COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 47 (2023) e13322

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13322

The Multiple Perspectives Theory of Mental States in Communication

Daphna Heller,[a] Sarah Brown-Schmidt[b]

[a]*Department of Linguistics, University of Toronto*
[b]*Psychology & Human Development, Vanderbilt University*

Received 23 January 2021; received in revised form 11 June 2023; accepted 2 July 2023

Abstract

Inspired by early proposals in philosophy, dominant accounts of language posit a central role for mutual knowledge, either encoded directly in common ground, or approximated through other cognitive mechanisms. Using existing empirical evidence from language and memory, we challenge this tradition, arguing that mutual knowledge captures only a subset of the mental states needed to support communication. In a novel theoretical proposal, we argue for a cognitive architecture that includes separate, distinct representations of the self and other, and a cognitive process that compares these representations continuously during conversation, outputting both similarities and differences in perspective. Our theory accounts for existing data, interfaces with findings from other cognitive domains, and makes novel predictions about the role of perspective in language use. We term this new account the Multiple Perspectives Theory of mental states in communication.

Keywords: Alignment; Assertions; Common ground; Communication; Conversation; Mutual knowledge; Questions; Reference

In dominant accounts of natural language, philosophers have long posited that the notion of mutual knowledge is required to support the most basic aspects of communication (e.g., Lewis, 1969; Schiffer, 1972). Mutual knowledge has standardly been encoded in a representation known as *common ground* (Stalnaker, 1978). While the utility of mutual knowledge was criticized early on (Grice, 1975; Johnson-Laird, 1982; Sperber & Wilson, 1982),

Correspondence should be sent to Daphna Heller, Department of Linguistics, University of Toronto, 100 St. George St., Room 4088, Toronto, ON M5S 3G3, Canada. E-mail: daphna.heller@utoronto.ca

common ground has been widely adopted as the representation that supports communication, in both linguistics (e.g., Beaver, 2001, Farkas & Bruce, 2010, Heim, 1982) and psychology (e.g., Brennan, Galati, & Kuhlen, 2010; Clark, 1996; Clark & Marshall, 1981; Gibbs, Mueller, & Cox, 1988; Nadig & Sedivy, 2002). Some psychologists have rejected the idea that common ground plays a role in real-time language use, arguing that this complex representation is only used as a last resort (e.g., Horton & Keysar, 1996; Keysar, Barr, Balin, & Brauner, 2000). Others have argued that it is possible to derive effects of mutual knowledge through emergent shared representations (e.g., Horton & Gerrig, 2005; Pickering & Garrod, 2004; 2021). Importantly, however, the focus on mutual knowledge as the type of mental state supporting communication has generally remained unchallenged.

The goal of the current paper is to introduce the Multiple Perspectives Theory (MPT), a cognitive architecture of mental states in communication. We motivate our theory by demonstrating that mutual knowledge captures only a subset of the mental states needed to support communication. First, one type of mental state that has received limited attention is private information: while private information of the self has been studied empirically in comparison to mutual knowledge, the assumed private information of the other has been completely overlooked, and theories have not considered how it is estimated, represented, or used. We argue that in order to represent the private information of the other, a theory must include a representation of the other which is separate and independent from the representation of self, motivating a theory of multiple perspectives. Second, by focusing primarily on mutual knowledge, existing theories fail to capture the fact that communication also depends on partners having *different* perspectives. Existing theories have characterized communication as the growing of common ground or as increased alignment between partners; however, no proposal exists for how *differences* in perspective are computed or represented to enable the exchange of information. Importantly, as we elaborate below, the mechanisms proposed for encoding mutual knowledge, which is a type of similarity in perspective, cannot simply be extended to accommodate the representation of perspective *differences*.

We begin by reviewing the relevant background on the role of mutual knowledge in communication, and the theories that have been proposed to support these phenomena (Section 1). In Section 2, we discuss cases of existing empirical evidence from language and memory that reveal that mutual knowledge is only one of the types of mental states needed to support communication. In discussing these cases, we compare potential accounts under existing theories to the account of our proposed architecture, which includes a representation of self, a representation of other, and a cognitive process that compares the two to output similarities and differences in perspective. We further detail our theory in Section 3, and in Section 4, we discuss additional linguistic and conversational phenomena that become relevant under our new cognitive architecture, and identify directions for future research.

1. Mutual knowledge, common ground, and other shared representations

Communicative success has long been argued to depend on the ability to operate on the information that conversational partners can assume they both know, a concept known as

mutual knowledge (cf. *shared*, where sharing is not acknowledged and could be accidental). Building on earlier work by Lewis (1969) and Schiffer (1972), Stalnaker (1978) introduced the notion of *common ground*: the set of propositions that encode mutual knowledge. Against this model of conversation, Stalnaker argues that an assertion should encode information that is *not* already common ground: uttering *I let the dog out* is appropriate when this information is not already taken to be known by both partners; following the assertion, this information would be added to common ground. Clark and colleagues (e.g., Clark & Marshall, 1978, 1981; Clark, 1996) pointed out that information enters common ground not just through language, but also through perception (e.g., if both partners are aware that they both witnessed me letting the dog out). In addition, information can be assumed to be common ground due to background world knowledge: for example, if you find out your interlocutor also attended the University of Toronto, you can assume they also know the names of campus buildings (Isaacs & Clark, 1987).

The ability to determine what information can be classified as mutual knowledge was criticized early on, because this computation was argued to require infinite recursion (e.g., Bach & Harnish, 1979). Specifically, for a proposition p to count as mutual knowledge, it is not sufficient to assume that your conversational partner knows the proposition p: they also need to know that you know that they know p, and you need to know that they know that you know that they know p, and so on, ad infinitum. Clark and Marshall (1978, 1981) suggested that conversational partners do not engage in such computations, and instead use *heuristics* (based on shared physical and linguistic experiences) to draw *inferences* about what information can be considered mutual knowledge (e.g., if we attended a movie together, we can infer the layout of the theater and the names of the characters to be mutual knowledge). Later work demonstrated that conversational partners sometimes produce utterances that are intended to signal to their partner that they have understood and accepted what the partner has said (e.g., "*OK*," "*Yeah*"), and are taking it to be mutual knowledge, a behavior known as *grounding* (Clark & Schaefer, 1989; Clark & Krych, 2004; Brennan, 2005).

In the rest of this section, we first review empirical findings about linguistic phenomena that have been argued to be sensitive to mutual knowledge (Section 1.1); these findings need to be accounted for independent of whether and how mutual knowledge is encoded or represented. We then turn to discuss three prominent theories that provide cognitive mechanisms that can account for these linguistic patterns (Section 1.2), either by directly encoding mutual knowledge (in a representation of common ground), or by positing alternative mechanisms that aim to derive the same effects.

1.1. Mutual knowledge in language use: Empirical findings

The bulk of the data for the role of mutual knowledge in language use comes from the *referential communication* task, in which partners work together to arrange a series of images or objects. Successfully referring to objects, events, and other entities depends on the assumption that the mapping between entities and linguistic forms is the same for both partners. For example, a definite description (e.g., *the dog*) is used to refer to an entity when the speaker can reasonably assume the addressee would consider the referent to be "uniquely identifiable"

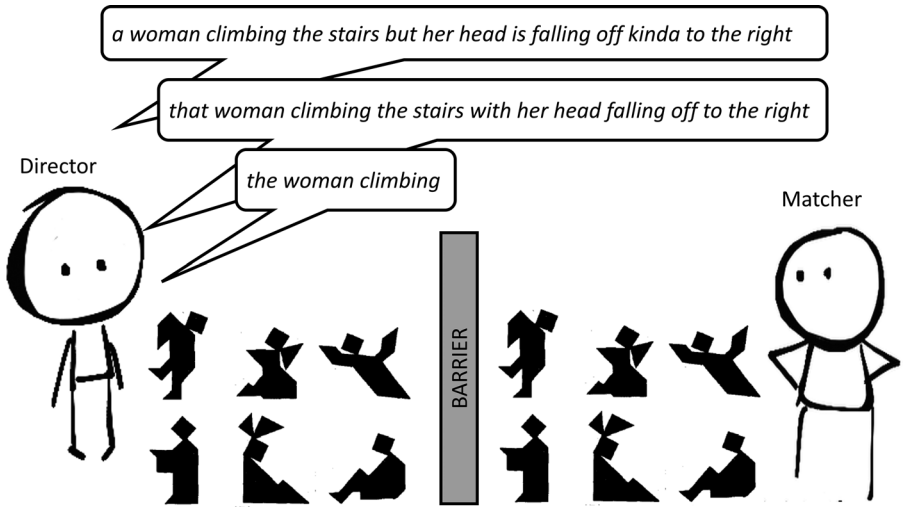

Fig. 1. Example tangrams used in a referential communication task (adapted from Clark and Wilkes-Gibbs, 1986), along with the referring expressions used across several rounds to refer to the top-left tangram.

in the context (Gundel, Hedberg, & Zacharski, 1993). Thus, referring expressions in general, and definite descriptions in particular, are a natural test bed for studying the effects of mutual knowledge.

A classic example of the way referring expressions are thought to make use of mutual knowledge is the way they change over time. Krauss and Weinheimer (1964) noted that repeated reference to the same entity leads to a decrease in the *length* of referring expressions (see also Ariel, 1990), and relate this decrease to Zipf's (1935) generalization about the relationship between frequency and word length (see Krauss & Weinheimer, 1966; Krauss, Garlock, Bricker, & McMahon, 1977). This is illustrated in Fig. 1 which includes a set of tangrams—geometrical shapes that lack conventional labels—along with an example of how referring to the leftmost tangram in the top row shortens over time (see Clark & Wilkes-Gibbs, 1986).

This shortening effect is partner-specific: When speakers switch to a new partner, they tend to use longer labels; however, these longer labels are not as long as in the initial use, indicating some partner-independent effects (Brennan & Clark, 1996; Wilkes-Gibbs & Clark, 1992; Yoon & Brown-Schmidt, 2014; 2018). While the partner-independent shortening of referential forms can be taken to reflect the salience or accessibility of the referent (e.g., Ariel, 1990; Gundel et al., 1993), these accounts are not sensitive to the identity of the partner, and therefore, cannot account for the length difference between an old and a new partner.

What remains stable over repeated reference are the *concepts* encoded in the referring expressions. For example, when a certain abstract figure is referred to using language that conceptualizes it as a woman climbing, this conceptualization continues to be used with the same partner, rather than a re-encoding as, for example, as a dancer (Fleming & Darley, 1991; Mason, 2004; Clark & Schaefer, 1987).

Other past experiences also affect the form of referring expressions (Yoon & Brown-Schmidt, 2014, 2019) as well as their interpretation (Orena & White, 2015; Barr & Seyfeddinipur, 2010). For example, speakers choose a name for an object if they assess that their partner would know this name, whether because the partners learned the names together (Wu & Keysar, 2007; Heller, Gorman, & Tanenhaus, 2012; Gorman, Gegg-Harrison, Marsh, & Tanenhaus, 2013), or because the partner exhibits expertise on the topic (Isaacs & Clark, 1987). A related effect is observed in describing events: speakers are less likely to mention information (such as the instrument used to perform a specific action) when they assess that this information can be inferred by their partner (e.g., because it is typical: Lockridge & Brennan, 2002; cf. Brown & Dell, 1987).

Mutual knowledge is also relevant to listeners' interpretation of referring expressions: listeners are sensitive to whether or not referents are mutually known (Nadig & Sedivy, 2002; Hanna, Tanenhaus, & Trueswell, 2003; Sikos, Tomlinson, Heins, & Grodner, 2019; Rubin, Brown-Schmidt, Duff, Tranel, & Cohen, 2011), how the shared versus privileged status of some referents affects labels for other, mutually known referents (Heller, Grodner, & Tanenhaus, 2008), and what labels have been previously used with a particular partner (Metzing & Brennan, 2003; Brown-Schmidt, 2009b). For example, in a situation where one glass is visible to both the listener and the speaker and a second glass is visible to the listener alone, listeners rapidly identify the mutually visible glass as the intended referent (Nadig & Sedivy, 2002; Hanna et al., 2003; Sikos et al., 2019; Rubin et al., 2011).

In sum, both the production and the comprehension of referring expressions exhibit sensitivity to mutual knowledge (independent of how this is encoded or represented), and this information may be inferred based on different types of cues, including the physical environment, prior conversation, and background assumptions (see Clark, 1996's discussion of physical copresence, linguistic copresence, and community membership).

1.2. Theories of mutual knowledge: Common ground and alternatives

One way to encode mutual knowledge is to represent it as such in a common ground model. On this approach (e.g., Clark, 1996), conversational partners create and continuously update a representation that includes both the discourse record and nonlinguistic information (e.g., the physical environment). Because we learn about mutual knowledge from indirect cues, some have argued that common ground representations are *probabilistic* (e.g., Hanna et al., 2003; Brown-Schmidt, 2012), meaning that information is considered to be common ground with a certain probability—if there is a lot of evidence that something is common ground, the probability will be high; if there is little evidence for common ground, that probability will be low. In contrast, the one-bit model of common ground (Galati & Brennan, 2010; Brennan & Hanna, 2009) encodes this information as binary (e.g., encoding whether the speaker has discussed this topic with this particular addressee before, or not).

The fact that labels are retrieved from an existing representation of common ground, rather than being constructed on the fly, accounts for the repeated use of established conceptualizations, known as *conceptual pacts* (Brennan & Clark, 1996). For example, with repeated reference to the top-left tangram from Fig. 1 with the same partner, the speaker will retrieve

a label from the common ground representation developed earlier with that partner, therefore, reusing the conceptualization of a climbing woman (in a probabilistic representation, the entity-label link needs to be above a certain probability threshold). Note, however, that this explanation does not account for why referring expressions tend to become shorter over time.[1] When the partner is new, there is no existing common ground representation from which labels can be retrieved: this could explain the use of longer descriptions, but not why the same conceptualization is often used with a new partner who lacks common ground for it. In sum, a representation of common ground can account for *some* of the partner-specific effects in the production of referring expressions (but not all).

