Empathy Test was presented with E-prime (version 2.0; Schneider et al., 2002), using the keyboard (numbers 1 through 9) to record responses. The Visual Perspective-taking Task was presented with DMDX (Forster and Forster, 2003), using a Logitech G502 HERO mouse with a polling frequency of 1,000 Hz to record reaction times.

Procedure

The experiment took place in the Humanities Lab of Radboud University. Upon entering the lab, participants signed for informed consent. Then, participants filled in the IRI and ART questionnaire as well as two other questionnaires not relevant

to the purposes of the current study on a laptop. After that, participants were tested on their eye dominance, and received instruction on the eye-tracking part of the experiment, which took place in a soundproof booth. After calibration of the eye tracker, participants read the narrative at their own pace while one eye was being tracked. After having finished reading, the participants completed the MET and VPT on the same computer in the soundproof booth. Then, the participants moved to the laptop outside the booth to complete the STOMP and the measure of shallow comprehension, as well as one other question not relevant for the purposes of the current study. Finally, participants were debriefed about the goal of the experiment and compensated for their time. Participants took between 60 to 90 min to complete the entire experiment. As described above, three participants were excluded because they were not able to finish the experiment within the available time.

Data Analysis

Pre-processing of Eye-Tracking Data

Eye-tracking data were pre-processed in RStudio using popEye (Schroeder, 2019). PopEye is an R package that can be used to clean, pre-process, and analyze data from reading experiments. The default parsing algorithm from EyeLink was used for the parsing of fixations, saccades, and blinks. During the first stage of data pre-processing, fixations < 80 ms were merged with any longer fixations within a 1-letter distance. In the second stage, fixations < 40 ms were merged with any longer fixations within a 3-letter distance. Fixations that were more than 20% away from the text area were removed. Fixations were automatically aligned on the vertical axis to the lines of the text using the SpakovII algorithm (Špakov et al., 2019).

After the automatic pre-processing, all sections from all participants were inspected visually to check the quality of the automatic vertical alignment. If the automatic alignment of a section was incorrect because the underlying data were too noisy (e.g., horizontal drift) the section was rejected (i.e., removed from all analyses). If more than 30% of the sections of a participant had to be rejected, that participant was excluded. This led to the exclusion of 21 participants (see above). Of the remaining included participants, 317 sections (6.29%) had to be removed on this ground. If the automatic alignment of a section was incorrect but the quality of the underlying data was sufficient, the same pre-processing steps described above were applied again except this time outliers were not removed and vertical alignment was done manually. That is, fixations were visualized per section and for each sequence of fixations it was determined to which line the sequence belonged. This was done for 705 (13.99%) sections. For the remaining 4,018 (79.72%) sections, the automatically pre-processed and aligned data was of sufficient quality. After pre-processing, data from at least 40 of the original 56 sections was available for each participant ($M = 52.48$, $SD = 4.14$).

From the pre-processed data, eye-tracking measures were calculated. In line with recommendations by Kliegl and Laubrock (2017), Orquin and Holmqvist (2018), and von der Malsburg and Angele (2017), we decided against analyzing all of these measures, as this would greatly increase the risk of a Type-I error. Instead, we chose to focus on a small number of measures that covered both early and late processing: skip rate, gaze duration, and rereading rate. Skip rate, a binary variable that indicates whether a word has been fixated at any point during reading (skip rate $= 0$) or not (skip rate $= 1$), is usually associated with low-level word characteristics such as word length and word frequency (Inhoff and Radach, 1998; Brysbaert et al., 2005). However, it has also been found to be related to word predictability and context constraints (Brysbaert et al., 2005), making it an interesting candidate for our study, as viewpoint characteristics of words are both a lexical as well as a contextual phenomenon. Moreover, skip rate has been found to vary between readers (Faber et al., 2020), making it an interesting measure to detect individual differences.

Gaze duration reflects the total duration of fixations made on a word when it is first encountered and has been associated both with "later stages of word processing" (Radach and Kennedy, 2013, p. 431), as well as the "upper bound of early processing" (Kliegl and Laubrock, 2017, p. 77). As such, gaze durations might reflect the possible interaction between lexical characteristics (such as the viewpoint marker categories) and higher level processes (such as social-cognitive abilities). Moreover, gaze duration has often been found to be sensitive to individual differences between readers during narrative reading (van den Hoven et al., 2016; Mak and Willems, 2018; Vries et al., 2018; Eekhof et al., 2021).

Finally, rereading rate is a measure of late processing and reflects whether a word has been fixated again after the first run of reading (rereading rate $= 1$) or not (rereading rate $= 0$). The fact that this measure has been described as being relevant for cognitive processes that take place at the discourse level of texts (Rayner and Liversedge, 2011) makes it especially interesting for our study, as engaging with characters' viewpoints takes place at the discourse level as well.

In keeping with cautions expressed by Orquin and Holmqvist (2018), and Rayner and Liversedge (2011), we do not make direct qualitative assumptions about the connection between these eye-tracking measures and the exact linguistic or cognitive processes that they may reflect. However, in line with previous studies, we do assume that decreased skip rates and longer gaze durations reflect slower processing, potentially induced by processing difficulties (see e.g., Ashby et al., 2005; Rayner et al., 2011; Slattery and Yates, 2018; Gordon et al., 2020; Hessel and Schroeder, 2020). Rereading rate is relatively understudied, but Hessel and Schroeder (2020) found that words that were inconsistent with the context were reread more often, suggesting that increased rereading rate also reflects processing difficulties.

As a final cleaning step during pre-processing, gaze durations more than 3 standard deviations away from the subject-specific means were removed from all analyses. In addition, data from the first word of each section were removed from all analyses for each of the three measures. Function words were disregarded from all analyses, except function words that were part of a lexical unit that was coded as a viewpoint marker during application of the ViewPoint Identification Procedure. After data cleaning, content words had a mean skip rate of 0.27 ($SD = 0.44$), a mean gaze duration of 244.83 ms ($SD = 103.42$ ms), and a mean re-reading rate of 0.21 ($SD = 0.41$).

TABLE 2 | Descriptive statistics for measures of social-cognitive abilities and reading-related individual differences.

Measure	*M* (*SD*)	Cronbach's α
Interpersonal Reactivity Index—Perspective Taking	4.98 (0.85)	0.75
Interpersonal Reactivity Index—Personal Distress	3.57 (0.86)	0.74
Interpersonal Reactivity Index—Fantasy	5.04 (1.09)	0.85
Multifaceted Empathy Test—Emotional Empathy	5.02 (1.25)	0.95
Visual Perspective-taking Task—Altercentric Intrusion (ms)	28.80 (83.05)	
Visual Perspective-taking Task—Egocentric Intrusion (ms)	86.90 (85.20)	
Spontaneous Theory of Mind Protocol (%)	36.17 (10.04)	
Author Recognition Test	6.61 (3.28)	

Interpersonal Reactivity Index scores could vary between 1 and 7 for all subscales, scores on the Multifaceted Empathy Test could vary between 1 and 9, Altercentric Intrusion (Visual Perspective-taking Task) varied between −279.71 and 269.69 ms, Egocentric Intrusion (Visual Perspective-taking Task) varied between −77.49 and 305.99 ms, scores on the Spontaneous Theory of Mind Protocol could vary between 0 and 100, and scores on the Author Recognition Test could vary between −12 and 30.

Because we wanted to control for possible confounding effects of word length and word frequency, all words were annotated for the number of letters and lemma frequency, taken from the SUBTLEX-NL corpus (Keuleers et al., 2010).

Statistical Analyses

Statistical analyses were performed in RStudio (RStudio version 1.3.959, R version 4.0.0; R Core Team, 2020). We calculated mean scores per participant for the four subscales of the IRI. Reliability was acceptable for all subscales (see **Table 2**), except the Empathic Concern subscale ($\alpha = 0.69$). Consequently, the Empathic Concern subscale was not included in the analyses.

ART scores were calculated by taking the number of correctly identified author names and subtracting the number of wrongly identified names (see **Table 2**).

Emotion recognition scores for the Multifaceted Empathy Test were calculated by adding up the number of correct answers per participant for the emotion recognition trials. However, reliability turned out to be unacceptable ($\alpha = 0.43$). As reliability did not increase to be above 0.70 even after dropping half of the items, we decided to exclude this measure from the analyses. The reliability of the emotional empathy trials of the Multifaceted Empathy Test, on the other hand, was excellent ($\alpha = 0.95$). Scores per participant were calculated by averaging over the 40 items (see **Table 2**).

In line with Samson et al. (2010), we only analyzed data from matching trials (i.e., trials in which the number of circles visible from the cued perspective matches the number cue) and correct trials (i.e., trials with incorrect responses were excluded) of the Visual Perspective-taking Task. Egocentric Intrusion was calculated by subtracting the mean response time for congruent other-trials from incongruent other-trials per participant. As such, the measure reflects the extra time needed to take up the altercentric perspective in the presence of a conflicting egocentric perspective, compared to when the altercentric perspective is congruent with the egocentric perspective. High scores on this measure thus indicate a poor ability to separate the two different perspectives and suppress the egocentric perspective in favor of the altercentric perspective. Altercentric Intrusion was calculated by subtracting the mean response time for congruent self-trials from incongruent self-trials per participant. As such, the measure reflects the extra time needed to take up the egocentric perspective in the presence of a conflicting altercentric perspective, compared to when the egocentric perspective is congruent with the altercentric perspective. High scores on this measure thus indicate a poor ability to separate the two different perspectives and suppress the altercentric perspective in favor of the egocentric perspective. Mean scores for both measures are reported in **Table 2**.

Participants' responses on the STOMP task were chunked by the first author based on Rice and Redcay's (2015) procedure of dividing sentences into clauses that represent individual units of information (for a full description of the chunking rules, see **Supplementary Materials**). These chunks were then coded by the first author as being either external descriptions (i.e., physical descriptions and descriptions of physical inferences) or internal descriptions (descriptions of emotions, intentions, and mental states), using a translated and enriched version of the original STOMP coding guide that contained definitions, examples, and key words for the two types of descriptions. An independent researcher then coded 20% of the data to assess the reliability of the coding. As inter-rater reliability was almost perfect (93.31%, $\kappa = 0.86$), the codes of the first author were used in further analyses. A STOMP score was calculated for each subject by taking the percentage of internal descriptions per subject. Seven participants indicated that they had seen one of the movies of which the excerpts were taken, in which case the STOMP score was only based on responses to the other excerpt. One participant had seen both movies, and as a result no STOMP score was calculated. Mean scores are reported in **Table 2**.

We used the *lme4* package (Bates et al., 2015) to fit linear mixed models for the continuous eye-tracking data (gaze duration) and generalized linear mixed models with a logit link function for the binary eye-tracking data (skip rate and rereading rate). In addition, we used the *lmerTest* package to estimate degrees of freedom and statistical significance for the linear mixed models (Kuznetsova et al., 2017). Variance Inflation Factors (VIFs) for the models were calculated with a function reported online (R-hack/mer-utils R, 2014). Predictors were scaled and centered for all analyses. In addition, lemma frequency was log-transformed. The eye-tracking data were analyzed at the word level. As the VPIP scores were available on the level of lexical units, these scores were transformed to the word level by giving all words belonging to a single lexical unit the same score.

We used an identical model structure for the analyses of skip rate, gaze duration, and rereading rate: all models included fixed effects of word length (continuous), word frequency (continuous), viewpoint marker category (factor with four levels: non-viewpoint marker, perceptual viewpoint marker, cognitive viewpoint marker, or emotional viewpoint marker), the measures

of social cognition (the three IRI subscales, MET Emotional Empathy scores, Altercentric and Egocentric Intrusions taken from the VPT, and STOMP scores; all continuous), and ART scores (continuous), as well as interaction terms for the two-way interactions between viewpoint marker category and the measures of social cognition and between viewpoint marker category and ART scores. Finally, the models included by-subject random intercepts. Note that we did not add random slopes for viewpoint marker category, as this lead to convergence issues. Hence, the formula for the models was as follows:

$$
\begin{aligned}
&DV \sim \textit{word length} + \textit{word frequency} \\
&+ \textit{viewpoint marker category} * \textit{ART score} \\
&+ \textit{viewpoint marker category} * \textit{IRI Perspective Taking score} \\
&+ \textit{viewpoint marker category} * \textit{IRI Personal Distress score} \\
&+ \textit{viewpoint marker category} * \textit{IRI Fantasy score} \\
&+ \textit{viewpoint marker category} * \textit{STOMP score} \\
&+ \textit{viewpoint marker category} * \textit{MET score} \\
&+ \textit{viewpoint marker category} * \textit{Altercentric Intrusion} \\
&+ \textit{viewpoint marker category} * \textit{Egocentric Intrusion} \\
&+ (1|\textit{subject})
\end{aligned}
$$

We used dummy coding for the categorical predictor viewpoint markers category, using non-viewpoint markers as a reference level. Hence, for the main effect of viewpoint marker category, each level of viewpoint marker category (perceptual, cognitive, emotional) was compared to the non-viewpoint markers. With this type of contrast coding, the intercept represents the mean of the dependent variable for the reference level, i.e., non-viewpoint markers. Similarly, the estimates of the main effects of the other continuous predictors represent the effect estimate for the reference level, i.e., non-viewpoint markers. Estimates for the interactions between the other continuous predictors and the categorical variable viewpoint marker category indicate the difference between the estimate of the effect of the continuous variable for the reference level, i.e., non-viewpoint markers, and the estimate of the effect of the continuous variable for other levels of the categorical variable, i.e., the different categories of viewpoint markers.