Because constructing a representation of mutual knowledge requires integrating information from multiple sources and dynamically updating the representation over time, some researchers have proposed that such representations are not used in real-time language processing. Early alternatives were inspired by findings from the judgment and decision-making literature that people favor highly available information when making decisions (Tversky & Kahneman, 1973). For language, the proposal was that, due to its availability in the mind, speakers use their own "egocentric" perspective as a first approximation of mutual knowledge, and only resort to using common ground if communication fails (Horton & Keysar, 1996). Over time, however, empirical evidence has accumulated, demonstrating that mutual knowledge is, in fact, used during real-time language processing (see Brown-Schmidt & Heller, 2018, for a review). This evidence triggered a family of proposals that aim to explain how *apparent* use of mutual knowledge can fall out of simpler, nondedicated mechanisms, and without maintaining an explicit representation of common ground.

One proposal for how mutual knowledge might play a role without calling on specialized common ground representations appeals to ordinary memory (Horton & Gerrig, 2005; 2016). On the ordinary-memory view, one's partner acts as a retrieval cue which can automatically activate shared information via a resonance process. When referring to an entity, speakers access their memory for that entity, which may include a memory trace for an event that involved their conversational partner. For example, when accessing the memory representation for the top-left tangram from Fig. 1, an old conversational partner would serve as a cue for the conceptualization of a climbing woman, whereas a new partner would not. Horton and Gerrig posit that these determinations about what is and is not mutual knowledge can inform language use through either automatic or strategic cognitive processes, leading speakers to utter *the climbing woman* with some partners but not others. Horton and Brennan (2016) consider several aspects of the ordinary-memory view that are in line with the one-bit view (Galati & Brennan, 2010). Most relevant to the current proposal is the notion of metarepresentations: a representation of their partner's representation of something (e.g., my partner cannot see my workspace). Horton and Brennan (2016) suggest that while forming a metarepresentation of a partner's mental state might be time-consuming to calculate initially, once that *metarepresentation* has been calculated, it is accessed in subsequent conversation with relative ease.

A second alternative to the standard common ground model is Interactive Alignment (Pickering & Garrod, 2004; 2013) which explicitly rejected the idea that communication requires modeling the partner's mind, and instead posited that partners achieve aligned

representations through the activation of linguistic forms produced by the self and partner (Pickering & Garrod, 2004). In this theory, conversational partners converge on a conceptualization for an object because they reuse their own and/or their partner's linguistic forms (also see Garrod & Anderson, 1987). For the example in Fig. 1, activation of linguistic forms can explain the reuse of the words "*climbing*" and "*woman*" with the same partner, but this does not account for why labels become shorter over time. Thus, the development of aligned representations can account for a subset of the linguistic phenomena that are attributed to mutual knowledge.

Pickering and Garrod further develop alignment in the more recent Shared Workspace framework (Pickering & Garrod, 2021), where alignment between partners is achieved using information from the "shared workspace"—the publicly accessible communicative behavior and other shared aspects of the context. While this framework also rejects the idea of representing the other's mind, some information about the other is nevertheless encoded via "m-tagging," whereby interlocutors augment their own situation model to reflect what information is shared with a particular partner (see also Westra & Nagel, 2021) and how certain one is that it is shared (akin to Hanna et al., 2003 and Brown-Schmidt, 2012). In this framework, repeated conceptualization with the same partner is a result of a situation model where this conceptualization is m-tagged as shared with the old partner. While such tagging is not available for the new partner, the same conceptualization is predicted to be used because of its availability in one's own situation model. The shortening of referring expressions over time, meanwhile, is attributed to the limited capacity of the "shared workspace" which is said to dictate efficiency. In addition, information from the "shared workspace" is also used to create meta-representations of alignment and misalignment with the partner: a positive response from the partner (e.g., *yeah*, *OK*, *mm*) leads to a meta-representation of alignment, whereas a negative response from the partner (e.g., *eh?*, *what?*, *mm?*) leads to a meta-representation of misalignment (Pickering & Garrod, 2021: section 8.2).

The idea of partner alignment is shared by other researchers. For example, the phenomenon of coordinated eye-gaze in a shared visual world reflects attentional alignment between partners (Richardson, Dale, & Kirkham, 2007; Richardson, Dale, & Tomlinson, 2009; 2012; Dale, Fusaroli, Duran, & Richardson, 2013). Relatedly, neuroimaging findings suggest that persons with similar interpretations of narratives exhibit alignment at the neural level (Nguyen, Vanderwal, & Hasson, 2019). These findings support the idea that language, attention, and even brain activity align as partners converse, consistent with the idea that conversation is based on aligned representations.

More generally, the alternative approaches are similar to common ground theory in their goal of deriving the effects of mutual knowledge. But unlike the common ground theory which posits, for each partner, an explicit representation of mutual knowledge, these other theories achieve this goal without positing a dedicated mechanism that encodes mutual knowledge, or any other direct representation of the other. This difference is reminiscent of the debate in the developmental literature, where Heyes (2014) argues that what appears to be "automatic" reasoning about the belief states of others may instead be captured using domain-general representations of association (see also Westra, 2017; Apperly & Butterfill, 2009). Important for our purposes is the fact that, despite disagreements on mechanisms, the

different theories all share the idea that mutual knowledge is the cognitive basis for successful communication.

2. Going beyond mutual knowledge, or why we need a new theory of mental states

In this section, we pivot to discuss existing empirical findings from language and memory that reveal the limited utility of mutual knowledge as the representational basis for communication, whether directly encoded in a common ground representation (Clark & Marshall, 1978; 1981; Hanna et al., 2003), or derived indirectly via ordinary memory (e.g., Horton & Gerrig, 2005; 2016), aligned situation models (Pickering & Garrod, 2004, 2021), or any other emergent coordinated representations (e.g., Richardson, Dale, & Shockley, 2008; Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Coman, Momennejad, Drach, & Geana, 2016; Fay, Garrod, Roberts, & Swoboda, 2010). This discussion points to three types of mental states that need to be captured by a theory of mental states.

Our first topic is information questions (Section 2.1). Empirical findings concerning the production and comprehension of questions (Section 2.1.1) reveal that the private information of the other is represented and used as part of this basic conversational move: this is the information we assume the other knows without actually having those details ourselves (desideratum #1). In the literature, private information has been discussed primarily in contrast to mutual knowledge, and existing theories fail to detail how private information is calculated, represented, and used. This is especially true as it concerns the private information assumed about the other (we are aware of a single study on this topic, Hawkins, Gweon, & Goodman, 2021). As we demonstrate in Section 2.1.2, private information about the other is estimated and represented in some detail. A second point that arises from the way questions are processed (Section 2.1.1) is the need to represent *differences* between perspectives: this type of information does not have a clear place in theories where the other is only represented through the lens of mutual knowledge. The observation that differences need to be represented alongside similarities (desideratum #2) is novel and has not been previously discussed in the literature. In Section 2.2, we return to the topic of referring expressions, which has been the central empirical phenomenon on which theories of mental states are based (see again Section 1.1). We argue that a closer examination of existing empirical findings from both language comprehension (Section 2.2.1) and language production (Section 2.2.2) point to the need to *separately* represent the perspectives of self and other (desideratum #3); these findings cannot be explained when representing the other through the lens of the self. The need for separate representations of the self and other is further indicated by findings from the memory literature (Section 2.3) that identify asymmetries in representations of the conversation record. Despite ample evidence for these memory asymmetries, they have not been considered in developing the cognitive architecture that supports language use. In sum, these findings point to a theory that must fulfill three desiderata: (1) the (assumed) private information *of the other*, (2) *differences* in perspective between the self and other (alongside similarities which includes mutual knowledge), and (3) *separate* representations of self and other.

Our novel Multiple Perspectives Theory (MPT) is a cognitive architecture that is intended to capture the ways in which mental states are represented and used in communication. Some components of our theory are similar to aspects of other work, whereas other aspects are new. The novel contribution arises from specifying the calculation, representation, and use of mental states that have not been represented in previous theories. Our theory includes three components:

1. A representation of **self**. What do I know that is relevant to the current conversation? We propose that partners only activate representations that are *relevant* to the current exchange (Grice, 1975; Sperber & Wilson, 1986, 1995). Similar to the "egocentric" perspective discussed in some prior work (e.g., Keysar et al., 2000, 2003; Heller, Parisien, & Stevenson, 2016; Mozuraitis, Stevenson, & Heller, 2018), this is a representation of what you know. Different from prior work, we distinguish different types of self-knowledge (knowing what you do know vs. what you don't know), and use this representation in computing similarities and differences with the other.
2. A representation of **other**. What do *I assume* you know that is relevant to the conversation? As we do not have direct access to other minds, creating this representation depends on inferences based on *indirect cues* from multiple sources of information. As with the representation of self, we propose that partners only activate representations of the other that are *relevant* to the current exchange. This representation may seem familiar from other research. For example, in his seminal work on conversational implicatures, Grice (1975) suggests that implicatures are computed relative to the other's intentions (see also Sperber & Wilson's Relevance Theory: 1982, 1986, 1995); Grice's logic has been further formalized within the probabilistic Rational Speech Act framework (RSA: Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013). However, these theories have not generally aimed to account for phenomena that involves different perspectives like those reviewed here. Another line of research that posits a representation of the other is nonlinguistic mentalizing, or the Theory of Mind (ToM) literature outside language: we come back to this research in Section 2.2.3.
3. A cognitive **process** that actively compares the representations of self and other. This is the heart of our theory. We hypothesize that this comparison process occurs continuously over time as speaker meaning is computed. The output of the comparison process is transient, and it may reveal (i) *similarities* between self and other (including mutual knowledge), and (ii) *differences* between self and other.

The need for *separate* representations of the self and other has been recognized in the formal semantics literature since the groundbreaking work of Gunlogson (2003) on declarative questions. Since then, semantic theories have widely incorporated separate representations of the *discourse commitments* of the conversational partners (e.g., Farkas & Bruce, 2010, Faller, 2019 among others), which bear some similarity to our separate representations of self and other. However, these theories differ in their overall goal: while here we focus on mental states, semantic theory aims to model the conversation record, addressing the question of how given utterances change the context. Perhaps more importantly, semantic theories have also maintained a common ground that represents mutual knowledge, which is not part of the

cognitive architecture in the Multiple Perspectives Theory, a point we return to at the end of Section 2. Nevertheless, exploring connections between these literatures is a worthwhile goal for future work.

In discussing the empirical cases, we point out limitations of existing mental state accounts, specifically the standard common ground model of Clark and colleagues, the ordinary memory account (Horton & Gerrig, 2005; 2016), and the shared workspace framework (Pickering & Garrod, 2021), detailing how these cases are accounted for within the Multiple Perspectives Theory.

2.1. Questions

The communicative goal of information questions is to elicit information from the partner. It has long been observed that speakers ask about the information they themselves lack, and they address their question to an addressee they assume has the desired information[2] (Searle, 1969). In the sociologically oriented field of conversation analysis, this is known as determining one's *epistemic status* (Heritage, 2012a,b): whether your partner is more or less knowledgeable, which is crucial in order to decide whether to ask or tell, respectively. The fact that all 803 languages surveyed in the *World Atlas of Language Structures* (WALS) are reported to have questions (Dryer, 2013) reflects how basic this conversational move is in human interaction.

In the case of the Pig with the Hat (Section 2.1.1), we discuss empirical findings showing that speakers ask about the (assumed) private knowledge of the partner, and that listeners expect questions to be about their private knowledge (desideratum #1). The case of the Pig with the Hat illustrates a second point, which has not been previously noted in the literature: that interpreting a question requires not just a focus on one's own private knowledge (which is different from the perspective of the asker), but also the representation of *similarities* in perspective, which includes mutual knowledge. This observation suggests that similarities and differences in perspectives should be available all at once (desideratum #2). In Section 2.1.2, we turn to the case of the Knowledgeable Dentist, where we illustrate the level of detail at which the other's private knowledge needs to be represented (desideratum #1).

2.1.1. Empirical findings about questions: The case of the Pig with the Hat

The case of the Pig with the Hat provides empirical evidence that information questions require modeling and utilizing information that is private to the other. Brown-Schmidt, Gunlogson, and Tanenhaus (2008) examined the production and comprehension of information questions that were uttered in the context of a conversational game. Fig. 2 provides an example of a game board from two perspectives, where the color of the background indicates the shared versus private status of the objects (note that Brown-Schmidt et al. examined both physical displays with objects that were and were not mutually visible, as well as computerized displays like the one in Fig. 2). Let us imagine that one partner, Abbi, sees panel (a): it contains three objects with a white background which indicates that her partner, Ilana, can also see them, four objects with a gray background which indicates that only Abbi herself can see them, and one black square whose contents, she assumes, are only visible to Ilana. Ilana

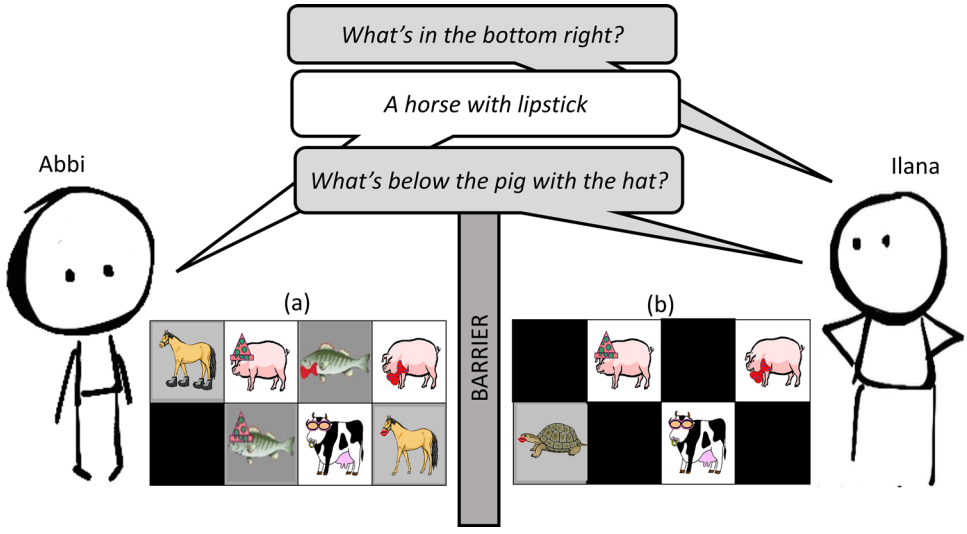

Fig. 2. The game boards used to elicit questions and examine their interpretation from the perspective of Abbi (a) and Ilana (b). The colored backgrounds reflect information status: White background for mutually visible objects, gray background for private objects, and black squares for objects only visible to the partner. All animals had accessories, thus requiring modifiers (e.g., "*with a hat*") to distinguish two of the same kind.

sees panel (b) which has three shared objects, one private object, and four objects visible only to her partner Abbi. In the game, the partners converse in order to figure out whether their combined board follows certain rules (e.g., are there two animals of the same type in adjacent squares?).