The *sjPlots* package (version 2.8.7; Lüdecke, 2021) was used to produce output tables from the linear mixed models.

RESULTS

The descriptive statistics for the measures of social cognition and reading-related individual differences are given in **Table 2**. The descriptive statistics for the eye-tracking measures by viewpoint marker category are given in **Table 3**.

Skip Rate

The estimates for the generalized linear mixed model predicting skip rate are given in **Table 4**. VIFs were below 2 for all predictors. As expected, there was a significant relationship between the control variables word length and word frequency, and skip rate for non-viewpoint markers. An increase in word length decreased the odds of skipping by 0.45 times (i.e., long words were skipped less often) and an increase in word frequency increased the odds of skipping by 1.13 times (i.e., high-frequent words were skipped more often).

There was also a significant relationship between viewpoint marker category and skip rate (see also **Table 3**). Compared to non-viewpoint markers, the odds of skipping perceptual viewpoint markers were increased by 1.12 times compared to non-viewpoint markers (i.e., these markers were skipped more often). On the other hand, the odds of skipping cognitive and emotional viewpoint markers were decreased by 0.71 and 0.88 times, respectively (i.e., these markers were skipped less often).

In addition, there were also significant main effects of IRI Perspective Taking scores, STOMP scores, and Egocentric Intrusion on skip rate for non-viewpoint markers. An increase in IRI Perspective Taking scores increased the odds of skipping non-viewpoint markers by 1.08 times. That is, readers with higher self-reported perspective-taking abilities were more likely to skip non-viewpoint markers. On the other hand, an increase in STOMP scores decreased the odds of skipping non-viewpoint markers by 0.92 times. That is, readers with a higher tendency to spontaneously mentalize, were less likely to skip non-viewpoint markers. Finally, an increase in Egocentric Intrusion also decreased the odds of skipping non-viewpoint markers by 0.92 times. That is, readers with higher Egocentric Intrusion scores, i.e., poor visual perspective takers, were less likely to skip non-viewpoint markers.

Next, we inspected the interactions between specific viewpoint markers and predictors measuring social-cognitive abilities, to see whether there was a difference between the effect of social-cognitive abilities on non-viewpoint markers and the effect of these abilities on specific types of viewpoint markers. In other words, the interactions allowed us to see whether there was a specific effect of certain social-cognitive abilities on the processing of viewpoint markers that surpasses the effect of these abilities on non-viewpoint markers.

There were significant interactions between viewpoint marker category (perceptual viewpoint markers) and both IRI Perspective Taking scores (see **Figure 1**) and IRI Fantasy scores (see **Figure 2**). There was a significantly more positive effect of both IRI Perspective Taking and IRI Fantasy scores on skip rate for perceptual viewpoint markers, compared to non-viewpoint markers. That is, for non-viewpoint markers, IRI Perspective Taking scores had a significantly positive effect and IRI Fantasy scores had a numerically positive, but non-significant effect on skip rate. For perceptual viewpoint markers, however, the effects of these scores were even more positive. In other words, for perceptual viewpoint markers the odds of skipping increased more as a result of being a reader with a high tendency to take the perspective of others than for non-viewpoint markers.

In addition, there was a significant interaction between viewpoint marker category (cognitive viewpoint markers) and Altercentric Intrusion, such that there was a significantly more negative effect of Altercentric Intrusion on skip rate for cognitive viewpoint markers, compared to non-viewpoint markers (see **Figure 3**). That is, for non-viewpoint markers, Altercentric Intrusion had a numerically positive, but non-significant effect

TABLE 3 | Descriptive statistics for the eye-tracking data by viewpoint marker category.

	Mean (*SD*)			Estimated marginal means (*SE*)		
Viewpoint marker category	**Skip rate**	**Gaze duration**	**Re-reading rate**	**Skip rate**	**Gaze duration**	**Re-reading rate**
Non-viewpoint markers	0.27 (0.44)	244.67 (103.44)	0.21 (0.41)	0.24 (0.01)	242.04 (2.99)	0.19 (0.01)
Perceptual viewpoint markers	0.29 (0.45)	238.47 (100.73)	0.20 (0.40)	0.26 (0.01)	238.28 (3.25)	0.19 (0.01)
Cognitive viewpoint markers	0.21 (0.41)	244.59 (101.05)	0.21 (0.41)	0.19 (0.01)	247.38 (3.14)	0.20 (0.01)
Emotional viewpoint markers	0.16 (0.37)	259.72 (110.46)	0.25 (0.43)	0.22 (0.01)	246.76 (3.35)	0.22 (0.01)

on skip rate, but for cognitive viewpoint markers the effect of Altercentric Intrusion was significantly more negative. In other words, for cognitive viewpoint markers the odds of skipping decreased more as a result of being an inflexible perspective taker than for non-viewpoint markers.

To follow up on these significant interactions, we ran two additional models on a subset of the data containing only perceptual viewpoint markers (for the interaction between perceptual viewpoint markers and the two IRI subscales) and a subset of the data containing only cognitive viewpoint markers (for the interaction between cognitive viewpoint markers and Altercentric Intrusion). The first follow-up model predicted skip rate for perceptual viewpoint markers with word length, word frequency, ART score, IRI Perspective Taking score and IRI Fantasy score as predictors, and by-subject random intercepts. This model confirmed that IRI Perspective Taking scores had a significant, positive effect on skip rate, such that an increase in IRI Perspective Taking scores increased the odds of skipping perceptual viewpoint markers by 1.13 times [$SE = 0.05$, CI (1.04–1.24), $z = 2.74$, $p = 0.006$; see **Supplementary Table 3**]. The effect of IRI Fantasy scores was not significant in this model [odds ratio = 1.07, $SE = 0.05$, CI (0.98–1.17), $z = 1.52$, $p = 0.129$; see **Supplementary Table 3**]. Hence, even though the effect of IRI Fantasy scores on skip rate for perceptual viewpoint markers differed significantly from the effect of IRI Fantasy scores on skip rate for non-viewpoint markers, there was by itself no significant effect of IRI Fantasy scores on skip rate when just looking at perceptual viewpoint markers. In other words, although the tendency to take the perspective of others did increase the odds of skipping perceptual viewpoint markers specifically, the tendency to fantasize did not.

The second follow-up model predicted skip rate for cognitive viewpoint markers with word length, word frequency, ART score, and Altercentric Intrusion as predictors, and by-subject random intercepts. This model revealed that there was in fact no significant effect of Altercentric Intrusion on skip rate for cognitive viewpoint markers [odds ratio = 0.95, $SE = 0.05$, CI (0.86–1.04), $z = -1.14$, $p = 0.255$; see **Supplementary Table 4**]. Hence, even though the effect of Altercentric Intrusion on skip rate for cognitive viewpoint markers differed significantly from the effect of Altercentric Intrusion on skip rate for non-viewpoint markers, there was by itself no significant effect of Altercentric Intrusion on skip rate when just looking at cognitive viewpoint markers. In other words, it was not the case that cognitive viewpoint markers specifically were skipped less often by readers who were poor visual perspective takers.

Note that even though there were significant effects of STOMP scores and Egocentric Intrusion on skip rate, there were no significant interactions between any of the viewpoint marker categories and Egocentric Intrusion. Hence, the effects of STOMP scores and Egocentric Intrusion on skip rate held for all content words, and was not specific to any category of viewpoint markers.

All in all, the results showed that perceptual viewpoint markers were skipped more often than non-viewpoint markers, whereas cognitive and emotional viewpoint markers were skipped less often than non-viewpoint markers. Furthermore, we found that STOMP scores and Egocentric Intrusion decreased the odds of skipping (i.e., readers with a high tendency to mentalize and poor visual perspective takers skip less often), but these effects were not specific to viewpoint markers. In addition, IRI Perspective Taking scores increased the odds of skipping (i.e., readers with high self-reported perspective-taking abilities skip more often) in general, and even more so for perceptual viewpoint markers in particular. That is, readers with higher IRI Perspective Taking scores were more likely to skip perceptual viewpoint markers, more so than non-viewpoint marking words in general. Although the effect of IRI Fantasy scores on skip rate was also significantly more positive for perceptual viewpoint markers than for non-viewpoint markers, a follow-up analysis revealed no significant main effect of IRI Fantasy scores on skip rate when just looking at perceptual viewpoint markers. Similarly, the effect of Altercentric Intrusion on skip rate was significantly more negative for cognitive viewpoint markers than for non-viewpoint markers, but a follow-up analysis revealed no significant main effect of Altercentric Intrusion on skip rate when just looking at cognitive viewpoint markers.

Gaze Duration

The estimates for the linear mixed model predicting gaze duration are given in **Table 5**. VIFs were below 2 for all predictors. As expected, there was a significant relationship between the control variables word length and word frequency, and gaze duration for non-viewpoint markers, such that an increase in word length increased gaze duration (i.e., longer words were read slower) and an increase in word frequency decreased gaze duration (i.e., words with a higher frequency were read faster) for non-viewpoint markers.

There was also a significant relationship between viewpoint marker category and gaze duration (see also **Table 3**). Compared to non-viewpoint markers, gaze durations were significantly decreased for perceptual viewpoint markers (i.e., faster reading),

TABLE 4 | Estimates for the generalized linear mixed model predicting skip rate.

Predictors	Odds ratios	*SE*	*CI*	*z*	*p*
(Intercept)	0.49	0.02	0.46–0.52	−20.61	<0.001***
Word length	0.45	0.00	0.44–0.46	−104.14	<0.001***
Word frequency	1.13	0.01	1.12–1.14	22.23	<0.001***
Viewpoint marker category (perceptual)	1.12	0.03	1.06–1.18	4.21	<0.001***
Viewpoint marker category (cognitive)	0.71	0.02	0.68–0.75	−14.09	<0.001***
Viewpoint marker category (emotional)	0.88	0.04	0.81–0.95	−3.16	0.002**
ART score	1.05	0.04	0.98–1.13	1.47	0.141
IRI—Perspective Taking score	1.08	0.04	1.00–1.17	2.05	0.041*
IRI—Personal Distress score	0.98	0.04	0.91–1.05	−0.56	0.573
IRI—Fantasy score	1.01	0.04	0.93–1.09	0.15	0.884
STOMP score	0.92	0.04	0.85–0.99	−2.18	0.029*
MET—Emotional Empathy score	0.96	0.04	0.89–1.05	−0.85	0.396
VPT—Altercentric Intrusion	1.01	0.04	0.94–1.09	0.33	0.741
VPT—Egocentric Intrusion	0.92	0.03	0.86–0.99	−2.23	0.026*
Viewpoint marker category (perceptual) * ART score	0.99	0.03	0.94–1.05	−0.34	0.737
Viewpoint marker category (cognitive) * ART score	1.04	0.02	0.99–1.09	1.62	0.106
Viewpoint marker category (emotional) * ART score	1.04	0.04	0.96–1.12	0.89	0.373
Viewpoint marker category (perceptual) * IRI—Perspective Taking score	1.07	0.03	1.01–1.14	2.21	0.027*
Viewpoint marker category (cognitive) * IRI—Perspective Taking score	0.99	0.03	0.94–1.05	−0.31	0.759
Viewpoint marker category (emotional) * IRI—Perspective Taking score	1.04	0.05	0.94–1.14	0.73	0.464
Viewpoint marker category (perceptual) * IRI—Personal Distress score	1.01	0.03	0.96–1.07	0.51	0.614
Viewpoint marker category (cognitive) * IRI—Personal Distress score	1.00	0.03	0.96–1.06	0.19	0.846
Viewpoint marker category (emotional) * IRI—Personal Distress score	1.01	0.04	0.93–1.10	0.29	0.773
Viewpoint marker category (perceptual) * IRI—Fantasy score	1.09	0.04	1.03–1.17	2.72	0.007**
Viewpoint marker category (cognitive) * IRI—Fantasy score	0.98	0.03	0.93–1.04	−0.68	0.494
Viewpoint marker category (emotional) * IRI—Fantasy score	1.04	0.05	0.94–1.14	0.70	0.484
Viewpoint marker category (perceptual) * STOMP score	1.01	0.03	0.95–1.07	0.19	0.850
Viewpoint marker category (cognitive) * STOMP score	1.02	0.03	0.97–1.08	0.82	0.414
Viewpoint marker category (emotional) * STOMP score	1.00	0.05	0.91–1.09	−0.07	0.947
Viewpoint marker category (perceptual) * MET—Emotional Empathy score	0.97	0.03	0.91–1.04	−0.85	0.395
Viewpoint marker category (cognitive) * MET—Emotional Empathy score	1.01	0.03	0.96–1.08	0.46	0.643
Viewpoint marker category (emotional) * MET—Emotional Empathy score	0.95	0.05	0.86–1.05	−1.02	0.308
Viewpoint marker category (perceptual) * VPT—Altercentric Intrusion	1.01	0.03	0.95–1.07	0.26	0.798
Viewpoint marker category (cognitive) * VPT—Altercentric Intrusion	0.94	0.02	0.90–0.99	−2.30	0.021*
Viewpoint marker category (emotional) * VPT—Altercentric Intrusion	1.04	0.05	0.96–1.14	0.98	0.327
Viewpoint marker category (perceptual) * VPT—Egocentric Intrusion	0.97	0.03	0.91–1.02	−1.24	0.214
Viewpoint marker category (cognitive) * VPT—Egocentric Intrusion	0.99	0.02	0.94–1.04	−0.53	0.597
Viewpoint marker category (emotional) * VPT—Egocentric Intrusion	0.92	0.04	0.84–1.00	−1.91	0.056

All continuous predictors were scaled and centered for analysis. Word frequency was log-transformed for analysis. Dummy coding was used for the categorical predictor Viewpoint Marker Category with non-viewpoint markers as the reference level. Hence, for the main effect of viewpoint marker category, all categories were compared to non-viewpoint markers. The intercept represents the mean odds ratios of skipping for non-viewpoint markers. The estimates of the other main effects represent the effect estimate for non-viewpoint markers. The estimates for the interaction terms represent the difference between the effect estimate for non-viewpoint markers and the effect estimate for that specific viewpoint marker category.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

whereas gaze durations were significantly increased for cognitive and emotional viewpoint markers (i.e., slower reading).

In addition, there were also significant main effects of STOMP and ART scores on gaze durations for non-viewpoint markers. Both an increase in ART and STOMP scores decreased gaze durations. That is, readers with higher ART scores, indicative of print exposure, and readers with higher STOMP scores, indicative of a tendency toward spontaneous mentalizing, fixated non-viewpoint markers for a shorter duration.

Again, we inspected the interactions between specific viewpoint markers and predictors measuring social-cognitive abilities, to see whether there was a difference between the effect of social-cognitive abilities on non-viewpoint markers and the effect of these abilities on specific types of viewpoint markers. In other words, the interactions allowed us to see whether there

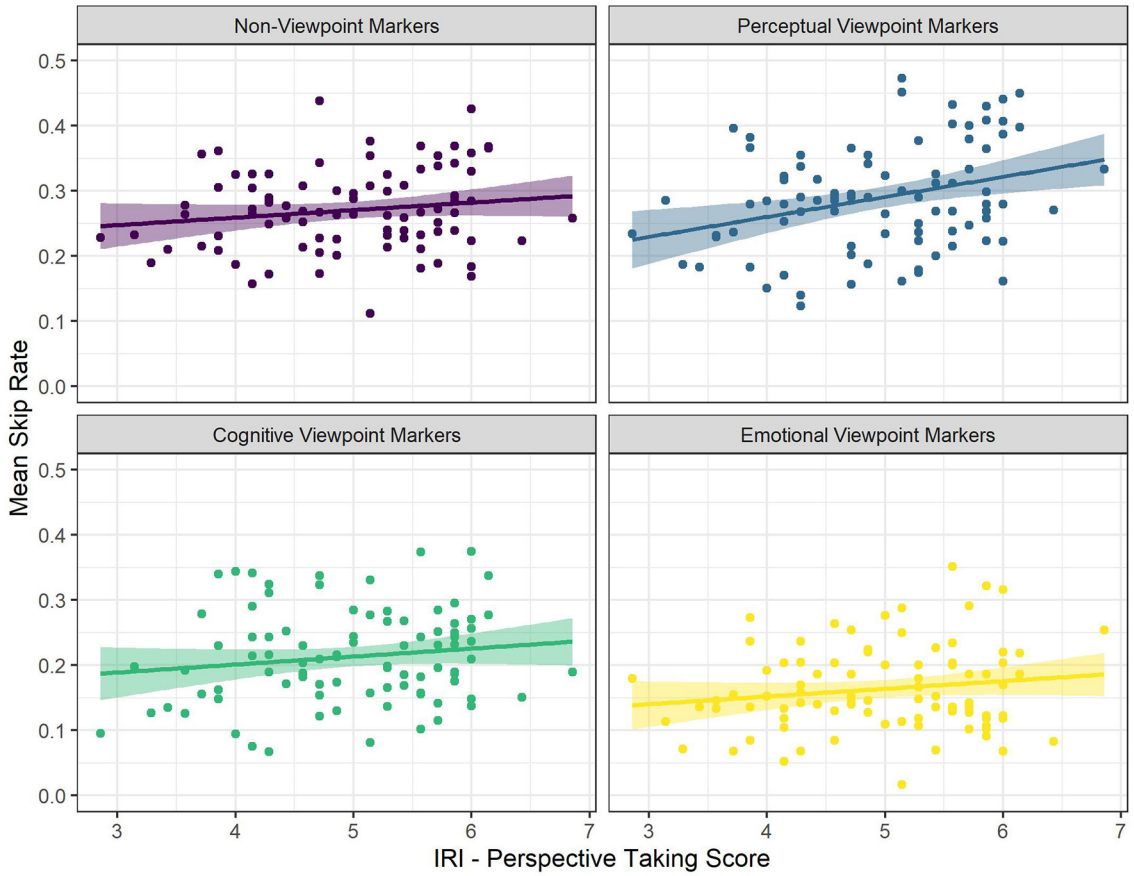


FIGURE 1 | The Relationships Between Mean Skip Rate and IRI Perspective Taking Score for the Different Categories of Viewpoint Markers. Each Dot Represents a Participant.

was a specific effect of certain social-cognitive abilities on the processing of viewpoint markers that surpasses the effect of these abilities on non-viewpoint markers.

There was a significant interaction between viewpoint marker category (emotional viewpoint markers) and the Fantasy score of the IRI (see **Figure 4**). There was a significantly more negative effect of the Fantasy score on gaze duration for emotional viewpoint markers, compared to non-viewpoint markers. That is, for non-viewpoint markers the Fantasy score had a numerically positive, but non-significant effect on gaze duration, but for emotional viewpoint markers the effect of the Fantasy score was significantly more negative. In other words, for emotional viewpoint markers gaze durations decreased more as a result of being a reader with a high tendency to fantasize than for non-viewpoint markers.

To follow up on this significant interaction, we ran an additional model on a subset of the data containing only emotional viewpoint markers, predicting gaze duration for these markers with word length, word frequency, ART score, and IRI Fantasy score as predictors, and by-subject random intercepts. This model revealed that there was in fact no significant effect of IRI Fantasy score on gaze duration when just looking at emotional viewpoint markers [estimate = −3.19, SE = 4.17, CI (−11.36–4.97), t = −0.77, p = 0.443; see **Supplementary Table 5**]. Hence, even though the effect of the IRI Fantasy score on gaze duration for emotional viewpoint markers differed significantly from the effect of the IRI Fantasy score on gaze duration for non-viewpoint markers, there was by itself no significant effect of the IRI Fantasy score on gaze duration for emotional viewpoint markers. In other words, it was not the case that emotional viewpoint markers specifically were read faster by readers with higher IRI Fantasy scores.

Note that even though there were significant effects of STOMP and ART scores on gaze duration, there were no significant interactions between any of the viewpoint marker categories and these scores. Hence, the effect of STOMP and ART scores held for all content words, and was not specific to any category of viewpoint markers.

In sum, perceptual viewpoint markers were read relatively fast, whereas cognitive and emotional viewpoint markers were read relatively slow compared to non-viewpoint markers. In addition, we found that ART and STOMP scores decreased gaze durations overall. Although the effect of IRI Fantasy score on gaze duration was significantly more negative for emotional

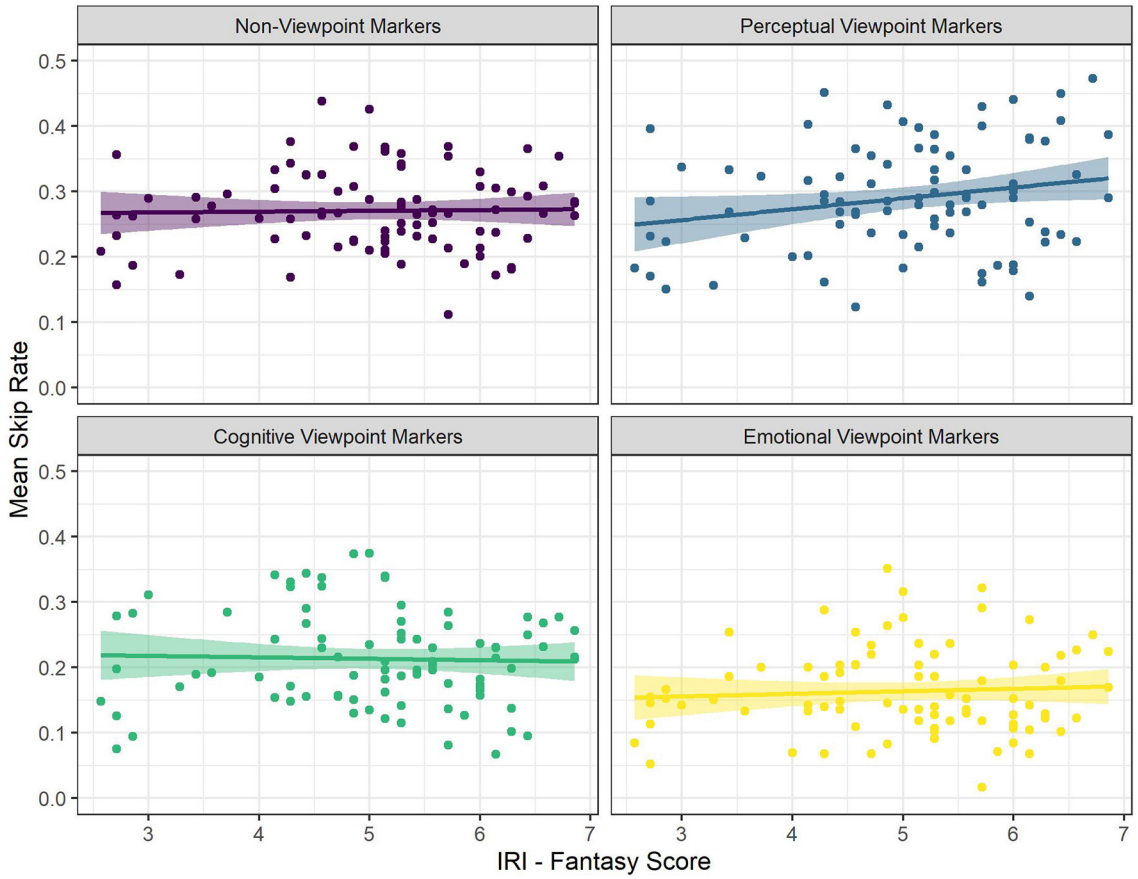


FIGURE 2 | The Relationships Between Mean Skip Rate and IRI Fantasy Score for the Different Categories of Viewpoint Markers. Each Dot Represents a Participant.

viewpoint markers compared to non-viewpoint markers, a follow-up analysis revealed that there was in fact no specific effect of IRI Fantasy scores on gaze duration when just looking at emotional viewpoint markers.