To begin, Ilana might ask, *What's in the bottom right?* The question is felicitous because Ilana does not know the answer, and, at the same time, she can use the cues in the display to assume that Abbi *does* know the answer. Empirical findings indicate that speakers consistently inquire about information which they assume to be known to the partner, but not to themselves. If Abbi responded, *A horse with lipstick*, then the existence of the horse in this cubby becomes mutual knowledge. Next, Ilana may ask, *What's below the pig with the hat?* In interpreting this question, Abbi expects Ilana to ask about objects that Ilana does not know about.

Empirical findings show that listeners interpret information questions as inquiring about information that is known to themselves but not the speaker (Nurmsoo & Bloom, 2008; Brown-Schmidt et al., 2008; Brown-Schmidt, 2009a; Brown-Schmidt & Fraundorf, 2015). Specifically, eye-tracking data reveal that as the listener hears *What's below the pig with …* (but has not yet heard the end), listeners direct attention toward the pig with the hat, presumably because below this pig is a fish the asker does not know about (whereas below the pig with the bow is the horse that became mutual knowledge in the previous conversational turn).

Despite the simplicity of this example, the case of the Pig with the Hat illustrates two important points about mental states in communication. First, it provides clear empirical evidence that information questions require representing the (assumed) *private* knowledge *of the*

other (desideratum #1). But what are the cognitive mechanisms that allow speakers to determine that the addressee knows—or is likely to know—the answer to a question? First, we note that a common ground model would only allow a speaker to infer that the answer to the question is not mutually known. But the set of information that is not mutual knowledge is unbounded and undifferentiated, and thus is not sufficient to determine whether the addressee is likely to know the answer; similar reasoning applies to the ordinary memory account which tags shared information in memory. While we acknowledge that neither theory has aimed to account for information questions, the fact that these theories are a nonstarter in trying to model questions reveals the need for a new theory.

The only existing theory to consider questions—albeit briefly—is Pickering and Garrod's (2021) Shared Workspace framework. They discuss a single example (p. 172), proposing that questions are asked when there is a meta-representation of *mis*alignment between partners. Can the questions in the case of the Pig with the Hat be explained using a meta-representation of misalignment? Recall that in the Shared Workspace framework, misalignment is gleaned from information in the "shared workspace"; however, in the case of the Pig with the Hat, there is no point at which the partner responds negatively that could be taken to indicate misalignment. As the "shared workspace" is said to include other contextual information, one may wish to extend their account, suggesting that a meta-representation of misalignment can be gleaned from the properties of the physical display. However, Pickering and Garrod are clear that the "shared workspace" only includes *shared* information, and so it would not include information about those animals that are not mutually visible (see further discussion in Section 2.2.2: the case of the Deceitful Lego). Thus, while meta-representations of misalignment could trigger questions in some situations, this framework does not have the necessary components to capture the perspective *asymmetries* illustrated in the case of the Pig with the Hat, or any other information question beyond the narrow cases in which misalignment is detected based on a partner's negative response. This is because this framework eschews any representation of the other, and so it cannot capture cases where the other is assumed to know information that you do not, which constitutes the informational basis for asking questions (note that an m-tagged situation model only encodes the opposite situation: information you do know and the other does not).

In the Multiple Perspectives Theory, the (assumed) private information of the other is encoded directly as part of the representation of the other, which exists alongside the representation of self; we discuss these representations in detail in Sections 3.1 and 3.2. The question is prompted by the speaker's comparison of their own visual perspective with their partner's perspective which reveals an information asymmetry between them.

The second point illustrated by the case of the Pig with the Hat arises from the fact that while the question word "*what*" delimits the space of answers to the addressee's private knowledge, the definite description "*the pig with the hat*" is nevertheless relativized to the set of referents that are mutually known to both partners. The fact that both phrases occur in the same question demonstrates that listeners cannot simply focus on their own perspective, even for a (relatively) small window of interpreting a single sentence. A strategy of adopting the perspective of the other—either fully, or in a serial adjustment process (Epley, Morewedge, & Keysar, 2004; Epley, Keysar, Van Boven, & Gilovich, 2004; cf., Eyal, Steffel, & Epley,

2018), would also fail here, as this would not allow the listener to find the answer. Instead, what we need is a representation of the similarities in perspective between partners alongside the differences between them (desideratum #2).

Because this point has not previously been made in the literature, we can only speculate on how other theories could be extended to account for it. The standard common ground model, as well as the ordinary memory model, are both dedicated to the encoding of similarities, representing mutual knowledge directly, or tagging information in one's own memory as mutual (this also applies to m-tagging in the Shared Workspace framework). We note that these mechanisms cannot be extended to the representation of perspective differences because the relevant information for which we hold different perspectives is not simply equal to the set of information that is *not* similar. In the Multiple Perspectives Theory, similarities and differences are both calculated by the comparison process that is hypothesized to continuously compare the representations of self and other—we discuss this process in detail in Section 3.3.

2.1.2. Tracking sources of evidence for the other: The case of the Knowledgeable Dentist

This section presents a thought experiment that illustrates how speakers choose an appropriate addressee for a question. This case reveals that conversational partners not only represent *what* private information their partners may have, but, in addition, they keep track of the assumed *source* of this information. This example provides further evidence for the direct encoding of the private information of the other by revealing the level of detail needed to be represented in the (assumed) private knowledge of the other (desideratum #1).

The case of the Knowledgeable Dentist concerns a situation where you take your son, James, to the dentist. There you are, talking to the dentist about James' teeth. When you seek an answer to the question of whether James has brushed his teeth that morning, who do you ask? (Let us assume for now that everyone provides truthful answers). The intuitive answer is clear: you ask James, because you know that he has *direct evidence* for the answer (he was there), whereas the dentist only has *indirect evidence* (she can draw an inference based on the examination). This simple example where speakers are able to choose the best addressee for a question reveals that speakers are tracking potential information sources and using these information sources to construct the representation of the other.

How can we apply the mechanisms of existing theories to account for this case? On the standard common ground model (e.g., Clark & Schaefer, 1989), your past conversation with the dentist—including on the topic of James' oral hygiene—was recorded into the common ground, whereas on the ordinary memory account you retrieve memory traces shared with the potential addressees that relate to the topic (e.g., Horton & Gerrig, 2005); this also applies to the m-tagged situation model of Pickering and Garrod (2021). If we extend the function of these representations to choosing an addressee for a question (which, admittedly, goes beyond their original function), we believe that these views would wrongly point to the dentist as the chosen addressee, because the dentist is the stronger association on the topic of James' oral hygiene. The skeptical reader may counter that, for this topic, you actually have more experience talking to James (because you nag him twice a day to brush his teeth): this would mean that you have more common ground with James or that the associations with James are stronger than the ones with the dentist, and so those accounts would correctly predict

that you would ask James and not the dentist. Be that as it may, these accounts miss the fact that an addressee for a question is not intuitively chosen based on how much *past* experience you share with them, but rather based on who you expect to have the better evidence for the answer. This can be further illustrated by a situation where your neighbor Liz takes James to the dentist: despite never having had a conversation about James' oral hygiene with either James or the dentist (or anyone else, for that matter), the neighbor would intuitively still ask James. In other words, the case of the Knowledgeable Dentist demonstrates the limited utility of mutual knowledge as the basis for mental states in communication, as it is unable to explain a basic conversational move like choosing an appropriate addressee for a question.

What about the Shared Workspace framework? Recall that this framework proposes that speakers ask questions when there is a meta-representation of misalignment. As in the case of the Pig with the Hat, here too there is no negative response from the partner that is used in that framework to indicate misalignment. But this case actually contains no other situational cues to misalignment, and so in the absence of a representation of the other, the Shared Workspace framework lacks mechanisms that could detect misalignment. Crucially, even if the framework was extended to detect misalignment in this case, there are no mechanisms that would distinguish between the misalignment with James and the misalignment with the dentist, meaning there is no basis for choosing between the two potential addressees.

In sum, the case of the Knowledgeable Dentist illustrates the important role that sources of information play in communication, and in doing so, reveals the level of detail needed when representing the private information of the other (desideratum #1). In the Multiple Perspectives Theory, (assumed) information sources are encoded in the representation of the other (and also in the representation of self), and so this information will become available as part of the comparison process that reveals the other's epistemic advantage.

2.2. Reference

This section returns to the linguistic phenomenon that has been the focus of the perspective-taking literature, namely, reference. Mutual knowledge has had explanatory success in the domain of reference for both comprehension (e.g., Nadig & Sedivy, 2002; Heller et al., 2008; Metzig & Brennan, 2003; Brown-Schmidt, 2009b; Brown-Schmidt, 2012) and production (e.g., Isaac & Clark, 1987; Wardlow Lane & Ferreira, 2008; Heller et al., 2012; Gorman et al., 2013; Vanlangendonck, Willems, Menenti, & Hagoort, 2016). This is because the act of referring requires coordination: a speaker should choose a label the addressee can map onto the intended referent.

Yet, we demonstrate that, even in for reference, there are empirical findings that point to the need to *separately* represent the perspectives of the partners (desideratum #3), rather than having the representation of the other tied to the representation of self, either in common ground or via tagging (see Breheny & Zheng, 2020 for arguments about other limitations of common ground in explaining behavior in referential communication tasks). This point is illustrated by the case of the Empty Martini Glass for language comprehension (Section 2.2.1) and by the case of the Deceitful Lego for production (Section 2.2.2). Since both cases involve representations of false belief, a common topic in the ToM literature, we conclude by commenting on the relationship between language use and ToM (Section 2.2.3).

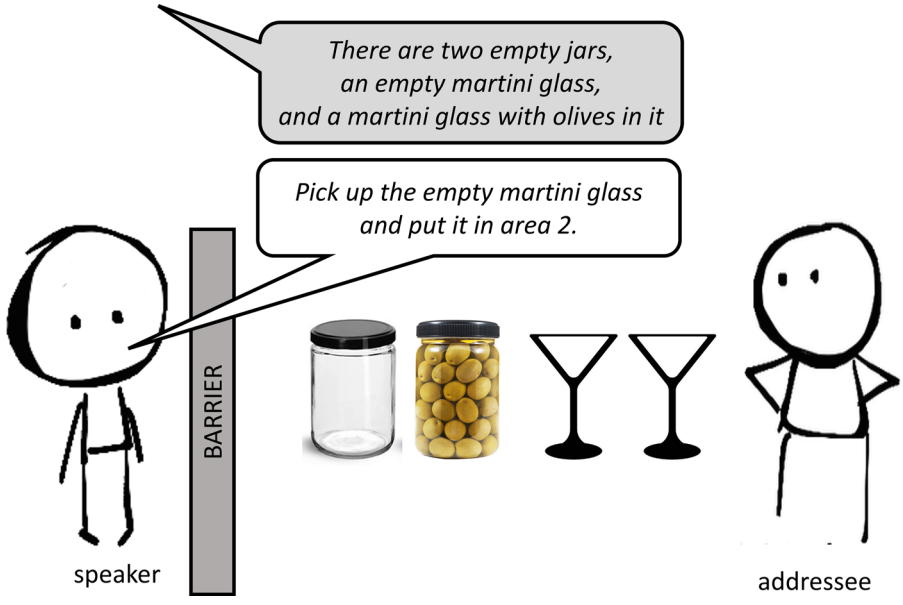

Fig. 3. Example trial from Hanna et al. (2003). Only the participant addressee saw the physical display, but both the speaker and addressee heard the experimenter's description: "*there are two empty jars*, *an empty martini glass*, *and a martini glass with olives in it*." In processing the speaker's instruction, addressees showed sensitivity to both linguistic and physical context.

2.2.1. Reference comprehension: The case of the Empty Martini Glass

In the case of the Empty Martini Glass, there is a discrepancy between the physical context and information expressed linguistically—see Fig. 3. Hanna et al. (2003, Exp. 2) created situations where the partner giving instructions (the speaker) could not see the objects in the display and only heard the experimenter describe these objects, whereas the partner following instructions (the addressee) both saw the physical display and heard the experimenter's linguistic description. Crucially, the description given by the experimenter was incorrect, creating a situation where the speaker and addressee had different—and inconsistent—assumptions about the display. In addition, the addressee was in a position to assume that the speaker had a false belief about the display due to the experimenter's faulty description.

In a referential task such as this one, a referring expression starting with "*the empty*" tends to create the expectation that the referent would be the empty jar; "*empty*" is less likely to be used for a martini glass because this information would not distinguish between the two glasses (see Heller, 2020 for a review). In contrast, the linguistic information given by the experimenter incorrectly described the situation as containing *two* empty jars, one empty martini glass, and one martini glass with olives. Against this (counterfactual) linguistic context, a referring expression starting with "*the empty*" would instead create the expectation that the speaker is referring to a martini glass, because "*empty*" distinguishes between the two martini glasses. Thus, depending on what contextual information is used—physical (real) or

linguistic (counterfactual)—the expression "*the empty*" leads to different, inconsistent expectations about the upcoming referent (note this uncertainty ends at the noun "*martini glass*").

Since definite referring expressions have been argued to be interpreted relative to mutual knowledge (e.g., Gundel et al., 1993; Clark & Marshall, 1978; Horton & Gerrig, 2005), we ask what counts as mutual knowledge between the speaker who is giving instructions and the addressee who needs to follow that instruction (see again Fig. 3). The answer could *not* be the physical context, because that information is only available to one partner, the addressee. Instead, the mutual information here is the (counterfactual) description given by the experimenter. Thus, theories that rely on what is mutual would predict that the unfolding expression (e.g., "*the empty*") would be processed relative to the linguistic information that was heard by both partners, and as such, addressees should anticipate the empty martini glass to be the referent. This prediction holds independent of whether this information is encoded directly in a representation of common ground (a la Clark and colleagues) or whether it is shared information retrieved from memory (as per Horton & Gerrig, 2005) or represented in a situation model (m-tagging in Pickering & Garrod, 2021).

The empirical findings of Hanna et al., however, are inconsistent with this prediction. Instead, the pattern observed was *intermediate* between an interpretation based on the mutual counterfactual information (anticipating a martini glass), and an interpretation based on the addressee's private knowledge of the physical context (anticipating an olive jar; see also Heller et al., 2008 for a related intermediate pattern).