Rereading Rate

The estimates for the generalized linear mixed model predicting rereading rate are given in **Table 6**. VIFs were below 2 for all predictors. As expected, there was a significant relationship between the control variables word length and word frequency, and rereading rate for non-viewpoint markers, such that an increase in word length increased the odds of rereading by 1.14 times (i.e., long words were reread more often) and an increase in word frequency decreased the odds of rereading by 0.93 times (i.e., high-frequent words were reread less often) for non-viewpoint markers.

There was also a significant relationship between viewpoint marker category and rereading rate (see **Table 3**). Compared to non-viewpoint markers, the odds of rereading cognitive and emotional viewpoint markers were increased by 1.07 and 1.16 times, respectively (i.e., these markers were reread more often). There was no significant effect of perceptual viewpoint markers on rereading rate compared to non-viewpoint markers.

In addition, there was also a significant main effect of ART scores on non-viewpoint markers. An increase in ART score increased the odds of rereading non-viewpoint markers by 1.10 times. That is, readers with higher ART scores, indicative of higher print exposure, reread non-viewpoint markers more often.

Again, we inspected the interactions between specific viewpoint markers and predictors measuring social-cognitive abilities, to see whether there was a difference between the effect of social-cognitive abilities on non-viewpoint markers and the effect of these abilities on specific types of viewpoint markers. In other words, the interactions allowed us to see whether there was a specific effect of certain social-cognitive abilities on the processing of viewpoint markers that surpasses the effect of these abilities on non-viewpoint markers.

There were significant interactions between viewpoint marker category (cognitive viewpoint markers) and Egocentric Intrusion (see **Figure 5**), and between viewpoint marker category (emotional viewpoint markers) and ART scores (see **Figure 6**). There was a significantly more positive effect of Egocentric Intrusion on rereading rate for cognitive viewpoint markers, compared to non-viewpoint markers. That is, for non-viewpoint markers, Egocentric Intrusion had a numerically positive, near-significant effect on rereading rate, and this effect was

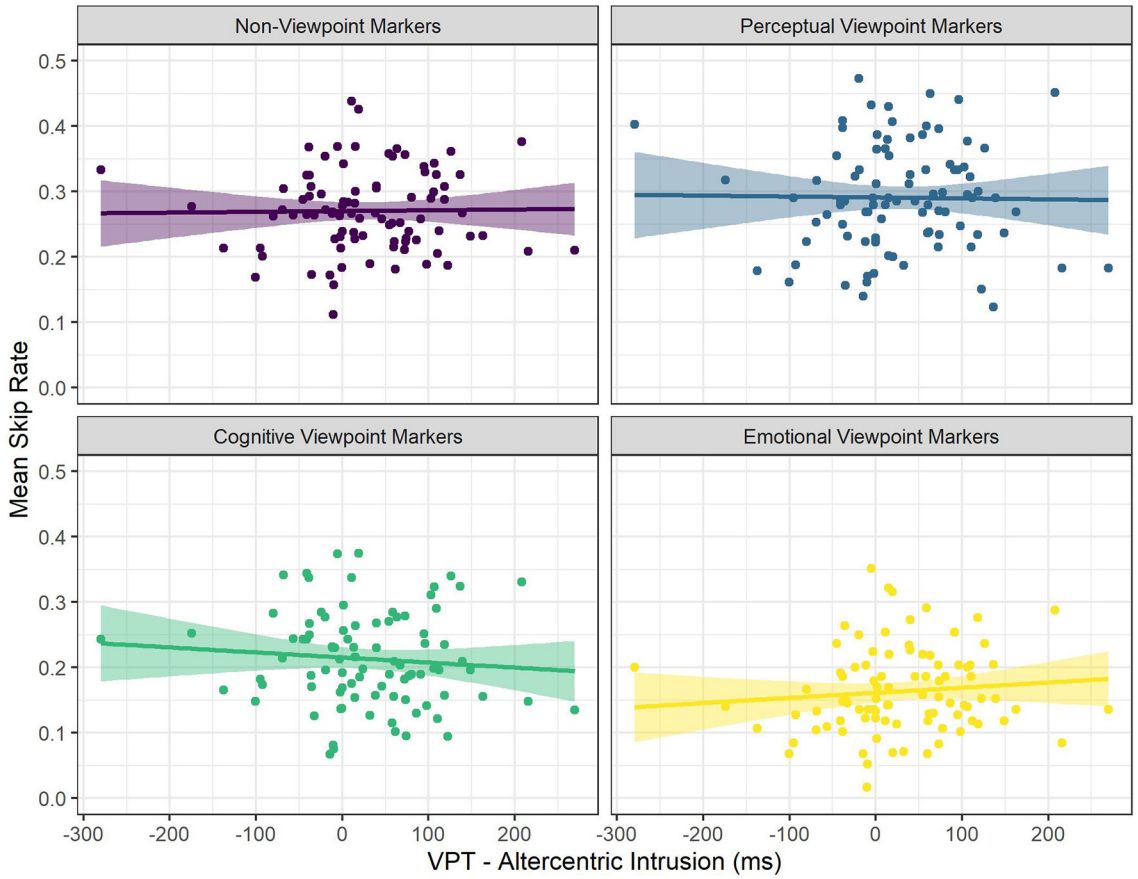


FIGURE 3 | The Relationship Between Mean Skip Rate and Altercentric Intrusion for the Different Categories of Viewpoint Markers. Each Dot Represents a Participant.

significantly more positive for cognitive viewpoint markers. In other words, for cognitive viewpoint markers the odds of rereading increased more as a result of being a poor visual perspective taker than for non-viewpoint markers.

In addition, there was a significantly more negative effect of ART score on rereading rate for emotional viewpoint markers, compared to non-viewpoint markers. That is, for non-viewpoint markers, ART score had a significantly positive effect on rereading rate, but the effect of ART score was significantly more negative for emotional viewpoint markers, essentially meaning that contrary to non-viewpoint markers, rereading rate for emotional viewpoint markers was not affected by ART score.

To follow- up on the first significant interaction, we ran an additional model on a subset of the data containing only cognitive viewpoint markers. This model predicted rereading rate for cognitive viewpoint markers with word length, word frequency, ART score, and Egocentric Intrusions as predictors, and by-subject random intercepts. This model confirmed that Egocentric Intrusion had a significant, positive effect on rereading rate, such that an increase in Egocentric Intrusion increased the odds of rereading cognitive viewpoint markers by 1.14 times [$SE = 0.05$, CI (1.05–1.24), $z = 3.03$, $p = 0.002$; see **Supplementary Table 6**]. In other words, readers with poor visual perspective-taking abilities were more likely to reread cognitive viewpoint markers specifically.

To sum up, cognitive and emotional viewpoint markers were found to be reread more often than non-viewpoint markers, whereas perceptual viewpoint markers did not differ significantly from non-viewpoint markers. In addition, we found that ART score increased the odds of rereading (i.e., readers with higher print exposure reread more often), except for emotional viewpoint markers. Finally, Egocentric Intrusion increased the odds of rereading for cognitive viewpoint markers specifically (i.e., poor visual perspective takers reread cognitive viewpoint markers specifically more often).

The most important results are also schematically summarized in **Figure 7**.

DISCUSSION

In this article we set out to investigate the relationship between the processing of markers of narrative viewpoint, and social cognition. Specifically, we investigated how the linguistic processing of perceptual, cognitive, and emotional viewpoint markers during narrative reading is modulated by individual differences in social-cognitive abilities. We first looked at the

TABLE 5 | Estimates for the linear mixed model predicting gaze duration.

Predictors	Estimates	*SE*	*CI*	*t*	*p*
(Intercept)	230.78	2.99	224.91 – 236.65	77.11	<0.001***
Word length	11.41	0.28	10.86 – 11.96	40.53	<0.001***
Word frequency	−7.04	0.27	−7.56 – −6.51	−26.07	<0.001***
Viewpoint marker category (perceptual)	−3.88	1.33	−6.49 – −1.27	−2.91	0.004**
Viewpoint marker category (cognitive)	5.35	1.05	3.30 – 7.40	5.11	<0.001***
Viewpoint marker category (emotional)	4.79	1.56	1.74 – 7.85	3.08	0.002**
ART score	−10.70	3.04	−16.66 – −4.75	−3.52	0.001***
IRI—Perspective Taking score	−1.69	3.43	−8.40 – 5.03	−0.49	0.624
IRI—Personal Distress score	−0.91	3.11	−7.00 – 5.19	−0.29	0.771
IRI—Fantasy score	0.29	3.59	−6.74 – 7.33	0.08	0.935
STOMP score	−7.38	3.38	−14.00 – −0.75	−2.18	0.032*
MET—Emotional Empathy score	5.11	3.70	−2.14 – 12.36	1.38	0.170
VPT—Altercentric Intrusion	−0.74	3.07	−6.76 – 5.29	−0.24	0.811
VPT—Egocentric Intrusion	2.46	3.08	−3.58 – 8.49	0.80	0.427
Viewpoint marker category (perceptual) * ART score	2.07	1.35	−0.58 – 4.71	1.53	0.125
Viewpoint marker category (cognitive) * ART score	−0.66	1.05	−2.72 – 1.40	−0.63	0.530
Viewpoint marker category (emotional) * ART score	−1.78	1.56	−4.85 – 1.29	−1.14	0.256
Viewpoint marker category (perceptual) * IRI—Perspective Taking score	2.21	1.52	−0.78 – 5.20	1.45	0.147
Viewpoint marker category (cognitive) * IRI—Perspective Taking score	0.70	1.18	−1.61 – 3.00	0.59	0.554
Viewpoint marker category (emotional) * IRI—Perspective Taking score	1.91	1.79	−1.59 – 5.42	1.07	0.285
Viewpoint marker category (perceptual) * IRI—Personal Distress score	−1.66	1.38	−4.37 – 1.04	−1.21	0.228
Viewpoint marker category (cognitive) * IRI—Personal Distress score	−1.35	1.07	−3.45 – 0.75	−1.26	0.208
Viewpoint marker category (emotional) * IRI—Personal Distress score	−1.11	1.61	−4.27 – 2.04	−0.69	0.489
Viewpoint marker category (perceptual) * IRI—Fantasy score	−2.74	1.60	−5.88 – 0.39	−1.71	0.087
Viewpoint marker category (cognitive) * IRI—Fantasy score	−1.41	1.24	−3.84 – 1.02	−1.14	0.256
Viewpoint marker category (emotional) * IRI—Fantasy score	−4.90	1.87	−8.57 – −1.23	−2.62	0.009**
Viewpoint marker category (perceptual) * STOMP score	1.10	1.51	−1.86 – 4.05	0.73	0.468
Viewpoint marker category (cognitive) * STOMP score	−2.14	1.17	−4.44 – 0.16	−1.82	0.069
Viewpoint marker category (emotional) * STOMP score	−2.72	1.77	−6.19 – 0.74	−1.54	0.123
Viewpoint marker category (perceptual) * MET—Emotional Empathy score	−0.14	1.65	−3.38 – 3.10	−0.08	0.934
Viewpoint marker category (cognitive) * MET—Emotional Empathy score	1.47	1.29	−1.07 – 4.01	1.14	0.256
Viewpoint marker category (emotional) * MET—Emotional Empathy score	−1.83	1.95	−5.64 – 1.99	−0.94	0.348
Viewpoint marker category (perceptual) * VPT—Altercentric Intrusion	−0.32	1.39	−3.03 – 2.40	−0.23	0.819
Viewpoint marker category (cognitive) * VPT—Altercentric Intrusion	−0.93	1.09	−3.06 – 1.20	−0.86	0.391
Viewpoint marker category (emotional) * VPT—Altercentric Intrusion	−0.65	1.61	−3.81 – 2.50	−0.41	0.685
Viewpoint marker category (perceptual) * VPT—Egocentric Intrusion	2.09	1.36	−0.58 – 4.76	1.54	0.124
Viewpoint marker category (cognitive) * VPT—Egocentric Intrusion	−1.34	1.07	−3.43 – 0.75	−1.26	0.209
Viewpoint marker category (emotional) * VPT—Egocentric Intrusion	−1.49	1.58	−4.59 – 1.60	−0.95	0.344

All continuous predictors were scaled and centered for analysis. Word frequency was log-transformed for analysis. Dummy coding was used for the categorical predictor viewpoint marker category with non-viewpoint markers as the reference level. Hence, for the main effect of viewpoint marker category, all categories were compared to non-viewpoint markers. The intercept represents the mean gaze duration for non-viewpoint markers. The estimates of the other main effects represent the effect estimate for non-viewpoint markers. The estimates for the interaction terms represent the difference between the effect estimate for non-viewpoint markers and the effect estimate for that specific viewpoint marker category.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

effect of different types of viewpoint markers on eye movements and found diverging patterns of reading behavior for perceptual viewpoint markers on the one hand, and cognitive and emotional viewpoint markers on the other. Crucially, we also found that social-cognitive abilities modulated the effect of different viewpoint markers on eye movements, suggesting that the processing of narrative viewpoint engages these abilities during reading. In what follows, we will first discuss the differences in reading behavior for the three types of viewpoint markers. We will then focus on the role of social-cognitive abilities and the implications of our findings.