The reader may argue that this pattern is, in fact, consistent with the standard common ground model, and it reflects processing relative to two representations: common ground and the addressee's egocentric perspective (we note that this alternative is not available in theories where one's own representation is tagged for what is shared, because those theories only have a single representation). This alternative account of these data may point to interference from the egocentric perspective during interpretation relative to common ground, or alternatively, the idea that the combination of the two perspectives is the mechanism that underlies perspective-taking (e.g., Heller et al., 2016). While this alternative will correctly account for the case of the Empty Martini Glass, it suffers from a conceptual limitation. This is because the information encoded in common ground—the counterfactual description coming from the experimenter—is actually just the speaker's own perspective (which is different from the addressee's perspective who also has visual information about the display). Thus, the case of the Empty Martini Glass is the first to suggest that referential processing during language comprehension, which has motivated much of the research on common ground, does not, in fact, rely on *mutual* information. Instead, on the Multiple Perspectives Theory, interpretation proceeds relative to the two *separate* perspectives of self and other (see also Ryskin, Stevenson, & Heller, 2020), correctly predicting this intermediate pattern.

2.2.2. Reference production: The case of the Deceitful Lego

In the case of the Deceitful Lego, it is the addressee who holds a false belief, and this false belief concerns object function. Mozuraitis et al. (2018) examined the production of referring expressions in contexts that included a visually misleading object, such as a crayon shaped to look like a Lego block, or a yoyo shaped to look like a baseball. While the speaker and

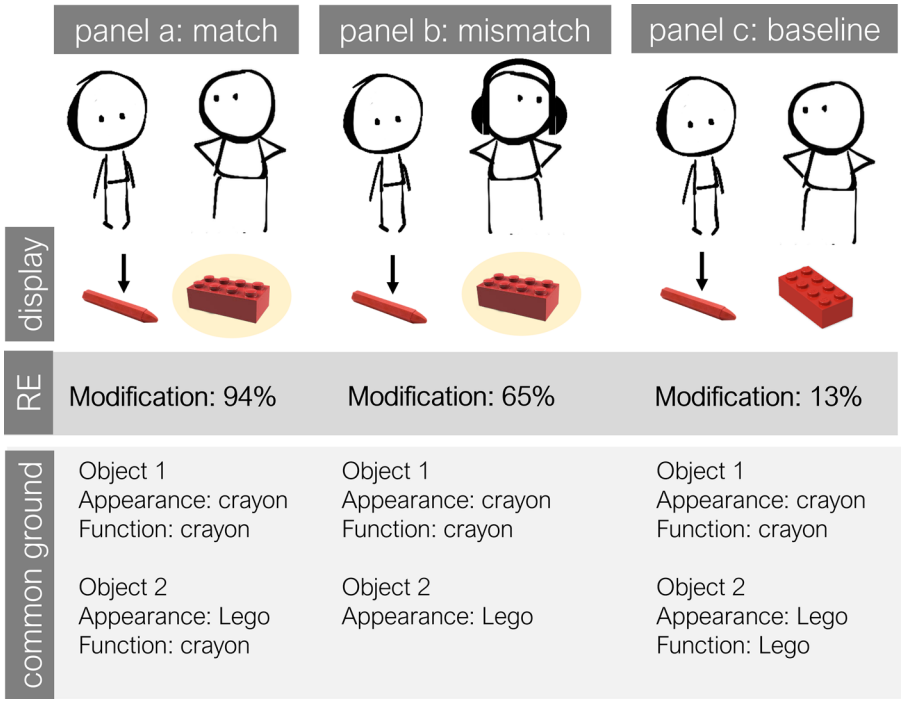

Fig. 4. The case of the Deceitful Lego: the conversational partners, the physical display (the referent is indicated with an arrow; a deceitful object is marked with a yellow background), the linguistic behavior of speakers, and the potential representation of common ground for each of the three conditions: match (panel a: a crayon and a visually misleading Lego-crayon), mismatch (panel b: a crayon and a visually misleading Lego-crayon), and baseline (panel c: a crayon and a regular Lego).

addressee could both *see* the visually misleading object, only the speaker knew its deceitful function (e.g., because the experimenter demonstrated that what looks like a Lego could actually be used for drawing). The addressee, who was not shown the surprising function of the object, presumably inferred—as people normally do—that the object's function was consistent with its appearance (i.e., that what looks like a Lego is indeed a Lego). Importantly, in this context, there is no *mutual knowledge about the function* of the visually misleading object: the speaker assumes it is a crayon, whereas the addressee assumes it is a Lego (see also Mozuraitis, Chambers, & Daneman, 2015; Baltaretu & Chambers, 2017).

Mozuraitis et al. (2018) exploited the fact that when speakers refer to an object, they must take into account what other objects exist in the scene (Olson, 1970; Osgood, 1971; Pechmann, 1989), and so they asked about the status of the deceitful object by looking at how speakers referred to a *different* object in the display. What referring expressions did speakers choose? In one context, where the display included a typical crayon (the referent) and the deceitful crayon and both partners knew the deceitful function of the Lego-crayon (the match condition: Fig. 4, panel a), speakers produced a modified expression (e.g., *the darker crayon*) about 94% of the time: this modification usually distinguished the referent

(typical crayon) from the visually misleading object (crayon that looks like a Lego). The behavior changed in a second context where the display included the same objects (a typical crayon and the Lego-crayon), but only the speaker knew about the deceitful function (the mismatch condition: Fig. 4, panel b): here modification rates were significantly lower at 65%. This difference in modification rate between the two situations that only differ in the knowledge state of the addressee indicates that speakers took into account the fact that the addressee did not know about the surprising function of the Lego.

In a third context where the display did not include a visually misleading object, and instead included two typical objects, such as a typical crayon and a regular Lego (the baseline condition: Fig. 4, panel c), modification rates were even lower at 13%. Note that this low baseline modification rate was obtained when the two objects in the display did not share *any* properties (e.g., one crayon and one Lego). Importantly, this baseline condition is parallel to the addressee's perspective in the mismatch condition: if one does not know the deceitful function of the Lego-crayon (Fig. 4, panel b), then one assumes that the display contains a crayon (the referent) and a Lego. Therefore, the different modification rate between the baseline condition (panel c) and the critical mismatch condition (panel b) reveals that in the mismatch case, speakers did not simply *adopt* their partner's perspective, but they were also affected by their own knowledge about the misleading function of the object.

What do existing theories predict in this case? We first consider what would be encoded in a common ground model. In the match condition, where the deceitful nature of the Lego is shown to both partners, the appearance and function of both the typical and deceitful objects are mutual knowledge, and as such would be encoded in the common ground (Fig. 4, panel a). In the mismatch condition, the appearance of the visually misleading object is mutual knowledge, but its function is not (both are mutual knowledge for the typical object, the crayon). One way to encode this situation in a common ground model is to assume that common ground is not specified in this case with respect to object function (Fig. 4, panel b); interestingly, an "empty workspace" is what Pickering and Garrod (2021) propose to for cases of perspective mismatch such as Wittgenstein's (1953) ambiguous duck-rabbit image.[3] The proposed contrast in representations between the match case (panel a) and the mismatch case (panel b) with respect to object function may initially seem satisfactory, as it can account for the difference in modification rates between the two conditions: Modification is needed more in the match case (a) in order to distinguish two objects that share the property [function: crayon], compared to the mismatch case (b) where function information is not encoded (see again Fig. 4). However, this account cannot predict the difference between the mismatch case (panel b) and the baseline condition (panel c): since neither representation encodes shared properties between the two objects (see again Fig. 4), it will wrongly predict that both will give rise to the same modification rate.

As in the previous section, the skeptical reader may counter that linguistic behavior can be modeled as a *combination* of the common ground representation and the egocentric perspective of the speaker. Unlike the case of the Empty Martini Glass, here such an account is not conceptually awkward, as the common ground representation we proposed above is faithful to its label in representing *mutual* knowledge (and not just the other's perspective). However, this account misses a central aspect of the situation, namely that the addressee draws

an inference that what looks like a Lego is indeed a Lego, and that the speaker is sensitive to this inference in choosing a referring expression. This inference does not belong in the common ground representation, because it only characterizes the addressee; similarly, it cannot be represented as shared information in memory (Horton & Gerrig, 2005, 2016) or by tagging one's own situation model (Pickering & Garrod, 2021). Thus, the reference production pattern is best explained as a combination of the *separate* and *distinct* perspectives of the speaker and addressee, as argued by Mozuraitis et al. (2018), an account we adopt in the Multiple Perspectives Theory (see Ferguson & Breheny, 2012 and Hawkins et al., 2021 for related findings).

2.2.3. *False belief in language use and nonlinguistic perspective taking: Taking stock*

The empirical findings discussed in this section demonstrate that modeling the production and comprehension of referring expressions requires the integration of the separate, *distinct* perspectives of the partners. These findings come from situations where the conversational partners have *different* representations of the same situation (the case of the Empty Martini Glass, Section 2.2.1) or of the same object (the case of the Deceitful Lego, Section 2.2.2). In the ToM literature, this type of knowledge mismatch is known as a Level 2 ToM representation, which include considerations of *different* beliefs (e.g., knowing that your partner mistakenly thinks a certain object is a Lego), which are considered distinct from Level 1 ToM representations, which include the *absence* of knowledge (e.g., knowing that your partner cannot see a jar that is visible to you). On several accounts, Level 1 and 2 representations have distinct developmental trajectories (Flavell, Everett, Croft, & Flavell, 1981), and involve distinct cognitive processes (Surtees, Apperly, & Samson, 2016; Surtees, Butterfill, & Apperly, 2012, Westra & Nagel, 2021).

In contrast to these cases, the bulk of the literature on language use focuses on the absence of belief—a Level 1 ToM representation. Examples include cases where information is given visually (e.g., Nadig & Sedivy, 2002; Wardlow Lane & Ferreira, 2008; Heller et al., 2008), cases where mutual knowledge relies on background information (Isaac & Clark, 1987) or on novel shared experiences (Wu & Keysar, 2007; Bromme, Jucks, & Wagner, 2005; Heller et al., 2012; Gorman et al., 2013; Hilliard, O'Neal, Plumert, & Cook, 2015; Hilliard & Cook, 2016), and cases where information is introduced linguistically (Metzing & Brennan, 2003; Hanna et al., 2003). Only a handful of studies have examined language processing in the context of Level 2 ToM, namely, false-belief (e.g., Hanna et al., 2003: Exp. 2; Keysar, Lin, & Barr, 2003: Exp. 2; Ferguson, Scheepers, & Sanford, 2010; Ferguson & Breheny, 2012; Mozuraitis et al., 2015; 2018). As we saw in this section, it is those Level 2 cases that critically demonstrate the need for separate, distinct representations of the perspectives of conversational partners.

Research on ToM outside language, however, often focuses on Level 2 representations that involve false-belief. For example, the cognitive development literature has been concerned with when the capacity to reason about false belief emerges in childhood (e.g., Dorrenberg, Rakoczy, & Liszkowski, 2018; Baillargeon, Buttelmann, & Southgate, 2018; Westra & Carruthers, 2017). Furthermore, research on (nonlinguistic) perspective-taking in adults provides evidence that the encoding of the other's perspective alongside one's own emerges spontaneously, providing further evidence for desiderata #3 (separate representations of the self and

other). However, in this literature, common parlance includes "adopting" or "taking another person's perspective" (Eyal et al., 2018; Savitsky, Keysar, Epley, Carter, & Swanson, 2011; Kampis & Southgate, 2020). In other words, behavior is framed as the adoption of a *single* perspective, either the self (i.e., egocentric) or of the other (i.e., altercentric). Relatedly, communicative success has been argued to involve the inhibition of one's own perspective in order to consider the mind of the other (e.g., Hartwright, Apperly, & Hansen, 2012; Vogeley et al., 2001), especially in research where this process is argued to be effortful (Savitsky et al., 2011; Epley et al., 2004; Epley et al., 2004). However, a theory where a single perspective is used does not fulfill desideratum #2: computing similarities and differences between the perspectives. Of note, Deschrijver and Palmer (2020) recently proposed that the perspective-taking literature outside language should shift from a representation of the other to a relational view of conflict monitoring. However, in the context of language use, differences in perspectives are a natural situation that drives information exchange rather than an undesirable situation that needs to be resolved (the use of the word "conflict" is revealing here).

2.3. Memory for conversation, and why it matters for a theory of communication

The processes that shape language use in the moment depend on memory representations of past experiences. These conversational processes, in turn, impact subsequent language use, both later in the same conversation and in subsequent conversations. For example, memory for the people and places you encounter in daily life allows you to understand references to them in conversation, and allows you to continue to talk about those people and places in the future. Thus, in developing theories of communication, we must consider findings and insights from studies that directly probe memory for conversation.

To illustrate, we begin by drawing upon an impactful paper by Ross and Sicoly (1979), which examined biases in memory for past interactions. In one problem-solving task, participants discussed a psychological case study, and then recalled that conversation after a 3- to 4-day delay. Comparing the recalls to the original recordings revealed that participants recalled 5.6% of their own statements, but only 2.6% of their partner's statements. This study illustrates, first, that conversational partners were unable to recall many details of the conversation, and, in addition, what they did recall was biased in favor of their own contributions.

The finding that *very little* of a conversation can be accurately recalled in detail has been reported many times since: After delays of several minutes to several weeks, participants can accurately recall a mere 0%−40% of the total "*idea units*" expressed in the conversation (Stafford & Daly, 1984; Stafford, Burggraf, & Sharkey, 1988; Ross & Sicoly, 1979; Pezdek & Prull, 1993; Samp & Humphreys, 2007). While coarse-grained or gist information is remembered best (e.g., Christiaansen, 1980; Conway, Cohen, & Stanhope, 1991; Bransford & Franks, 1971), the observation that memory for conversation is limited constrains the *upper bound* of what could functionally constitute mutual knowledge for the discourse history. Furthermore, memory is affected by different characteristics of the conversational experience, including its content and richness (Campos & Alonso-Quecuty, 2006; Pezdek & Prull;, 1993; Keenan, MacWhinney, & Mayhew, 1977; Kintsch & Bates, 1977; MacWhinney, Keenan, &

Reinke, 1982), suggesting that there could be considerable *variability* across partners in what is remembered.