Diverging Patterns of Reading Behavior for Different Viewpoint Dimensions

As expected, we found that cognitive and emotional viewpoint markers were skipped less, fixated longer, and reread more often

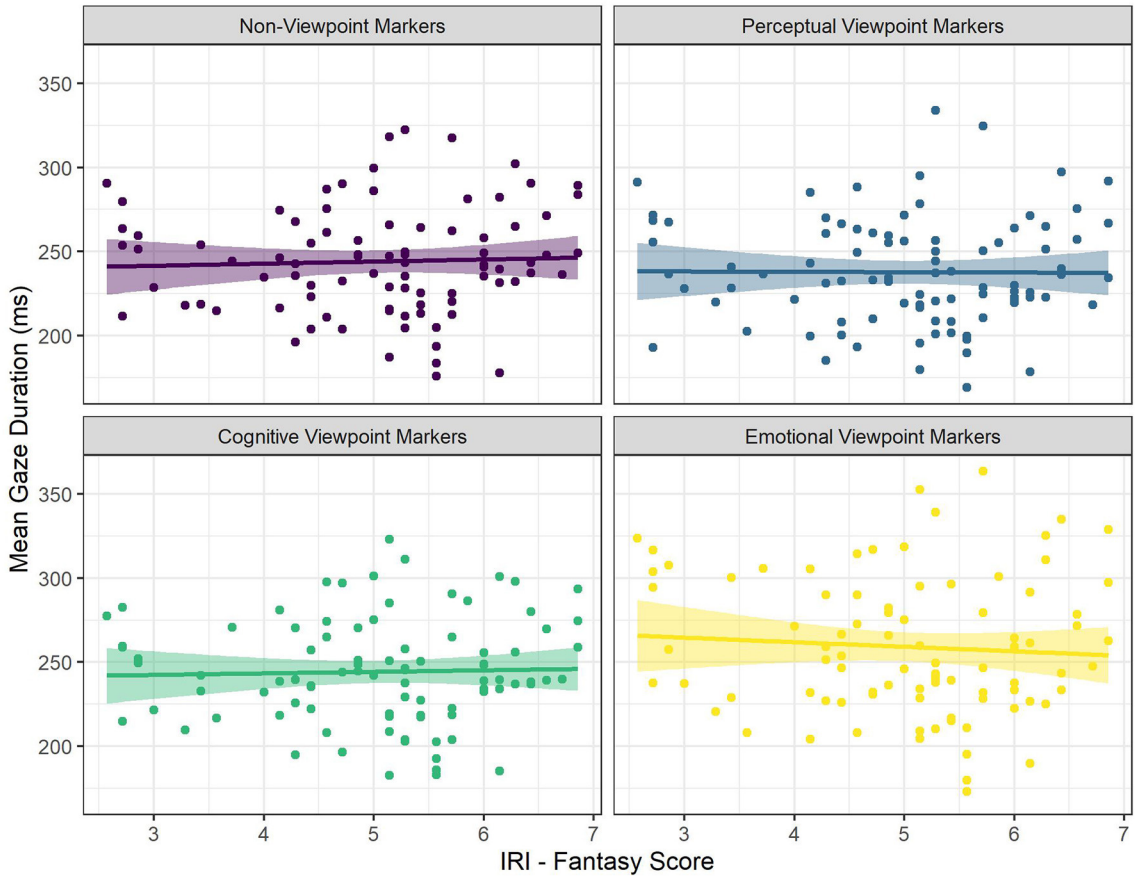


FIGURE 4 | The Relationship Between Mean Gaze Duration and IRI Fantasy Score for the Different Categories of Viewpoint Markers. Each Dot Represents a Participant.

compared to non-viewpoint marking content words. By contrast, however, perceptual viewpoint markers were fixated shorter and skipped more often than other non-viewpoint marking content words, and did not differ in terms of rereading rate from other non-viewpoint marking content words. In other words, whereas cognitive and emotional viewpoint markers were processed relatively slow, perceptual viewpoint marker were processed relatively fast compared to other content words, suggesting that the processing of perceptual narrative viewpoint is linguistically and/or conceptually simpler compared to the processing of cognitive and emotional narrative viewpoint (see also Mak and Willems, 2018).

When looking at the linguistic side of perceptual vs. cognitive and emotional perspective taking, it should first be noted that we controlled for differences in word length and word frequency in our analyses. Hence, the differences between perceptual viewpoint markers on the one hand, and cognitive and emotional viewpoint markers on the other, cannot be explained in terms of these basic linguistic characteristics. However, there might be additional semantic and syntactic differences between these viewpoint dimensions that could lead to differences in processing. For example, perceptual verbs such as *see* and *hear* are often classified as factive verbs (e.g., Givón, 1972), that is, expressing information that is assumed to be true, whereas most cognitive verbs such as *think* and emotional verbs such as *feel* are non-factives. Expressions of perception are thus one-dimensional in that they are implicative of the "truth" of what they express, while expressions of cognition and emotion are semantically multidimensional, referring to the speaker's stance toward the "truth" of what they express. Furthermore, in English, verbs of cognition have been found to be used with a sentential complement (*I think that it's raining*) more often than verbs of perception (*I see that it's raining*), which are more commonly used in simpler syntactic frames, such as in combination with direct objects (*I see rain*; Davis and Landau, 2020). As such, verbs of cognition and emotion might be semantically and syntactically more complex and thus take more time to process. In line with this hypothesis, Davis and Landau (2020) found that regardless of syntactic frame, children between 2 and 5 years old produced more verbs of perception (e.g., *to see, to hear*) than verbs of cognition (e.g., *to know, to think*). This finding is also in line with accounts of theory of mind and language acquisition that argue that children's perceptual understanding develops first, and subsequently serves as a model for understanding

TABLE 6 | Estimates for the generalized linear mixed model predicting rereading rate.

Predictors	Odds ratios	*SE*	*CI*	*z*	*p*
(Intercept)	0.21	0.01	0.19–0.23	−36.63	<0.001***
Word length	1.14	0.01	1.12–1.15	18.25	<0.001***
Word frequency	0.93	0.01	0.91–0.94	−10.96	<0.001***
Viewpoint marker category (perceptual)	1.00	0.04	0.93–1.07	−0.10	0.921
Viewpoint marker category (cognitive)	1.07	0.03	1.01–1.13	2.40	0.017*
Viewpoint marker category (emotional)	1.16	0.04	1.08–1.25	3.89	<0.001***
IRI—Perspective Taking score	0.91	0.04	0.83–1.00	−1.95	0.051
IRI—Personal Distress score	0.97	0.04	0.89–1.05	−0.75	0.455
IRI—Fantasy score	0.97	0.05	0.88–1.07	−0.60	0.548
STOMP score	1.08	0.05	0.98–1.18	1.53	0.127
MET—Emotional Empathy score	1.08	0.06	0.97–1.20	1.47	0.141
VPT—Altercentric Intrusion	1.04	0.05	0.95–1.13	0.80	0.426
VPT—Egocentric Intrusion	1.09	0.05	1.00–1.18	1.90	0.057
ART score	1.10	0.05	1.01–1.20	2.28	0.023*
Viewpoint marker category (perceptual) * IRI—Perspective Taking score	1.01	0.04	0.93–1.10	0.29	0.769
Viewpoint marker category (cognitive) * IRI—Perspective Taking score	1.02	0.03	0.96–1.08	0.62	0.535
Viewpoint marker category (emotional) * IRI—Perspective Taking score	0.98	0.04	0.90–1.07	−0.42	0.672
Viewpoint marker category (perceptual) * IRI—Personal Distress score	0.98	0.04	0.91–1.05	−0.57	0.567
Viewpoint marker category (cognitive) * IRI—Personal Distress score	1.03	0.03	0.97–1.09	1.02	0.309
Viewpoint marker category (emotional) * IRI—Personal Distress score	1.02	0.04	0.94–1.10	0.48	0.630
Viewpoint marker category (perceptual) * IRI—Fantasy score	0.99	0.04	0.92–1.08	−0.13	0.899
Viewpoint marker category (cognitive) * IRI—Fantasy score	0.96	0.03	0.90–1.02	−1.28	0.199
Viewpoint marker category (emotional) * IRI—Fantasy score	1.02	0.05	0.93–1.12	0.48	0.633
Viewpoint marker category (perceptual) * STOMP score	0.97	0.04	0.90–1.05	−0.67	0.500
Viewpoint marker category (cognitive) * STOMP score	0.95	0.03	0.89–1.01	−1.65	0.098
Viewpoint marker category (emotional) * STOMP score	0.97	0.04	0.89–1.06	−0.60	0.549
Viewpoint marker category (perceptual) * MET—Emotional Empathy score	0.97	0.04	0.89–1.05	−0.81	0.417
Viewpoint marker category (cognitive) * MET—Emotional Empathy score	1.02	0.03	0.95–1.09	0.53	0.597
Viewpoint marker category (emotional) * MET—Emotional Empathy score	0.99	0.05	0.90–1.08	−0.31	0.755
Viewpoint marker category (perceptual) * VPT—Altercentric Intrusion	0.95	0.04	0.89–1.03	−1.28	0.199
Viewpoint marker category (cognitive) * VPT—Altercentric Intrusion	1.01	0.03	0.95–1.06	0.18	0.855
Viewpoint marker category (emotional) * VPT—Altercentric Intrusion	1.04	0.04	0.96–1.12	0.87	0.384
Viewpoint marker category (perceptual) * VPT—Egocentric Intrusion	1.00	0.03	0.93–1.07	−0.10	0.919
Viewpoint marker category (cognitive) * VPT—Egocentric Intrusion	1.07	0.03	1.01–1.13	2.42	0.016*
Viewpoint marker category (emotional) * VPT—Egocentric Intrusion	0.99	0.04	0.92–1.07	−0.29	0.768
Viewpoint marker category (perceptual) * ART score	1.01	0.04	0.94–1.08	0.26	0.794
Viewpoint marker category (cognitive) * ART score	1.04	0.03	0.98–1.09	1.31	0.191
Viewpoint marker category (emotional) * ART score	0.92	0.03	0.85–0.99	−2.32	0.020*

All continuous predictors were scaled and centered for analysis. Word frequency was log-transformed for analysis. Dummy coding was used for the categorical predictor viewpoint marker category with non-viewpoint markers as the reference level. Hence, for the main effect of viewpoint marker category, all categories were compared to non-viewpoint markers. The intercept represents the mean odds ratios of rereading for non-viewpoint markers. The estimates of the other main effects represent the effect estimate for non-viewpoint markers. The estimates for the interaction terms represent the difference between the effect estimate for non-viewpoint markers and the effect estimate for that specific viewpoint marker category.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

more abstract mental states such as beliefs (Gopnik et al., 1994).

Nonetheless, in our study, viewpoint markers were not only verbs, but also other types of content words. Another potential linguistic difference between the different types of viewpoint markers is therefore the distribution of word classes. For example, whereas the class of perceptual viewpoint markers contained mostly verbs, emotional viewpoint markers were rarely verbs and more often nouns and adjectives. However, perceptual and cognitive viewpoint markers were very similar in their proportion of different word classes, and yet differed in terms of reading behavior. All in all, more research is needed to understand how perceptual perspective taking, on the one hand, and cognitive and emotional perspective taking, on the other, differ, both linguistically and conceptually.