These limitations in the *quantity* of information that can be recalled raise the question about the type of representation that supports, for example, the shortening of referring expressions in the tangram task (see Fig. 1). Neuropsychological evidence provides key insights: Individuals with bilateral hippocampal damage and severe amnesia show profound deficits in memory for new events (episodic memory) and facts, yet they successfully acquire referential labels for tangram images in referential communication tasks (Duff, Hengst, Tranel, & Cohen, 2006). At the same time, basic linguistic processes are notably impaired in these individuals over timescales of less than a minute (Duff & Brown-Schmidt, 2017; Kurczek et al., 2013; Covington et al. 2020). For example, if information is mentioned linguistically, and referenced again after ∼40 s, individuals with hippocampal amnesia fail to treat that information as mutual knowledge (Rubin et al., 2011). This pattern of sparing and impairment clearly shows that memory representations support language processing. Yet, the fact that individuals with amnesia successfully acquire brief referential labels in the tangram task suggests that the changes in referential form observed in this task are unlikely to depend on the ability to recall the conversation that gave rise to those labels. Thus, rather than representing the label "*the dancer*" in common ground with the partner, individuals may come to conceptualize a certain abstract figure as "*the dancer*" (Yoon, Duff, & Brown-Schmidt, 2017). These findings are relevant in the context of the standard common ground model that assumes it is an episodic memory record of the discourse history that supports inferences about common ground (i.e., Clark & Marshall, 1978). Together with findings that non-brain-injured adults recall less than 40% of what was said in a conversation after brief delays, this suggests that episodic memory for what was said per se may not be an essential ingredient in the phenomenon of referential shortening.

The second important finding is the *asymmetry* in what conversational partners remember. This asymmetry arises from the fact that memory is generally superior for what a person has *said*, compared to was *heard* (Fischer, Schult, & Steffens, 2015; Miller, deWinstanley, & Carey, 1996; Ross & Sicoly, 1979; Yoon, Benjamin, & Brown-Schmidt, 2016; McKinley, Brown-Schmidt, & Benjamin, 2017; Isaacs, 1990; cf. Stafford & Daly, 1984; Stafford et al., 1988 for a different pattern of results). This memorial asymmetry is reminiscent of well-established findings that generating or producing information enhances memory for it, phenomena known as the "generation effect" and "production effect" (Slamecka & Graf, 1978; MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). Relatedly, speakers find it more rewarding to disclosure personal information compared to stating facts, suggesting further asymmetries in the reward systems associated with producing information of different types (Tamir & Mitchell, 2012). Similarly, in storytelling, speakers expect novel stories will be more enjoyable, whereas listeners prefer to hear familiar stories (Cooney, Gilbert, & Wilson, 2017).

The memory asymmetry between speakers and listeners has also been demonstrated following a referential communication task—the task most often used to study perspective-taking in conversation. McKinley et al. (2017) had two partners complete a referential communication task in which they discussed a series of images. A subsequent recognition memory

test indicated that the odds of correctly recognizing a picture was 2.64 times greater if the person had described that picture themselves, compared to when their partner described it to them. Critically, this asymmetry in memory held even for items for which partners had gone through the process of grounding and converged on a short label (which has been argued to cement the representation in common ground, and has been adopted in the Shared Workspace framework as signaling alignment—see again Section 1). Notably, these memory effects emerge not only after a conversation is over, but they also influence language *during* conversation: when a person repeats something in conversation, they are significantly more likely to repeat what they themselves said than what their partner has said to them (Knutsen & Le Bigot, 2014). Interestingly, repeated information is better remembered, and people correctly guess that their partner will remember repeated information (Knutsen & LeBigot, 2021).

What do these memory findings tell us about theories of communication? These asymmetries in memory between partners highlight the importance of constructing and maintaining *separate* representations, one for each partner (desideratum #3). If conversational partners accumulate information as part of common ground (Schober & Clark, 1989), their common ground representations will be asymmetric in a systematic way, with each person's common ground representation biased toward their own contribution. This could be accommodated by assuming separate, asymmetric representations of common ground for each partner. However, when considered alongside the findings from questions (Section 2.1) and reference (Section 2.2), these findings emphasize the need for a new theory that is not centered on mutual knowledge.

These memory findings also offer a strong critique of the process of interactive alignment (Pickering & Garrod, 2004) which also plays a key role in the Shared Workspace framework (Pickering & Garrod, 2021). These theories argue that language production and language comprehension both draw on, and are mutually influenced by, the same representations, and are responsible for the alignment between conversational partners. To accommodate the findings about asymmetry in memory representations, these theories could be modified such that one's own produced linguistic forms would be *more* activated compared to those heard as a listener. However, the result would be that partners would be systematically mismatched in terms of which linguistic forms are most activated. Further, introducing such an asymmetry goes against the main goal of alignment, namely to account for how conversational partners converge on similar linguistic forms at different levels of representation. In sum, taken together with the limitations of the Shared Workspace framework as it concerns questions and reference under false belief, this offers strong evidence against the framework as a whole.

To summarize Section 2: this section reviewed existing empirical evidence, both linguistic (questions, reference) and nonlinguistic (ToM, memory). We pointed out the limitations of existing theories: from common ground that encodes mutual knowledge, to ordinary memory where one's own memory is used to activate shared information with the partner, to the Shared Workspace framework which, in addition to a tagged situation model, also adds metarepresentations of (mis)alignment (note: metarepresentations of mental states were first proposed by Horton & Brennan, 2016). This evidence points to three desiderata for a theory of mental states—the private information about the other (desideratum #1), similarities and

differences between the self and other (desideratum #2), separate representations of the self and other (desideratum #3)—that do not have a clear place in any of the existing theories.

We outlined how our theory fulfills these desiderata, and successfully handles the representations of mental states that support the use of questions, reference under false belief, and memory asymmetries using three components: the representation of self, the representation of other, and the comparison process. What about including a representation of common ground as part of this inventory? While this representation has proven useful in half a century of language research, we depart from this tradition and argue that this type of representation is not, in fact, part of the cognitive architecture that supports language use. First, because the information encoded in the common ground is already available elsewhere, adding this representation as a fourth component would result in a less parsimonious theory. More important, it is unclear how this representation would be used. For example, would some definite descriptions be processed relative to common ground (e.g., "*the pig with the hat*": see Section 2.1.1), while others relative to the distinct perspectives of the self and other? (see the case of the Empty Martini Glass and the case of the Deceitful Lego in Section 2.2). Similarly, how would a question be processed (e.g., "*What's below the pig with the hat?*"): relative to the two perspectives (that highlight both differences and similarities in perspective), or relative to the common ground which licenses the definite? In our theory, information about mutual knowledge is just one of the possible outputs of the comparison process.

3. The Multiple Perspectives Theory of mental states in communication

In this section, we develop the three components of our theory in more detail: the representation of self, the representation of other, and the comparison process.

3.1. Representation of self

Recall that the first component of the Multiple Perspectives Theory is an estimated representation of one's own perspective. We note that the representation of self is related to metacognitive proposals, including work from the memory literature that examines assessments of learning, and judgments of one's own knowledge (e.g., Dunlosky & Nelson, 1992; Tullis & Benjamin, 2012), such as a feeling of knowing (Smith & Clark, 1993; Brennan & Williams, 1995). Even early efforts to develop artificial agents anticipated the need for agents to represent the self, to allow them to participate in turn-taking with humans (Wachsmuth, 2008).

3.1.1 Relevance

In the representation of self, one accesses the *relevant* bits of information for the purposes of the current conversational turn, with relevance being dynamically updated. The notion of relevance is shared with both Grice (1975) and *Relevance Theory* (e.g., Sperber & Wilson, 1986, 1995), and is similar to ideas in formal linguistics that utterances are interpreted relative to the Question Under Discussion (QUD: Roberts, 1996/2004). Cognitively, however, one

must keep in mind that relevance exhibits biases that temper the influence of what is truly relevant: findings in the judgment and decision-making literature suggest that people under-rely on relevant information, and over-rely on irrelevant information (i.e., *relevance insensitivity*, Hsee, Yang, & Li, 2019).

3.1.2 Sources of information

The case of the Knowledgeable Dentist demonstrated the role of sources of information to choosing an addressee for a question (Section 2.1.2), but sources of information are also relevant to assertions. Making an assertion involves an assessment of one's level of certainty about this information, and possibly the evidence for it. The typological literature has argued that about 25% of the world's languages mark sources of evidence using *grammatical* elements, known as evidential markers (Chafe & Nichols, 1986; Willett, 1988; Aikhenvald, 2004; de Haan, 2013), and these markers have more recently been further analyzed in the formal semantics literature (e.g., AnderBois, 2014; Korotkova, 2016; Faller, 2019). Evidential markers are used in assertions to indicate how the speaker knows the information they are reporting. For example, in Turkish, verbs in the past tense include a suffix that marks the source of information as either direct perception or indirect evidence (e.g., Ünal & Papafragou, 2016, 2020). Thus, if we were to rewrite the case of the Knowledgeable Dentist in Turkish, the different individuals would use different evidential markers to report that James has brushed his teeth. James' brother, who saw James brushing his teeth, would use the marker of direct perception (*fırçaladı* "brushed-direct"), and so will James himself, whereas the dentist—who did not witness the event and instead infers this information from the examination—or her colleague who read the notes, would use the marker of indirect evidence (*fırçalamış* "brushed-indirect"). Other languages (e.g., Korean) encode more fine-grained distinctions, such as distinguishing information that is *inferred* based on direct evidence (e.g., the dentist) and information that is *reported* by another person (e.g., the colleague). The existence of these grammatical elements further suggests that sources of information are encoded in one's perspective alongside the information acquired.

3.1.3 Full-color versus placeholder representations

We propose that information can be represented to different levels of specificity: alongside detailed or full-color representations that support assertions, we posit the existence of *placeholder* representations. One type of placeholder representation is knowing that you do not know: To ask a question, the speaker must identify a *gap* in their own knowledge, what Phillips and Norby (2019) call "egocentric ignorance," with the goal of the question being to replace the placeholder with the desired information. This proposal is consistent with the view in formal linguistics that the meaning of a question is defined by the set of its potential answers (e.g., Hamblin, 1973; Groenendijk & Stokhof, 1984).

We come back to the representation of self in discussing the comparison process of self and other in section 3.3.

3.2. Representation of other

The second component of the Multiple Perspectives Theory is a representation of the other, an *estimate* of the other person's mental state that is inferred from indirect cues. As with the representation of self, one accesses aspects of the partner's perspective that are *relevant* to the purposes of the current conversation (Section 3.1); the representation of the other also participates in the comparison process.

3.2.1. Indirect cues

The cues used to infer the representation of other have been extensively discussed in the context of inferences about mutual knowledge (e.g., Clark, 1996; Mason, 2004; Clark & Schaefer, 1987; Isaacs & Clark, 1987). These include visual and linguistic cues that are part of shared experiences, and the social groups one belongs to. As our cases in Section 2 demonstrate, the same cues are used to infer information that is not part of the self-perspective (e.g., the incorrect function of the deceitful Lego), including private information attributed to the other (e.g., the identity of the animal below the pig with the hat or whether James has brushed his teeth that morning). Because information about the perspective of the other is inherently uncertain, we propose that it is represented as probabilistic (cf. Hanna et al., 2003; Brown-Schmidt, 2012): if there is a lot of evidence that the other assumes a certain piece of information, the probability will be high, and if evidence is weak or limited, that probability will be low.

Empirical evidence demonstrates that the other's perspective is sometimes inferred even when this information is not immediately relevant, both for visual perspective (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010) and false belief (Schneider, Bayliss, Becker, & Dux, 2011). Relatedly, Savitsky et al. (2011) propose that people actively monitor strangers' perspectives more than friends' perspectives, because they anticipate that they might diverge more from their own perspective. Similarly, Nigro and Neisser (1983) show that personal memories can be encoded either from one's own perspective or from the perspective of an observer, suggesting the intriguing possibility that the same event may be encoded differently in the representations of self and other—we discuss this possibility in more detail in the case of the Wet Balcony (Section 4.2). Of course, the representation of other may suffer from systematic biases: for example, both the language and memory literatures indicate that such estimates tend to be biased in the direction of one's own knowledge (Fussell & Krauss, 1991; 1992; Tullis & Fraundorf, 2017).

3.2.2. Sources of information

As in the representation of self, information in the representation of other may be marked for its source. Including the source allows us to account for how you choose an appropriate addressee in the case of the Knowledgeable Dentist. As discussed for the representation of self, some languages encode sources grammatically, with speakers tracking the likely sources of information that others have, even when they themselves do not have that information. Specifically, when evidential markers appear in questions, they often exhibit a "flip" in the

information source: the asker of the question marks the information source from which the addressee is assumed to have learned the answer (e.g., Aikhenvald, 2004; Korotkova, 2016).

3.2.3. Full-color versus placeholder representations

Like in the representation of self, information about the other can be represented to different levels of specificity. Detailed or *full-color representations* involve the attribution of fine-grained knowledge to the other. For example, in the case of the Pig with the Hat, Ilana attributes to Abbi the ability to see the pig with the hat (Section 2.1.1). Not all the information attributed is true: as discussed in the literature on false belief (e.g., Hanna et al., 2003; Wimmer & Perner, 1983), one may attribute false information, such as the false assumption that what looks like a Lego is indeed a Lego (The case of the Deceitful Lego, Section 2.2.2). Importantly, because information about the other is inferred, one may sometimes make an incorrect attribution. For example, Abbi may (incorrectly) assume that Ilana—who is Canadian—is fluent in French.

Importantly, one may also attribute *placeholder* representations. This includes attributing information that is private to the other. For example, in the case of the Pig with the Hat, Ilana attributes to Abbi knowledge of what animal is below the pig with the hat, without herself knowing which animal it is—in this case, the attribution is based on the visual cue of seeing a black square (in that experimental setup, the partner always has an animal "behind" the black square). These attributions can also be based on social categories: Ilana can attribute to Abbi knowledge about hiking in the Smokey mountains, whereas Abbi can attribute to Ilana knowledge about Toronto restaurants, despite the fact that neither has fine-grained details. Another type of placeholder representations may be attributed when you know that your partner has *an absence* of knowledge: in the case of the Pig with the Hat, Abbi attributes to Ilana absence of knowledge about the animals in the cubbyholes that are hidden from Ilana's view (see Keysar et al., 2003; Wimmer & Perner, 1983).

Our representation of the other contrasts with a common ground model and with tagged situations models and activated shared knowledge in that it does not tie the other's perspective to one's own perspective. Having *separate* representations of self and other provides the necessary tools to account for the cases discussed in Section 2: It supports asking questions (Section 2.1.1)—including choosing an appropriate addressee for a question (Section 2.1.2), it allows us to explain the production and comprehension of referring expressions in situations where the perspectives of the partners are inconsistent (Section 2.2), and it also allows for the representations of self and other to diverge in memory over time (Section 2.3). Critically, this type of representation frees us from the symmetry dictated by mutual knowledge, whether directly encoded in a common ground representation or indirectly represented through one's own memory or situation model.

Our cognitive architecture is similar to what has been assumed in other research traditions that are concerned with mentalizing the other. One area that assumes separate representations is perspective-taking (or mentalizing) outside language (Section 2.2.3), such as inferences about the belief, goal, and desire states of other agents (Samson et al., 2010; Schneider et al., 2011; Surtees et al., 2016; Surtees et al., 2012; Flavell et al., 1981; Thornton & Tamir, 2020). Similarly, the developmental and comparative literatures ask whether, how, and when

representations of the other minds develop (Wimmer & Perner, 1983; Premack & Woodruff, 1978; Drayton & Santos, 2016; Westra & Carruthers, 2017). Thus, one attractive consequence of modeling conversation using distinct representations of self and other is convergence with these research traditions (see also Apperly, 2018).

3.3. The comparison process

The representations of self and other concern the way mental state information is *stored*, but how is this information *used*? For example, one asserts information that comes from the representation of self, but the asserter must also consult the representation of the other in order to determine that the addressee is not likely to already know this information. When asking a question, the asker will identify a gap in their self-representation, but they must also identify a placeholder representation for the answer in the representation of the other. Furthermore, in order to tailor their message such that it would be interpretable for the particular addressee, a speaker must identify *similarities* between the perspectives of self and other. Thus, it is not sufficient to consult the representations of self and other *separately*, but the *relationship* between the two must also be determined.

At the heart of our Multiple Perspectives Theory is a cognitive process that compares the representations of self and other, with the transient output of this process being used in planning the next conversational move. What aspects of the representations are compared? The process targets the subset of information deemed relevant for the current exchange. This follows the notion of relevance assumed by Grice (1975) where contributions are evaluated "for the current purpose," as well as by Relevance Theory (Sperber & Wilson, 1982, 1995). When does the comparison process occur? As illustrated in Section 3.4, representations of self and other are hypothesized to be continuously compared throughout the conversation. The logic of this hypothesis is that the comparison process uses computations of speaker meaning which unfold over time as language is processed (Bögels, 2020; Eberhard et al., 1995); this hypothesis awaits future empirical testing.

The comparison process can output either similarities or differences in perspective; in both cases, the transient output is used in the moment to update the representations of self and other. If the representations of self and other for a given piece of information are *different*, the output of the comparison process reveals, first, the *type* of mismatch (note that because the only representations in the theory are of the self and other, it is the output of the comparison process that will reveal the existence of a Level 1 or a Level 2 ToM mismatch). There are three possible outcomes: (i) **Epistemic advantage**. You (i.e., the self) possess information the other does not, which could lead to an assertion (or an answer to a question). This situation is illustrated in the case of the Misleading Martini in the next section. (ii) **Epistemic ignorance**. You have placeholder information about knowledge you lack, and the comparison process indicates your partner is likely to have this information. In the case of the Pig with the Hat (Section 2.1.1), this type of outcome leads Ilana to ask Abbi "*What's below the pig with the hat?*". (iii) **Epistemic difference**. The comparison process determines that you and your partner have *distinct* representations. Such cases may involve false belief, as in the case of the Empty Martini Glass (Section 2.2.1) and the case of the Deceitful Lego (Section 2.2.2),

but also cases where partners have different perspectives: for example, a given tangram in Fig. 1 might be conceptualized by one partner as a superhero and by another as an angel (Yoon & Stine-Morrow, 2019). Such experimental situations map onto real-world settings where you may think of a certain person as your neighbor, whereas I may think of them as a famous cellist (E. Clark, 1997). Distinct perspectives are sometimes explicitly encoded in utterances, as in "*the rat-lobster*" for a case where one person sees a rat and the other a lobster (Clark & Wilkes-Gibbs, 1986; see Brown-Schmidt & Tanenhaus, 2008). Outputs indicating difference not only include the *type* of difference, but also the *level* of asymmetry between the partners (see Gunlogson, 2003; Heritage, 2012a). For example, Gunlogson (2003) points out that in the case of epistemic ignorance, different forms of questions reflect a different level of epistemic asymmetry: polar interrogatives (e.g., *Did you get a haircut?*) are uttered when there is greater epistemic asymmetry than rising declaratives (e.g., *You got a haircut?*). We argue that differences—or epistemic asymmetries—are the driver of information exchange in communication.

The second possible outcome of the comparison process is that the representations of self and other are the same with respect to some currently relevant piece of information. Thus, the comparison process may output *mutual knowledge* (but, unlike on a common ground model, this output is a transient representation used to update the representations of self and other). For example, in the case of the Pig with the Hat (Section 2.1.1), the comparison process reveals that the partners have the same knowledge about the two mutually visible pigs, allowing Ilana to use a definite description ("*the pig with the hat*"). Another relevant example is our tangram example (Fig. 1), where the comparison process determines that you and the first partner share the conceptualization of the top-left tangram as a woman climbing, whereas the comparison process with the second partner will output a difference. We argue that detecting similarities between perspectives is the basis for determining *how* to encode messages such that they can be successfully interpreted by the addressee.

It is important to note that not all similarities arise from the output of the comparison process. Conversational partners can be assumed to have similar surface representations of recent language that are active in memory, as recency entails that they would normally have the *same* representations of what was said. This representation of recent language plays a crucial role in supporting a range of linguistic phenomena, including contrastive accents and repetition reduction (see Wagner & Watson, 2010), syntactic persistence effects (e.g., *At what time do you close*? *At three o'clock*; Levelt & Kelter, 1982; Branigan, Pickering, & McLean, 2005), "surface" anaphora (VP ellipsis, sluicing, and gapping; see Hankamer & Sag, 1976; Tanenhaus & Carlson, 1990), and Echo questions (e.g., A: *I'm pregnant* B: *You're pregnant?!?*; Noh, 1998). In contrast to the information encoded in the representation of the self and other, these representations are linguistic in nature, and they decay after short delays (Brewer, 1977; Bransford & Franks, 1971; Bransford, Barclay, & Franks, 1972; Sachs, 1974); this activation of linguistic forms plays a central role in Interactive Alignment (Pickering & Garrod, 2004) and in the Shared Workspace framework (Pickering & Garrod, 2021). As time goes by, however, only some of this conversationally communicated information is likely to be recalled, particularly coarse-grained representations of what was discussed (e.g., Connine, Blasko, &

Hall, 1991; Christiansen & Chater, 2016), and salient or engaging information (Pezdek & Prull, 1993).

How does the comparison process compare with existing theories? Unlike the classic common ground model and ordinary memory theory, the comparison process of MPT outputs differences between perspectives, not just similarities. Unlike the Shared Workspace framework with its meta-representations of alignment and misalignment, the output of the comparison process indicates similarities and differences with respect to all relevant pieces of information. The output of similarities and differences contrasts with existing approaches in being transient and used to update the separate representations of self and other; it is this separate storage that allows the representations of self and other to diverge over time. Because comparing self and other does not depend on having the same type of representations between the partners, the comparison process can be extended to accounts of interaction with nonhuman animates and machines. Related argumentation can be found in discussion of the anthropomorphization (see Epley, Waytz, & Cacioppo, 2007; Jaeger & Levin, 2016; Baker, Hymel, & Levin, 2018) and simulation of and by nonhuman entities (Johnson & Demiris, 2005)[4].

3.4. Putting it all together: The case of the Misleading Martini

We illustrate the Multiple Perspectives Theory with the case of Misleading Martini, which takes place when Ilana and Abbi are attending a party (Fig. 5). Ilana observes Abbi drinking out of a martini glass, and the drink is almost gone—this is time point 1. Ilana knows that Abbi normally has multiple drinks, and asks "*Another martini*?" (a rising declarative that is biased toward a positive answer). On our theory, Abbi is processing the question using the perspectives of self and other, and, in addition, their comparison reveals a discrepancy: while the question reveals that from Ilana's perspective' Abbi's drink is a martini, in Abbi's self-perspective, it is a nonalcoholic mocktail (because she is pregnant). The identification of the discrepancy leads to a further inference: Because Abbi assumes that Ilana would not offer her a martini if she knew she was pregnant (based on Abbi's representation of Ilana), she now infers that Ilana does not know about her pregnancy (time-point 2). The output of the comparison process at time-point 2, Abbi's inference that Ilana does not know she is pregnant, is added to Abbi's representation of Ilana, and also to Abbi's representation of self (including potentially the inferential source of information). This information could simply be stored, but in this example, the newly identified asymmetry between perspectives leads Abbi to utter the assertion "*I'm pregnant*" (time-point 3), after which Abbi would update her representation of Ilana with the information that Abbi is pregnant (there is no need in this case to update the representation of self). This piece of information in the stored representation of the other might be revisited at a future date, for example, if Abbi wants to borrow some maternity clothes some months later, she can ask "*Do you have any maternity clothes I can borrow?*", without first informing Ilana why it is that she needs them.

While in this example the comparison process outputs that Ilana does not know that Abbi is pregnant, in other cases, the result of the comparison process may involve further embeddings. For example, if Ilana previously found out that Abbi was pregnant (and tried to keep this knowledge a secret), Abbi may discover this, and compare their mental states and determine

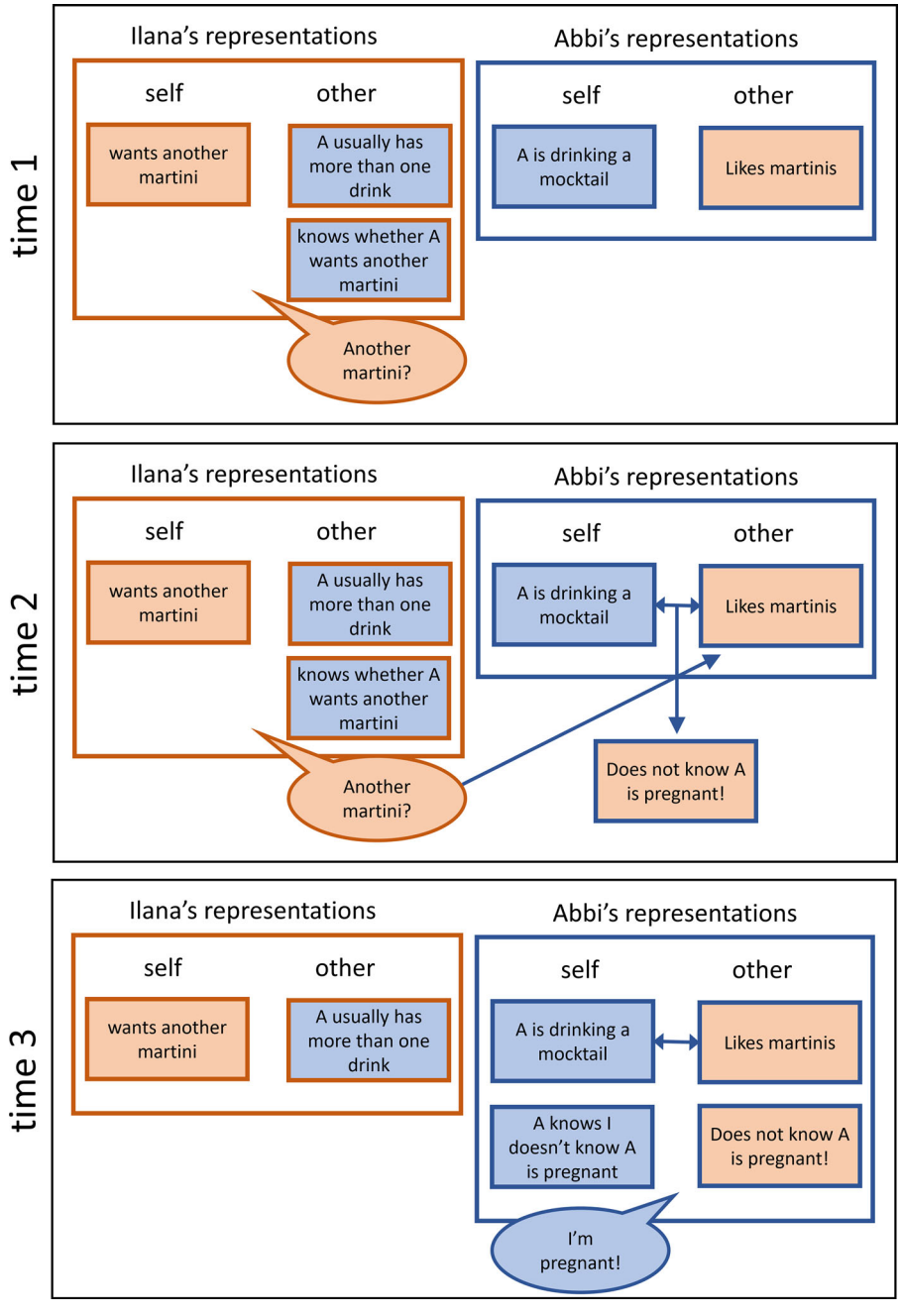

Fig. 5. Ilana's representation of the self and other, Abbi's representation of the self and other, and how they are updated over time in the case of the Misleading Martini.

that Ilana does not know that Abbi knows that Ilana knows that Abbi is pregnant. Upon making this comparison in mental states, Abbi could then utter, "*Ilana*, *I know you know*, *no need to pretend anymore*." This example demonstrates that embedded mental states—at least up to a certain level—are represented and tracked, counter to prior theories that assumed that embeddings are not computed (Section 1).

4. Similarities and differences in language use: Looking ahead

The cognitive architecture proposed in the Multiple Perspectives Theory sheds new light on language use in conversation. For example, the comparison process leads us to consider the role of differences of perspective in language, and the separate representations of self and other open the possibility to divergence over time. These novel aspects of or theory lead us to introduce novel phenomena in language use, with the goal of inviting new directions for future empirical research.