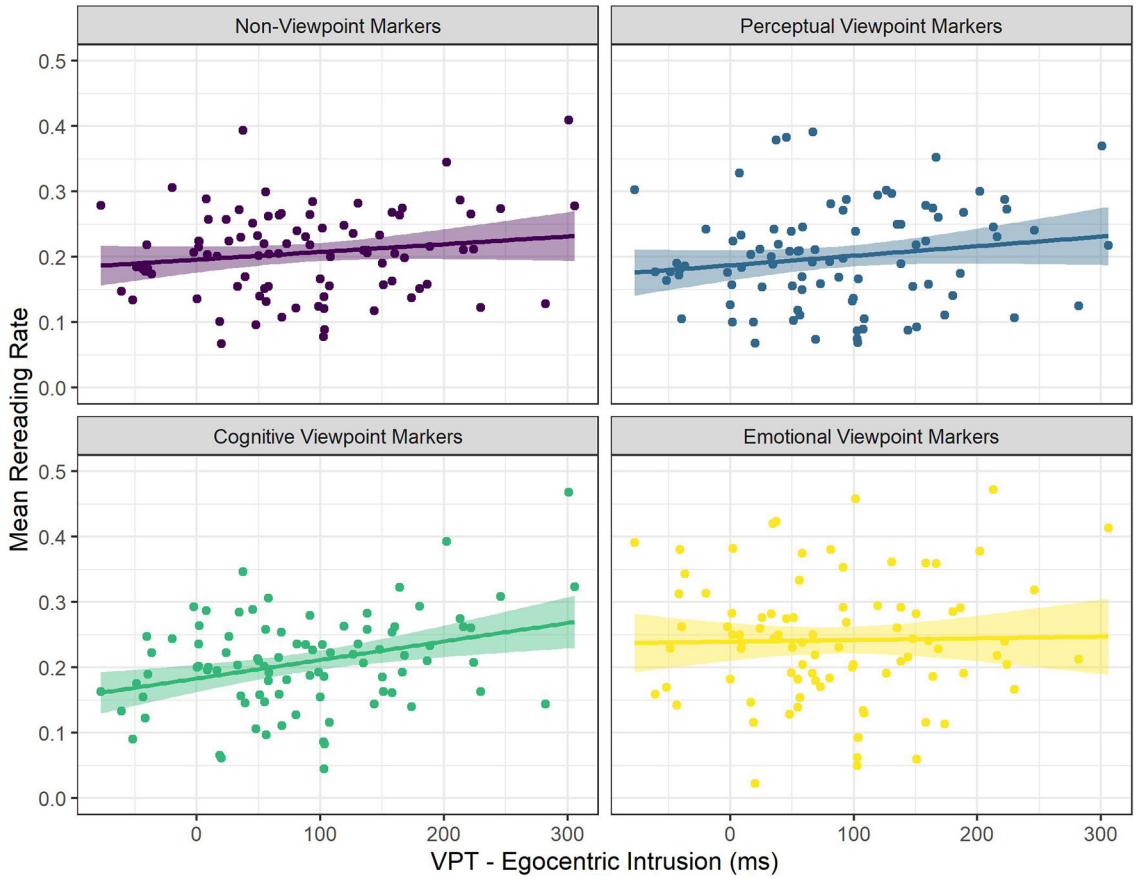


FIGURE 5 | The Relationships Between Mean Rereading Rate and Egocentric Intrusion for the Different Categories of Viewpoint Markers. Each Dot Represents a Participant.

The Role of Individual Differences in Social-Cognitive Abilities

Besides the differences in reading behavior for the different categories of viewpoint markers, we found that individual differences in social-cognitive abilities affected the processing of both words in general and, crucially, perceptual and cognitive viewpoint markers specifically. Firstly, we found that Egocentric Intrusion, a measure derived from the Visual Perspective-taking Task (Samson et al., 2010) that reflects the interference of one's own perspective when taking someone else's perspective, and scores on the Spontaneous Theory of Mind Protocol (Rice and Redcay, 2015), reflecting the spontaneous tendency to mentalize, decreased skip rate. That is, poorer perspective takers and readers with a high tendency to mentalize were less likely to skip words overall. In addition, scores on the Perspective Taking subscale of the Interpersonal Reactivity Index (Davis, 1983) increased skip rate, such that readers who are more likely to take the perspective of others, were more likely to skip words. Finally, scores on the Spontaneous Theory of Mind Protocol also decreased gaze durations for words overall, such that readers with high tendency to mentalize looked at words less long. Although the finding that STOMP scores decrease the odds of skipping words seems to contradict the other findings, the overall picture seems to be that readers with better social-cognitive abilities are faster readers (i.e., more skipping, shorter durations) in general.

The main aim of the study, however, was to see how social-cognitive abilities modulated the linguistic processing of viewpoint markers specifically. We found that readers with higher scores on the Perspective Taking subscale of the Interpersonal Reactivity Index were more likely to skip perceptual viewpoint markers in particular. Moreover, readers who experienced more egocentric intrusion and were thus less flexible perspective takers, were particularly more likely to reread cognitive viewpoint markers. These results cautiously suggest that besides a general facilitatory effect of social-cognitive abilities on linguistic processing, perspective-taking abilities facilitate the processing of at least perceptual and cognitive viewpoint markers. That is, the better these abilities (i.e., more self-reported perspective taking in daily life, more flexible visual perspective taking), the higher the likelihood that readers will skip perceptual viewpoint markers and not reread cognitive viewpoint markers.

What is puzzling, however, is why these two measures of perspective taking affect the processing of perceptual and cognitive viewpoint markers specifically, and not of all types of

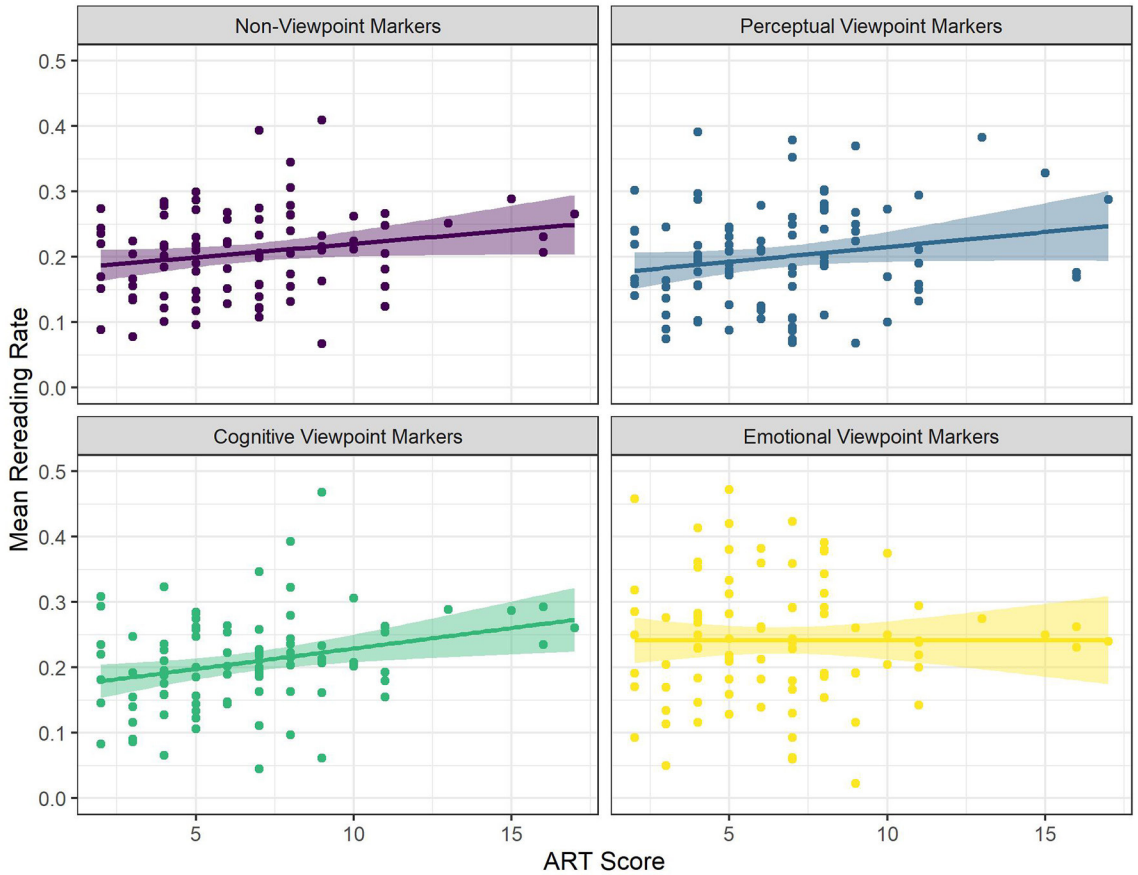


FIGURE 6 | The Relationships Between Mean Rereading Rate and ART Score for the Different Categories of Viewpoint Markers. Each Dot Represents a Participant.

viewpoint markers. This might first and foremost be an issue of power: the stimulus narrative contained more than twice as many cognitive viewpoint markers as emotional viewpoint markers. Hence, future studies could look at more balanced narratives that contain an equal amount of perceptual, cognitive, and emotional viewpoint markers to see whether in such a case social-cognitive abilities affect the processing of all viewpoint dimensions. Alternatively, it could be the case that specific aspects of social-cognitive abilities are in fact related to specific types of narrative viewpoint processing. More detailed studies are needed to further elucidate the details behind these relationships.

All in all, our findings provide a first, modest piece of evidence that processing narrative viewpoint engages social-cognitive abilities, and that a weakness in these abilities thus leads to a delay in processing. As such, our findings corroborate earlier studies that have shown that social-cognitive abilities play a role in narrative processing (e.g., Pelletier and Wilde Astington, 2004; Mason and Just, 2009; Pavias et al., 2016; Atkinson et al., 2017). Moreover, we extend these findings by showing that these abilities are specifically related to the linguistic processing of narrative viewpoint, furthering our understanding of the exact aspects of narrative reading that social-cognitive abilities are implicated in. Our findings also resonate with developmental studies on the relationship between theory of mind and narrative comprehension in general (Pelletier and Wilde Astington, 2004; Pavias et al., 2016; Atkinson et al., 2017), and the acquisition and processing of epistemic markers, verbs of cognition, and verbs of emotion specifically (Moore et al., 1990; Antonietti et al., 2006; Grazzani and Ornaghi, 2012; Ornaghi and Grazzani, 2013). Interestingly, our study reveals that the relationship between social cognition and the processing of viewpoint markers such as verbs of cognition holds into adulthood.

A possible explanation for the finding that social-cognitive abilities facilitate the linguistic processing of narrative viewpoint could be that readers with high social-cognitive abilities have better linguistic or reading skills, for example because they read more often (Mar et al., 2006; Djikic et al., 2013; Mumper and Gerrig, 2017; Lenhart et al., 2020) or because social cognition (partially) relies on language, as is often argued for the relationship between language development and theory of mind development (e.g., Baird and Astington, 2005). Note, however, that we controlled for print exposure, as measured with the Author Recognition Test (Stanovich and West, 1989), in our analyses. Hence, there seems to be a unique contribution of social-cognitive abilities to the linguistic processing of viewpoint, independently of print exposure. To completely rule