4.1. Cross-linguistic grammatical marking of similarities and differences in perspective

Because previous theories of mental states have focused on mutual knowledge, the literature has generally examined grammatical phenomena that are sensitive to similarity in perspective. The phenomenon that has been studied most extensively is referring expressions and especially definite descriptions (since Clark & Marshall, 1978), but other grammatical elements that trigger presuppositions have also been considered, such as the verb "*return*" (e.g., Chambers & San Juan, 2008). By contrast, a core aspect of the Multiple Perspectives Theory is that the same mechanism used to compute similarities between self and other is also used to compute *differences*. The Multiple Perspectives Theory, therefore, predicts the existence of linguistic phenomena that are sensitive to *differences*, or asymmetries in perspective. As discussed in Section 3, an asymmetry between conversational partners serves as the precondition for the most basic conversational moves—namely, assertions and questions, but our theory further predicts the existence of *grammatical* elements that are sensitive to perspective asymmetries, and these should be found alongside grammatical elements that are sensitive to symmetry in perspective.

One relevant class of grammatical elements are evidentials (first mentioned in Section 2.1.2), which have been introduced and analyzed in the fields of typology (Chafe & Nichols, 1986; Willett, 1988; Aikhenvald, 2004; de Haan, 2013) and formal semantics (AnderBois, 2014; Korotkova, 2016; Faller, 2019). For example, the particle "*yo*" in Japanese has been noted by Davis (2011) and McCready (2005) to mark cases where the speaker's assertion (which encodes information from the self-perspective) explicitly contrasts with the addressee's perspective, an asymmetry that may arise when the addressee's actions indicate that they do not know this information, when the addressee has uttered a sentence that contradicts this information, or when the speaker thinks the addressee had forgotten this information. We note that grammatical marking of sources of evidence may require retrieval of source information from memory; whether speakers are aware of this is left for future

research. Future empirical investigations of evidential markers will also have to consider that the knowledge itself can be dissociated from information about the origins of how you learned that information (e.g., Johnson & Raye, 1981; Foley, Foley, Durley, & Maitner, 2006), and who told you (e.g., Fischer et al., 2015; Gopie & MacLeod, 2009).

Grammatical elements that are sensitive to perspective (a)symmetry are not limited to sentence-level particles, which are somewhat akin to conversational moves, but are also found in the domain of referring expressions. One such example is the determiners and demonstrative pronouns in Turkish (Küntay & Özyürek, 2006): when the speaker and the addressee jointly attend to the referent (i.e., symmetry), two elements are available which encode the proximal versus distal contrast (*bu* vs. *o*). In contrast, when the addressee's attention to the referent has not been established (i.e., asymmetry), a different demonstrative form is used (*šu*). As reviewed in Evans, Bergqvist, and San Roque (2018a), other languages also exhibit sensitivity to perspective (a)symmetry in their inventory of demonstratives: Bininj Gun-wok, which tracks whether the addressee was previously interested in the referent, and the Athapaskan language Kaska which encodes the shared versus unshared status of referents.

More generally, the typology literature has documented a range of elements that are sensitive to (a)symmetry of perspectives. Evans, Bergqvist, and San Roque (2018a, 2018b) review a range of grammatical phenomena from the world's languages that are sensitive to (a)symmetries in both knowledge and attention. Such perspectival (a)symmetries are found across the grammar, from elements used to talk about entities (e.g., demonstratives), to elements used for events (e.g., verbs), propositions, and metapropositions (markers of evidentiality and certainty). For example, this includes languages of New Guinea (Foe, Wola, and Pole), languages of South America (Jaqaru and Southern Nambikuara), and the Tibeto-Burman language Kurtöp.

The prevalence of grammatical elements that are sensitive to differences in perspective provides preliminary support to the Multiple Perspectives Theory which considers perspective asymmetries as a central type of mental state in conversation. Moreover, the existence of pairs of grammatical elements that change minimally in response to perspective (a)symmetry suggests parallelism between perspective similarities and perspective differences: this provides support to our theory that derives both using the same mechanism of the comparison process (over any theory that treats symmetry differently from asymmetry).

4.2. Asymmetries in updating representations, and the case of the Wet Balcony

In theories that focus on mutual knowledge, communication is generally thought to lead to *convergence*: either because the common ground grows (Wilkes-Gibbs & Clark, 1992), or because partners become increasingly aligned in their representations (Pickering & Garrod, 2004, 2021). By contrast, in the Multiple Perspectives Theory, communication leads to the updating of two *distinct* representations: the self and the other. Our theory, therefore, makes the novel prediction that updating the two representations with the same message could in some cases lead to divergence in perspective rather than convergence.

We illustrate this prediction with the case of the Wet Balcony. Imagine you are having a party at your apartment. Your partner goes out onto the balcony, and when they come back in,

they say "*It's really wet out there*." Your friends infer that it is raining and update their representations accordingly. You, however, know that your balcony gets wet when your upstairs neighbor washes off their balcony, and, furthermore, you know they only do that on evenings with clear skies, and so you update your self-representation with this information. Importantly, because you do not expect your guests to draw the same inference, you apply a different update to the other representation (self: it is clear outside; other: it is raining). The case of the Wet Balcony thus demonstrates that the same sentence turn may lead to different updates to the self and other perspectives. Yoon and Brown-Schmidt (2019) make a similar point about utterances with polystable meanings, cases where language is intended to give rise to multiple meanings, as in metaphor or private keys (Fleming & Darley, 1991; Mason, 2004; Clark & Schaefer, 1987; van Boven, Kruger, Savitsky, & Gilovich, 2000). The case of the Wet Balcony demonstrates that polystable meanings can also arise when conversational partners draw different *inferences* due to different background knowledge.

The distinct updates are required in order to account for potentially distinct effects at a later time. To illustrate, imagine that, upon departure, your guests select a gift from an eccentric gift basket that includes an umbrella, breath mints, and some watercolor paintings by a local artist. While you—the host—would have chosen a watercolor painting, we predict you would expect the guests to be more likely to choose an umbrella (to protect from the rain) as compared to a painting (which would be ruined in the rain). At an even later time, the two representations may decay independently in memory, such that in some cases only one meaning may be accessible in memory. This prediction is inspired by findings from the memory literature that representations of one's own experiences and the experiences of others can become conflated over time (Principe & Schindewolf, 2012; Coman et al., 2016; Landau & Marsh, 1997; Hyman, Roundhill, Werner, & Rabiroff, 2014). This could result in a situation where you erroneously recall that it was raining on the night of the party. Note, furthermore, that because of the memory benefit for generated content in conversation, we make a further prediction that you would be more likely to make such a memorial error than your partner who produced the original utterance. Future empirical research will examine these predictions with respect to memory and action.

4.3. The emergence of asymmetries over time

In theories that encode mutual knowledge, whether as common ground or via tagging or activation of shared knowledge, memory for shared experiences is expected to be the same for both partners over time. In the Multiple Perspectives Theory, experiences that are shared with another person are encoded in the *separate* representations of self and other. Because of systematic biases in how information is encoded and recalled (Section 2.3), the Multiple Perspectives Theory makes the prediction that recent information in the conversation will likely be similar among conversational partners, whereas retrieval of more distant memories may be asymmetric, with each partner recalling more of what they said themselves, an asymmetry that can grow over time. We illustrate this prediction by considering the following exchange:

Abbi: I am going to have to leave a little early today. I have a ride booked.
Ilana: A ride?

Abbi: Yeah at Peloton, with my favorite instructor Christine D'Ercole
Ilana: What sort of music does she play?
Abbi: I like her 80's playlists, things like Pat Benatar…
Ilana: Oh, that sounds like fun.

Over a short time-scale—at the end of that same conversation—the following utterance would likely be felicitous:

Abbi: *OK, gotta go to my 80's ride.*

What is predicted over a longer time-scale? Consider that after a delay, conversational partners can accurately recall less than 40% of the specific ideas expressed in a conversation, with a bias toward one's own contributions. Thus, if Abbi and Ilana each expressed 100 ideas in their conversation, and each recalled only 40 after a delay, with a bias such that 3/4 of the recalled ideas were self-generated, then only 10 out of the original 200 ideas would be coremembered (and this liberally assumes they happen to coremember the same ideas). A week later, when Abbi produces *my 80's ride*, she may assume this definite to be easily interpretable to Ilana, whereas, in fact, only about 25% of those definites are contextually supported from Ilana's perspective, with the remaining 75% having to be accommodated, namely, requiring the addressee to fill in the details (Lewis, 1969). Indeed, Poesio and Viera (1998) report that only 30% of definites in a corpus were clearly anaphoric. We predict that the combination of limited and biased memory, along with egocentric biases in assumptions about others' knowledge, will lead to situations where speakers produce utterances that they assume addressees will readily interpret, but that actually require addressees to rely on inference to achieve interpretation. Future empirical research will examine whether such utterances exhibit the expected asymmetry between self and other.

5. Conclusion

In this paper, we have argued that mutual knowledge, whether encoded directly in common ground, retrieved from memory, or from a tagged situation model, captures only a subset of the mental states needed to support communication. In considering existing empirical findings from language (questions and reference) and memory, we argued that a theory of mental states in communication needs to fulfill three additional desiderata: (i) the private information that the other is assumed to hold, (ii) differences in perspective between conversational partners (alongside similarities), and (iii) separate perspectives for the self and other.

The Multiple Perspectives Theory provides a cognitive architecture that includes representations of the self and the other that are maintained and updated separately over time. At the heart of this theory is a process where the representations of self and other are continuously compared during conversation as speaker meaning is computed. The comparison process outputs both similarities between perspectives (including mutual knowledge), and differences between perspectives, which includes both absence of knowledge and other inconsistencies between perspectives. The output of this comparison process is transient representations that

are used in the moment; what is continuously maintained and updated over time is the separate representations of self and other. The separate representations may diverge over time, as supported by empirical findings from the memory literature.

We note that the action in communication lies in exchanging information that is not already shared, meaning that what drives communication are the *differences* between conversational partners, and not mutual knowledge (cf. Johnson-Laird, 1982). At the same time, communication crucially depends on mutual knowledge which is used by speakers to encode messages that can be successfully decoded by their addressee. We also demonstrated that similarities and differences are found within a single utterance: while the utterance as a whole depends on the difference between perspectives (e.g., a speaker asks a question about something the addressee is likely to know but they do not), this utterance may include linguistic elements such as definite descriptions whose form and denotation depend on the ability of both partners to coordinate—namely, a similarity in perspective. Empirically, the Multiple Perspectives Theory leads to novel predictions, expanding the domain of inquiry to a broader range of phenomena, including evidentials, demonstrative pronouns, asymmetries in information updating, and divergence in memory over time. Theoretically, the Multiple Perspectives Theory provides an opportunity to explore connections with other traditions, such as the nonlinguistic literature on ToM and the formal semantics literature that focuses on modeling the conversational record. Thus, by shifting the focus from mutual knowledge to a more comprehensive view of mental states, the Multiple Perspectives Theory creates exciting new avenues for future research.

Acknowledgments

This paper draws on research supported by the Social Sciences and Humanities Research Council of Canada (Insight Grant 435-2016-0742) to Daphna Heller, and by the National Science Foundation under Grant Nos. 19–21492 and 22–17478 to Sarah Brown-Schmidt. For excellent conversations and insightful comments, we would like to thank Craig Chambers, Gail Mauner, Katharina Rohlfing, Leon Bergen, Ray Jackendoff, Roger Levy, and Rachel Ryskin. Special thanks to Amit Almor for the case of the Wet Balcony, and also to Adrian Brasoveanu, Richard Breheny, and Susan Brennan. All errors are our own.

Notes

1 If this shortening is attributed to the increased salience or accessibility of the referent (e.g., Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993), the account would incorrectly predict the same shortening to happen with a new partner.

2 While other types of questions are of course possible (e.g., rhetorical questions), here we constrain ourselves to information questions. We note that, for information questions, it may be sufficient for the speaker to assume that the addressee can help them find a path to the answer. This is illustrated by the felicity of conditional answers (Tellings, 2019). For example, you have just installed a new printer driver and are not sure if you

did this correctly, so you call the I.T. desk.YOU: *Have I installed the printer correctly*? IT: *If the icon shows in your bar, you have*.Here, IT does not know the answer to the question—they do not have visual access to your screen—but they can direct you to the answer. Importantly, this phenomenon relies on one's ability to represent the other's private knowledge.

3 This proposal of an empty workspace (Pickering & Garrod, 2021: p. 150), which they also posit for cases where the partners hear different things (*Mary* vs. *Martha*: p. 154), is rather surprising when considered against other aspects of the Shared Workspace framework. First, having an empty workspace is inconsistent with the argument that the "shared workspace" is NOT a representation (e.g., p. 152), but rather "captures the publicly accessible communicative behaviors, together with relevant aspects of the context" (p. 79) and reflects "information that is in both interlocutors' attention" (p. 149). Second, this proposal is also inconsistent with the spirit of the Shared Workspace framework which aims to eschew any direct representation of the other: determining that the shared workspace is empty critically depends on representing the *different* perspective of the other.

4 Whether theory of mind type reasoning involves specialized processes that operate specifically on mental states, or on propositional representations more generally is debated (see Cohen, Sasaki, & German, 2015), and may prove relevant to whether the proposed comparison processes would be functionally distinct when applied to human mental states versus other types of nonsentient agents.

References

Aikhenwald, A. (2004). *Evidentiality*. Oxford: Oxford University Press.

AnderBois, S. (2014). On the exceptional status of reportative evidentials. in Proceedings of SALT 24, pp. 234–254.

Apperly, I. A. (2018). Mindreading and psycholinguistic approaches to perspective taking: Establishing common ground. *Topics in Cognitive Science*, *10*, 133–139.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Baillargeon, R., Buttelmann, D., & Southgate, V. (2018). Invited commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development*, *46*, 112–124.

Bach, K., & Harnish, R. M. (1979). *Linguistic communication and speech acts*. Cambridge, MA: MIT Press.

Baker, L. J., Hymel, A. M., & Levin, D. T. (2018). Anthropomorphism and intentionality improve memory for events. *Discourse Processes*, *55*(3), 241–255.

Baltaretu, A. A., & Chambers, C. G. (2017). Perceptual uncertainty effect and referential contrast. Poster presented at AMLaP.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, *25*(4), 441–455, https://doi.org/10.1080/01690960903047122

Beaver, D. (2001). *Presupposition and assertion in dynamic semantics*. CSLI Publications.

Bögels, S. (2020). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *Cognition*, *203*, 104347.

Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005). Priming prepositional-phrase attachment during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 468.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*(2), 193–209.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, *2*(4), 331–350. https://doi.org/10.1016/0010-0285(71)90019-3

Breheny, R., & Zheng, X. (2020). Common ground and theory of mind in on-line perspective taking: Arguments for a more limited, but still important, role for 'cognitive common ground'. University College London.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions* (pp. 95–129). Cambridge, MA: MIT Press.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. *Psychology of Learning and Motivation*, *53*, 301–344.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science (Special Issue on Joint Action)*, *1*, 274–291.

Brennan, S. E., & Williams, M. (1995). The feeling of another′ s knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*(3), 383–398.

Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition*, *5*(6), 673–678.

Bromme, R., Jucks, R., & Wagner, T. (2005). How to refer to 'diabetes'? Language in online health advice. *Applied Cognitive Psychology*, *19*(5), 569–586. https://doi.org/10.1002/acp.1099

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, *27*, 62–89.

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*(5), 441–472. https://doi.org/10.1016/0010-0285(87)90015-6

Brown-Schmidt, S. (2009a). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*(5), 893–900.

Brown-Schmidt, S. (2009b). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190.

Brown-Schmidt, S., & Fraundorf, S. (2015). Interpretation of informational questions modulated by joint knowledge and intonational contours. *Journal of Memory and Language*, *84*, 49–74.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–1134. https://doi.org/10.1016/j.cognition.2007.11.005

Brown-Schmidt, S., & Heller, D. (2018). Perspective taking during conversation. In G. Gaskell & S. A. Rueschemeyer (Eds), *Oxford Handbook of Psycholinguistics*. Oxford University Press.

Brown-Schmidt, S., & Tanenhaus, M.K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*, 643–684.

Campos, L., & Alonso-Quecuty, M. (2006). Remembering a criminal conversation: Beyond eyewitness testimony. *Memory*, *14*(1), 27–36.

Chafe, W., & Nichols, J. (1986). *Evidentiality: The linguistic coding of epistemology*. Norwood: Ablex Publishing Corporation.

Chambers, C. G., & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition*, *108*(1), 26–50.

Christiaansen, R. E. (1980). Prose memory: Forgetting rates for memory codes. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 611–619. https://doi.org/10.1037/0278-7393.6.5.611

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and brain sciences*, 39.

Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition*, *64*, 1–37.
Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81.
Clark, H. H., & Marshall, C. R. (1978). Reference diaries. In D. L. Waltz (Ed.), *TINLAP-2: Theoretical issues in natural language processing-2* (pp. 57–63). New York: Association for Computing Machinery.
Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge: Cambridge University Press.
Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language*, *26*(2), 209–225. https://doi.org/10.1016/0749-596X(87)90124-0
Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*(2), 259–294. https://doi.org/10.1207/s15516709cog1302_7
Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39.
Cohen, A. S., Sasaki, J. Y., & German, T. C. (2015). Specialized mechanisms for theory of mind: Are mental representations special because they are mental or because they are representations? *Cognition*, *136*, 49–63.
Coman, A., Momennejad, I., Drach, R. D., & Geana, A. (2016). Mnemonic convergence in social networks: The emergent properties of cognition at a collective level. *Proceedings of the National Academy of Sciences*, *113*(29), 8171–8176.
Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, *30*(2), 234–250.
Conway, M. A., Cohen, G., & Stanhope, N. (1991). On the very long-term retention of knowledge acquired through formal education: Twelve years of cognitive psychology. *Journal of Experimental Psychology: General*, *120*(4), 395–409. https://doi.org/10.1037/0096-3445.120.4.395
Cooney, G., Gilbert, D. T., & Wilson, T. D. (2017). The novelty penalty: Why do people like talking about new experiences but hearing about old ones? *Psychological Science*, *28*(3), 380–394. https://doi.org/10.1177/0956797616685870
Covington, N. V., Kurczek, J., Duff, M. C., & Brown-Schmidt, S. (2020). The effect of repetition on pronoun resolution in patients with memory impairment. *Journal of clinical and experimental neuropsychology*, *42*(2), 171–184.
Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The self-organization of human interaction. *Psychology of Learning and Motivation*, *59*, 43–95.
Davis, C. M. (2011). *Constraining interpretation: Sentence-final particles in Japanese* (Ph.D. thesis), University of Massachusetts Amherst.
Deschrijver, E., & Palmer, C. (2020). Reframing social cognition: Relational versus representational mentalizing. *Psychological Bulletin*, *146*(11), 941–969. https://doi.org/10.1037/bul0000302
Dorrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, *46*, 12–30.
Drayton, L. A., & Santos, L. R. (2016). A decade of theory of mind research on Cayo Santiago: Insights into rhesus macaque social cognition. *American Journal of Primatology*, *78*(1), 106–116.
Dryer, M. S. (2013). Position of interrogative phrases in content questions (Chapter 93). In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (pp. 378–379). Leipzig: Max Planck Institute for Evolutionary Anthropology.
Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience*, *9*(1), 140–146.
Duff, M. C., & Brown-Schmidt, S. (2017). Hippocampal contributions to language use and processing. In D. Hannula & M. Duff (Eds.), *The hippocampus from cells to systems* (pp. 503–536). Cham: Springer.
Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*(4), 374–380.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339. https://doi.org/10.1037/0022-3514.87.3.327

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*(6), 760–768. https://doi.org/10.1016/j.jesp.2004.02.002

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864.

Evans, N., Bergqvist, H., & San Roque, L. (2018a). The grammar of engagement I: Framework and initial exemplification. *Language and Cognition*, *10*, 110–140. https://doi.org/10.1017/langcog.2017.21

Evans, N., Bergqvist, H., & San Roque, L. (2018b). The grammar of engagement II: Typology and diachrony. *Language and Cognition*, *10*, 141–170.

Eyal, T., Steffel, M., & Epley, N. (2018). Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of Personality and Social Psychology*, *114*(4), 547–571. https://doi.org/10.1037/pspa0000115

Faller, M. (2019). The discourse commitments of illocutionary reportatives.

Farkas, D. F., & Bruce, K. B. (2010). On reacting to assertions and polar questions. *Journal of Semantics*, *27*(1), 81–118.

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*, 351–386. https://doi.org/10.1111/j.1551-6709.2009.01090.x

Ferguson, H. J., Scheepers, C., & Sanford, A. J. (2010). Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes*, *25*(3), 297–346. https://doi.org/10.1080/01690960903041174

Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology*, *48*, 257–263.

Fischer, N. M., Schult, J. C., & Steffens, M. C. (2015). Source and destination memory in face-to-face interaction: A multinomial modeling approach. *Journal of Experimental Psychology: Applied*, *21*(2), 195–204. http://dx.doi.org/10.1037/xap0000046

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*(1), 99–103. http://dx.doi.org/10.1037/0012-1649.17.1.99

Fleming, J. H., & Darley, J. M. (1991). Mixed messages: The multiple audience problem and strategic communication. *Social Cognition*, *9*, 25–46.

Foley, M. A., Foley, H. J., Durley, J. R., & Maitner, A. T. (2006). Anticipating partners' responses: Examining item and source memory following interactive exchanges. *Memory & Cognition*, *34*(7), 1539–1547.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998–998.

Fussell, S. R., & Krauss, R. M. (1991). Accuracy and bias in estimates of others' knowledge. *European Journal of Social Psychology*, *21*(5), 445–454.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, *62*(3), 378–391. https://doi.org/10.1037/0022-3514.62.3.378

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*(1), 35–51.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218.

Gibbs, Jr, R. W., Mueller, R. A., & Cox, R. W. (1988). Common ground in asking and understanding questions. *Language and Speech*, *31*(4), 321–335.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*(1), 173–184.

Gopie, N., & MacLeod, C. M. (2009). Destination memory: Stop me if I've told you this before. *Psychological Science*, *20*, 1492–1499.

Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (2013). What's learned together stays together: Speakers' choice of referring expression reflects shared experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 843–853. https://doi.org/10.1037/a0029467

Grice, H. P. (1975). *Logic and conversation*. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789004368811_003

Groenendijk, J. A. G., & Stokhof, M. J. B. (1984). Studies on the semantics of questions and the pragmatics of answers [PhD thesis], University of Amsterdam.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307. https://doi.org/10.2307/416535

Gunlogson, C. (2003). *True to form: Rising and falling declaratives as questions in English*. Routledge.

de Haan, F. (2013). Semantic distinctions of evidentiality (Chapter 77). In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, *10*(1), 41–53.

Hankamer, J., & Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry*, *7*(3), 391–428.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61. https://doi.org/10.1016/S0749-596X(03)00022-6

Hartwright, C. E., Apperly, I. A., & Hansen, P. C. (2012). Multiple roles for executive control in belief–desire reasoning: Distinct neural networks are recruited for self perspective inhibition and complexity of reasoning. *NeuroImage*, *61*(4), 921–930.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, *16*(2), 114–121.

Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive Science*, *45*, e12926.

Heim, I. (1982). The semantics of definite and indefinite NPs. *University of Massachusetts at Amherst dissertation*.

Heller, D. (2020). The production and comprehension of referring expressions: Definite descriptions. *Language and Linguistic Compass*, *14*, e12370. https://doi.org/10.1111/lnc3.12370

Heller, D., Gorman, K. S., & Tanenhaus, M. (2012). To name or to describe: Shared knowledge affects referential form. *Topics in Cognitive Science*, *4*(2), 290–305. https://doi.org/10.1111/j.1756-8765.2012.01182

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836. https://doi.org/10.1016/j.cognition.2008.04.008

Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition*, *149*, 104–120. https://doi.org/10.1016/j.cognition.2015.12.008

Heritage, J. (2012a). Epistemics in action: Action formation and territories of knowledge. *Research on Language and Social Interaction*, *45*(1), 1–29.

Heritage, J. (2012b). The epistemic engine: Sequence organization and territories of knowledge. *Research on Language and Social Interaction*, *45*(1), 30–52.

Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science*, *17*(5), 647–659.

Hilliard, C., O'Neal, E., Plumert, J., & Cook, S. W. (2015). Mothers modulate their gesture independently of their speech. *Cognition*, *140*, 89–94.

Hilliard, C., & Cook, S. W. (2016). Bridging gaps in common ground: Speakers design their gestures for their listeners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 91–103.

Horton, W. S., & Brennan, S. E. (2016). The role of metarepresentation in the production and resolution of referring expressions. *Frontiers in Psychology*, *7*, Article 1111. https://doi.org/10.3389/fpsyg.2016.01111

Horton, W. S., & Gerrig, R. J. (2005). Conversational common ground and memory processes in language production. *Discourse Processes*, *40*(1), 1–35. https://doi.org/10.1207/s15326950dp4001_1

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the memory-based processing approach to common ground. *Topics in Cognitive Science*, *8*, 780–795. https://doi.org/10.1111/tops.12216

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117. https://doi.org/10.1016/0010-0277(96)81418-1

Hsee, C. K., Yang, Y., & Li, X. (2019). Relevance insensitivity: A new look at some old biases. *Organizational Behavior and Human Decision Processes*, *153*, 13–26.

Hyman, Jr, I. E., Roundhill, R. F., Werner, K. M., & Rabiroff, C. A. (2014). Collaboration inflation: Egocentric source monitoring errors following collaborative remembering. *Journal of Applied Research in Memory and Cognition*, *3*(4), 293–299.

Isaacs, E. A. (1990). Mutual memory for conversation [Doctoral dissertation], Stanford University.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, *116*(1), 26.

Jaeger, C. B., & Levin, D. T. (2016). If asimo thinks, does roomba feel?: The legal implications of attributing agency to technology. *Journal of Human–Robot Interaction*, *5*(3), 3–25.

Johnson, M., & Demiris, Y. (2005). Perceptual perspective taking and action recognition. *International Journal of Advanced Robotic Systems*, *2*(4), 32.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67.

Johnson-Laird, P. (1982). Mutual ignorance: Comments on Clark and Carlson paper. In N. V. Smith (Ed.), *Mutual knowledge* (pp. 40–45). Academic Press.

Kampis, D., & Southgate, V. (2020). Altercentric cognition: How others influence our cognitive processing. *Trends in Cognitive Science*, *24*, 945–959.

Keenan, J. M., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 549–560.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38. https://doi.org/10.1111/1467-9280.00211

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7

Kintsch, W., & Bates, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(2), 150–159. https://doi.org/10.1037/0278-7393.3.2.150

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin & Review*, *21*(6), 1590–1599. https://doi.org/10.3758/s13423-014-0620-7

Knutsen, D., & Le Bigot, L. (2021). Estimating each other's memory biases in dialogue. *Discourse Processes*, *58*(2), 155–176. https://doi.org/10.1080/0163853X.2020.1837541

Korotkova, N. (2016). *Heterogeneity and universality in the evidential domain* [PhD thesis], UCLA.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*, 113–114.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*(3), 343–346. https://doi.org/10.1037/h0023705

Krauss, R. M., Garlock, C. M., Bricker, P. D., & McMahon, L. E. (1977). The role of audible and visible back-channel responses in interpersonal communication. *Journal of Personality and Social Psychology*, *35*(7), 523.

Küntay, A., & Özyürek, A. (2006). Learning to use demonstratives in conversation: What do language specific strategies in Turkish reveal? *Journal of Child Language*, *33*, 303–320.

Landau, J. D., & Marsh, R. L. (1997). Monitoring source in an unconscious plagiarism paradigm. *Psychonomic Bulletin & Review*, *4*(2), 265–270.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*, 78–106.

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, *9*, 550–557.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671.

MacWhinney, B., Keenan, J. M., & Reinke, P. (1982). The role of arousal in memory for conversation. *Memory and Cognition*, *10*, 308–317. https://doi.org/10.3758/BF03202422

McCready, E. (2005). *The dynamics of particles* [Doctoral Dissertation], University of Texas Austin.

McKinley, G. L., Brown-Schmidt, S., & Benjamin, A. S. (2017). Memory for conversation and the development of common ground. *Memory & Cognition*, *45*, 1281–1294.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213. https://doi.org/10.1016/S0749-596X(03)00028-7

Miller, J. B., deWinstanley, P., & Carey, P. (1996). Memory for conversation. *Memory*, *4*(6), 615–631.

Mozuraitis, M., Chambers, G. C., & Daneman, M. (2015). Privileged vs. shared knowledge about object identity in real-time referential processing. *Cognition*, *142*, 148–165.

Mozuraitis, M., Stevenson, S., & Heller, D. (2018). Modeling reference production as the probabilistic combination of multiple perspectives. *Cognitive Science*, *42