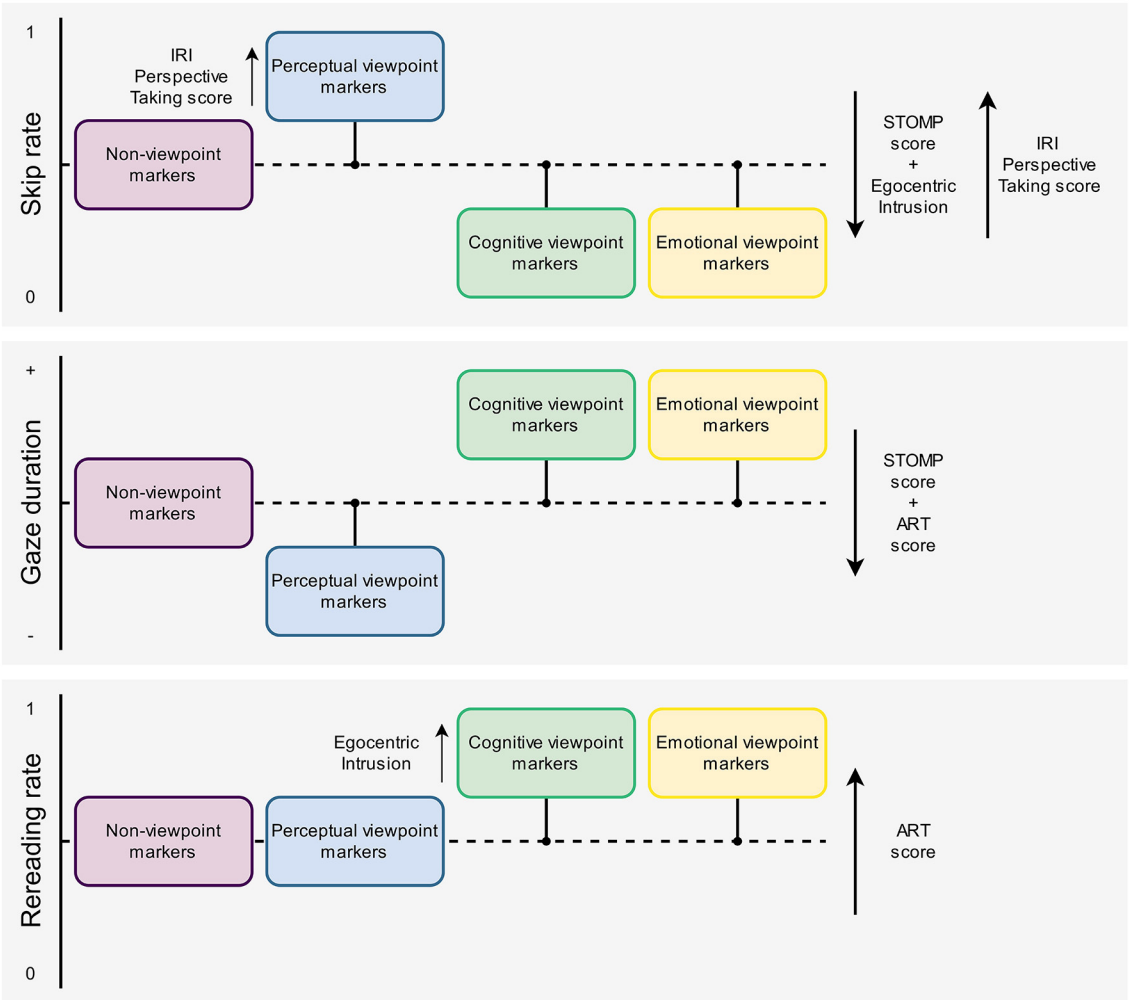


FIGURE 7 | Schematic summary of the results: compared to non-viewpoint markers, perceptual viewpoint markers were skipped more often, whereas cognitive and emotional viewpoint markers were skipped less often. Egocentric Intrusion and STOMP score decreased skip rate overall, and IRI Perspective Taking score increased skip rate overall, and for perceptual viewpoint markers specifically. Compared to non-viewpoint markers, perceptual viewpoint markers were fixated shorter, whereas cognitive and emotional viewpoint markers were fixated longer. STOMP score and ART score decreased gaze durations overall. Compared to non-viewpoint markers, perceptual viewpoint markers did not differ in terms of rereading rate, whereas cognitive and emotional viewpoint markers were reread more often. ART score increased rereading rate overall, and Egocentric Intrusion increased rereading rate for cognitive viewpoint markers specifically.

out this explanation, however, future research would benefit from including more measures of reading habits and skills.

Another conceivable explanation for the facilitatory effect of social-cognitive abilities could be that readers with high social-cognitive abilities process viewpoint markers faster because in light of these abilities viewpoint markers become (partially) redundant. That is, readers who can afford to do so might use their social-cognitive abilities to make sense of the viewpoints of characters, rather than depending too much on the linguistic cues that are provided in the text. In other words, these readers might use their social-cognitive abilities to decrease the demand on linguistic processing. On the other hand, readers with relatively poor social-cognitive skills might need to rely more on explicit markers of viewpoint, leading to slower linguistic processing. In other words, there might be trade-off between using social-cognitive or linguistic means to engage in narrative perspective taking.

This hypothesis is supported by a study on individual differences in perspective shifting: Duff (2018) found that, overall, readers were more likely to take the perspective of a character, rather than a narrator, when interpreting sentences that contained a verb of cognition (e.g., *know* or *believe*) compared to when the sentences contained no such predicate. However, this effect was found to interact with scores on the Autism Quotient questionnaire such that readers with high scores on this questionnaire were most sensitive to the presence of verbs of cognition. That is, unlike other participants, readers with high scores on the AQ took the perspective of the character almost exclusively when a verb of cognition was

present. This suggests that readers with poor social-cognitive abilities are more sensitive to linguistic expressions of perspective.

The explanation that readers with better social-cognitive abilities rely less on textual cues such as viewpoint markers than readers with poor social-cognitive abilities also raises new questions. For example, how do readers with varying levels of social-cognitive abilities process narratives in which viewpoint markers are largely lacking and the contents of characters' minds has to be inferred based on external descriptions of behavior? If the proposed explanation holds, we would expect that readers with poor social-cognitive abilities would be impeded in their attempts to understand or identify with the story characters, because their ability to do so largely depends on explicit markers of viewpoint. On the other hand, readers with high social-cognitive skills would be able to compensate for the lack of explicit viewpoint marking with their social-cognitive abilities. An experiment in which the presence or absence of viewpoint markers is manipulated within narratives could be designed to test this hypothesis.

All in all our results suggest that linguistic markers of narrative viewpoint play a role in engaging social-cognitive abilities during reading. This finding is also of relevance for research on the positive effects of narratives on social cognition. If markers of narrative viewpoint engage social-cognitive abilities, then these abilities might be strengthened through repeated exposure to and engagement by narratives (Mar, 2018). Hence, markers of narrative viewpoint might be an interesting candidate in the search for textual determinants of the social-cognitive potential of narratives (see also Koopman and Hakemulder, 2015). Note that a recent study did not find a difference in the effect of a *single* exposure to a narrative with or without direct access to the inner worlds of protagonists on social-cognitive abilities (internal vs. external focalization; Wimmer et al., 2021). By contrast, reasoning from the present findings, it may be hypothesized that a study that combines such a textual approach with the individual differences approach outlined here, might reveal interesting patterns of results.

In conclusion, our experiment showed that individual differences in social cognition affect the linguistic processing of narratives, and specifically narrative viewpoint. Future research will need to further unravel what this means for narrative processes such as narrative empathy and identification, and, ultimately, the impact of narratives on social cognition.

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 at: www.osf.io/xdjtp.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Assessment Committee Humanities, Radboud University (approval code: 2018-3568). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LE, KK, JS, and RW conceived, designed the experiment, reviewed, edited, and wrote the final manuscript. LE collected and analyzed the data and drafted the first manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by a Veni Grant awarded to KK (Project Number 275-89-038) and a Vidi Grant awarded to RW (Project Number 276-89-007) from the Netherlands Organization for Scientific Research (NWO).

ACKNOWLEDGMENTS

The authors wish to thank Susanne Brouwer for her advice on the statistical analyses and Stefanie Ramachers for her help with translating and data coding.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.698986/full#supplementary-material

REFERENCES

Abbott, H. P. (2008). *The Cambridge Introduction to Narrative*. Cambridge: Cambridge University Press.

Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., and Jacobs, A. M. (2012). The power of emotional valence—From cognitive to affective processes in reading. *Front. Hum. Neurosci.* 6:e00192. doi: 10.3389/fnhum.2012.00192

Antonietti, A., Liverta-Sempio, O., Marchetti, A., and Astington, J. W. (2006). "Mental language and understanding of epistemic and emotional mental states." in *Theory of Mind and Language in Developmental Contexts,* eds. A. Antonietti, O. Liverta-Sempio, and A. Marchetti (Berlin: Springer US), 1–30.

Ashby, J., Rayner, K., and Clifton, C. (2005). Eye movements of highly skilled and average readers: differential effects of frequency and predictability. *Quart. J. Exp. Psychol.* 58, 1065–1086. doi: 10.1080/02724980443000476

Atkinson, L., Slade, L., Powell, D., and Levy, J. P. (2017). Theory of mind in emerging reading comprehension: a longitudinal study of early indirect and direct effects. *J. Exp. Child Psychol.* 164, 225–238. doi: 10.1016/j.jecp.2017.04.007

Baird, J. A., and Astington, J. W. (eds.). (2005). "Introduction: why language matters." in *Why Language Matters for Theory of Mind,* (Oxford: Oxford University Press). Available online at https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780195159912.001.0001/acprof-9780195159912-chapter-1 (accessed September 7, 2021).

Bal, M. (2009). *Narratology: Introduction to the Theory of Narrative, 3rd Edn*. Toronto, ON: University of Toronto Press.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. *J. Child*

Psychol. Psychiatry Allied Discip. 42, 241–251. doi: 10.1017/S0021963001006643

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

Black, J. E. (2019). An IRT analysis of the reading the mind in the eyes test. *J. Pers. Assess.* 101, 425–433. doi: 10.1080/00223891.2018.1447946

Bruner, J. (1986). *Actual Minds, Possible Worlds, Revised Edn*. Cambridge, MA: Harvard University Press.

Brysbaert, M., Drieghe, D., and Vitu, F. (2005). "Word skipping: implications for theories of eye movement control in reading." in: *Cognitive Processes in Eye Guidance,* eds. G. Underwood. (Oxford: Oxford University Press), 53–78.

Davis, E. E., and Landau, B. (2020). Seeing and believing: the relationship between perception and mental verbs in acquisition. *Lang. Learn. Dev.* 17, 26–47. doi: 10.1080/15475441.2020.1862660

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126. doi: 10.1037/0022-3514.44.1.113

De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., and Davis, M. H. (2007). Measuring empathic tendencies: reliability and validity of the dutch version of the interpersonal reactivity index. *Psychol. Belgica.* 47:235. doi: 10.5334/pb-47-4-235

Den Boon, T., and Geeraerts, D. (2005). *Van Dale: Groot Woordenboek van de Nederlandse Taal*. Utrecht: Van Dale Lexicografie.

Djikic, M., Oatley, K., and Moldoveanu, M. C. (2013). Reading other minds: effects of literature on empathy. *Sci. Study Literat.* 3, 28–47. doi: 10.1075/ssol.3.1.06dji

Duff, J. (2018). *Somebody's fool: Theory of mind and the interpretation of epithets* (Amherst, MA: University of Massachusetts Amherst).

Dziobek, I., Preißler, S., Grozdanovic, Z., Heuser, I., Heekeren, H. R., and Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *NeuroImage* 57, 539–548. doi: 10.1016/j.neuroimage.2011.05.005

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., et al. (2008). Dissociation of cognitive and emotional empathy in adults with asperger syndrome using the multifaceted empathy test (MET). *J. Autism Dev. Disord.* 38, 464–473. doi: 10.1007/s10803-007-0486-x

Eekhof, L. S., Kuijpers, M. M., Faber, M., Gao, X., Mak, M., van den Hoven, E., et al. (2021). Lost in a story, detached from the words. *Disc. Proc.* 58, 595–616. doi: 10.1080/0163853X.2020.1857619

Eekhof, L. S., Van Krieken, K., and Sanders, J. (2020). VPIP: a lexical identification procedure for perceptual, cognitive, and emotional viewpoint in narrative discourse. *Open Libr. Human.* 6:18. doi: 10.16995/olh.483

Faber, M., Mak, M., and Willems, R. (2020). Word skipping as an indicator of individual reading style during literary reading. *J. Eye Movem. Res. 13*:3. doi: 10.16910/jemr.13.3.2

Fernández, C. (2013). Mindful storytellers: emerging pragmatics and theory of mind development. First *Language* 33, 20–46. doi: 10.1177/0142723711422633

Foell, J., Brislin, S. J., Drislane, L. E., Dziobek, I., and Patrick, C. J. (2018). Creation and validation of an english-language version of the multifaceted empathy test (MET). *J. Psychopathol. Behav. Assess.* 40, 431–439. doi: 10.1007/s10862-018-9664-8

Forster, K. I., and Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behav. Res. Methods Instr. Comp.* 35, 116–124. doi: 10.3758/BF03195503

Frith, C. D., and Frith, U. (2006). The neural basis of mentalizing. *Neuron* 50, 531–534. doi: 10.1016/j.neuron.2006.05.001

Fuchs, T. (2015). Pathologies of intersubjectivity in autism and schizophrenia. *J. Conscious. Stud.* 22, 191–214.

Givón, T. (1972). "Forward implications, backward presuppositions, and the time axis of verbs." in *Syntax and Semantics Volume,* eds J. Kimball. (Leiden: Brill) 29–50.

Goldman, A., and de Vignemont, F. (2009). Is social cognition embodied? *Trends Cogn. Sci.* 13, 154–159. doi: 10.1016/j.tics.2009.01.007

Gopnik, A., Slaughter, V., and Meltzoff, A. (1994). "Changing your views: how understanding visual perception can lead to a new theory of the mind." in *Children's Early Understanding of Mind: Origins and Development,* eds C. Lewis and P. Mitchell. (Mahwah, NJ: Lawrence Erlbaum Associates, Inc), (157–181).

Gordon, P. C., Moore, M., Choi, W., Hoedemaker, R. S., and Lowder, M. W. (2020). Individual differences in reading: Separable effects of reading experience and processing skill. *Mem. Cogn.* 48, 553–565. doi: 10.3758/s13421-019-00989-3

Grazzani, I., and Ornaghi, V. (2012). How do use and comprehension of mental-state language relate to theory of mind in middle childhood? *Cogn. Dev.* 27, 99–111. doi: 10.1016/j.cogdev.2012.03.002

Healey, M. L., and Grossman, M. (2018). Cognitive and affective perspective-taking: evidence for shared and dissociable anatomical substrates. *Front. Neurol.* 9:e00491. doi: 10.3389/fneur.2018.00491

Hessel, A. K., and Schroeder, S. (2020). Interactions between lower- and higher-level processing when reading in a second language: an eye-tracking study. *Disc. Process.* 57, 940–964. doi: 10.1080/0163853X.2020.1833673

Inhoff, A. W., and Radach, R. (1998). "Chapter 2—definition and computation of oculomotor measures in the study of cognitive processes." in *Eye Guidance in Reading and Scene Perception,* eds. G. Underwood. (Amsterdam: Elsevier Science Ltd), 29–53.

Keuleers, E., Brysbaert, M., and New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behav. Res. Methods.* 42, 643–650. doi: 10.3758/BRM.42.3.643

Kliegl, R., and Laubrock, J. (2017). "Eye-movement tracking during reading." in *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide,* eds A. M. B. De Groot and P. Hagoort. (Hoboken, NJ: Wiley), 68–88.

Koopman, E. M. (2015). Empathic reactions after reading: the role of genre, personal factors and affective responses. *Poetics* 50, 62–79. doi: 10.1016/j.poetic.2015.02.008

Koopman, E. M., and Hakemulder, F. (2015). Effects of literature on empathy and self- reflection: a theoretical-empirical framework. *J. Liter. Theory* 9, 79–111. doi: 10.1515/jlt-2015-0005

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). Lmertest package: tests in linear mixed effects models. *J. Stat. Software.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lenhart, J., Dangel, J., and Richter, T. (2020). The relationship between lifetime book reading and empathy in adolescents: examining transportability as a moderator. *Psychol. Aesth Creat. Arts*. doi: 10.1037/aca0000341

Lüdecke, D. (2021). *sjPlot: Data Visualization for Statistics in Social Science*. doi: 10.5281/zenodo.1308157

Mak, M., and Willems, R. M. (2018). Mental simulation during literary reading: individual differences revealed with eye-tracking. *Lang. Cogn. Neurosci.* 34, 511–535. doi: 10.1080/23273798.2018.1552007

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Ann. Rev. Psychol.* 62, 103–134. doi: 10.1146/annurev-psych-120709-145406

Mar, R. A. (2018). Evaluating whether stories can promote social cognition: introducing the social processes and content entrained by narrative (SPaCEN) framework. *Disc. Process.* 55, 454–479. doi: 10.1080/0163853X.2018.1448209

Mar, R. A., and Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspect. Psychol. Sci.* 3, 173–192. doi: 10.1111/j.1745-6924.2008.00073.x

Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., and Peterson, J. B. (2006). Bookworms versus nerds: exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *J. Res. Pers.* 40, 694–712. doi: 10.1016/j.jrp.2005.08.002

Mason, R. A., and Just, M. A. (2009). the role of the theory-of-mind cortical network in the comprehension of narratives. *Lang. Linguistics Compass.* 3, 157–174. doi: 10.1111/j.1749-818X.2008.00122.x

Mattan, B. D., Rotshtein, P., and Quinn, K. A. (2016). Empathy and visual perspective-taking performance. *Cogn. Neurosci.* 7, 170–181. doi: 10.1080/17588928.2015.1085372

Milligan, K., Astington, J. W., and Dack, L. A. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Dev.* 78, 622–646. doi: 10.1111/j.1467-8624.2007.01018.x

Moore, C., Pure, K., and Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Dev.* 61, 722–730. doi: 10.1111/j.1467-8624.1990.tb02815.x

Mumper, M. L., and Gerrig, R. J. (2017). Leisure reading and social cognition: a meta-analysis. *Psychol. Aesthetics Creat. Arts.* 11, 109–120. doi: 10.1037/aca0000089

Nielsen, M. K., Slade, L., Levy, J. P., and Holmes, A. (2015). Inclined to see it your way: do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process? *Q. J. Exp. Psychol.* 68, 1931–1951. doi: 10.1080/17470218.2015.1023206

Oakley, B. F. M., Brewer, R., Bird, G., and Catmur, C. (2016). Theory of mind is not theory of emotion: a cautionary note on the Reading the Mind in the Eyes Test. *J. Abnorm Psychol.* 125, 818–823. doi: 10.1037/abn0000182

Oatley, K. (1999). Why fiction may be twice as true as fact: fiction as cognitive and emotional simulation. *Rev. Gen. Psychol.* 3, 101–117. doi: 10.1037/1089-2680.3.2.101

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., and Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: toward a brief form for research and applied settings. *Front. Psychol.* 6:e01503. doi: 10.3389/fpsyg.2015.01503

Ornaghi, V., and Grazzani, I. (2013). The relationship between emotional-state language and emotion understanding: a study with school-age children. *Cogn. Emot.* 27, 356–366. doi: 10.1080/02699931.2012.711745

Orquin, J. L., and Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behav. Res. Methods.* 50, 1645–1656. doi: 10.3758/s13428-017-0998-z

Pavias, M., Broek, P., van den Hickendorff, M., Beker, K., and Leijenhorst, L. V. (2016). Effects of social-cognitive processing demands and structural importance on narrative recall: differences between children, adolescents, and adults. *Disc. Process.* 53, 488–512. doi: 10.1080/0163853X.2016.1171070

Pelletier, J., and Wilde Astington, J. (2004). Action, consciousness and theory of mind: children's ability to coordinate story characters' actions and thoughts. *Early Educ. Dev.* 15, 5–22. doi: 10.1207/s15566935eed,1501_1

R Core Team (2020). *R: A LANGUAGE and Environment for Statistical Computing*. R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed September 7, 2021)

Radach, R., and Kennedy, A. (2013). Eye movements in reading: some theoretical context. *Q. J. Exp. Psychol.* 66, 429–452. doi: 10.1080/17470218.2012.750676

Rayner, K., and Liversedge, S. P. (2011). "Linguistic and cognitive influences on eye movements during reading." in *The Oxford Handbook of Eye Movements*. (Oxford: Oxford University Press), 751–766.

Rayner, K., Slattery, T. J., Drieghe, D., and Liversedge, S. P. (2011). Eye movements and word skipping during reading: effects of word length and predictability. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 514–528. doi: 10.1037/a0020990

R-hack/mer-utils R (2014). Available online at: https://github.com/aufrank/R-hacks/blob/master/mer-utils.R (accessed September 7, 2021)

Rice, K., and Redcay, E. (2015). Spontaneous mentalizing captures variability in the cortical thickness of social brain regions. *Soc. Cogn. Affect. Neurosci.* 10, 327–334. doi: 10.1093/scan/nsu081

Ritter, K., Dziobek, I., Preißler, S., Rüter, A., Vater, A., Fydrich, T., et al. (2011). Lack of empathy in patients with narcissistic personality disorder. *Psychiatry Res.* 187, 241–247. doi: 10.1016/j.psychres.2010.09.013

Ryan, M.-L. (2007). "Toward a definition of narrative." in *The Cambridge Companion to Narrative,* eds D. Herman. (Cambridge: Cambridge University Press), 22–36.

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.l Psychol. Hum. Percept. Perform.* 36, 1255–1266. doi: 10.1037/a0018729

Sanders, J. (1994). *Perspective in Narrative Discourse.* Tilburg University. Available online at: https://research.tilburguniversity.edu/en/publications/fac1cba3-7042-4aeb-85b3-8784409c9473 (accessed September 7, 2021)

Sanders, J., and Redeker, G. (1996). "Perspective and the representation of speech and thought in narrative discourse." in *Spaces, Worlds and Grammars,* eds G. Fauconnier and E. Sweetser. (Chicago, IL: University of Chicago Press), 290–317.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-prime 2.0*. Sharpsburg, PA: Psychology Software Tools Inc.

Schroeder, S. (2019). *popEye—An R Package to Analyse Eye Movement Data From Reading Experiments*. Available online at: https://github.com/sascha2schroeder/popEye (accessed September 7, 2021)

Slattery, T. J., and Yates, M. (2018). Word skipping: effects of word length, predictability, spelling and reading skill. *Quart. J. Exp. Psychol.* 71, 250–259. doi: 10.1080/17470218.2017.1310264

Špakov, O., Istance, H., Hyrskykari, A., Siirtola, H., and Räihä, K.-J. (2019). Improving the performance of eye trackers with limited spatial accuracy and low sampling rates for reading analysis by heuristic fixation-to-word mapping. *Behav. Res. Methods.* 51, 2661–2687. doi: 10.3758/s13428-018-1120-x

Stanovich, K. E., and West, R. F. (1989). Exposure to print and orthographic processing. *Read. Res. Q.* 24, 402–433. doi: 10.2307/747605

Tamir, D. I., Bricker, A. B., Dodell-Feder, D., and Mitchell, J. P. (2016). Reading fiction and reading minds: the role of simulation in the default network. *Soc. Cogn. Affect. Neurosci.* 11, 215–224. doi: 10.1093/scan/nsv114

Toolan, M. J. (2001). *Narrative: A Critical Linguistic Introduction*. London: Routledge.

Turner, R., and Felisberti, F. M. (2017). Measuring mindreading: a review of behavioral approaches to testing cognitive and affective mental state attribution in neurologically typical adults. *Front. Psychol.* 8:e00047. doi: 10.3389/fpsyg.2017.00047

van den Hoven, E., Hartung, F., Burke, M., and Willems, R. (2016). Individual differences in sensitivity to style during literary reading: insights from eye-tracking. *Collabra Psychol.* 2:25. doi: 10.1525/collabra.39

van Krieken, K. (2019). "Literary, long-form, or narrative journalism." in *The International Encyclopedia of Journalism Studies,* eds. T. P. Vos and F. Hanusch (Hoboken, NJ: Wiley-Blackwell), 1–7.

van Krieken, K., Hoeken, H., and Sanders, J. (2017). Evoking and measuring identification with narrative characters – a linguistic cues framework. *Front. Psychol.* 8:e01190. doi: 10.3389/fpsyg.2017.01190

van Krieken, K., and Sanders, J. (2021). What is narrative journalism? A systematic review and an empirical agenda. *Journalism* 22, 1393–1412. doi: 10.1177/1464884919862056

van Krieken, K., Sanders, J., and Hoeken, H. (2015). Viewpoint representation in journalistic crime narratives: an analysis of grammatical roles and referential expressions. *J. Pragm.* 88, 220–230. doi: 10.1016/j.pragma.2014.07.012

Vandelanotte, L. (2017). "Viewpoint." in *The Cambridge Handbook of Cognitive Linguistics*, eds B. Dancygier (Cambridge: Cambridge University Press).

von der Malsburg, T., and Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *J. Mem. Lang.* 94, 119–133. doi: 10.1016/j.jml.2016.10.003

Vries, C., de, Reijnierse, W. G., and Willems, R. M. (2018). Eye movements reveal readers' sensitivity to deliberate metaphors during narrative reading. *Sci. Study Literat.* 8, 135–164. doi: 10.1075/ssol.18008.vri

Warnell, K. R., and Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition* 191:103997. doi: 10.1016/j.cognition.2019.06.009

Wimmer, L., Friend, S., Currie, G., and Ferguson, H. J. (2021). reading fictional narratives to improve social and moral cognition: the influence of narrative perspective, transportation, and identification. *Front. Commun.* 5:611935. doi: 10.3389/fcomm.2020.611935

Zunshine, L. (2003). Theory of mind and experimental representations of fictional consciousness. *Narrative* 11, 270–291. doi: 10.1353/nar.2003.0018

Zunshine, L. (2006). *Why We Read Fiction: Theory of Mind and the Novel*. Columbus, OH: Ohio State University Press.

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

Frontiers in Communication

Investigates the power of communication across culture and society

A cross-disciplinary journal that advances our understanding of the global communication revolution and its relevance across social, economic and cultural spheres.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact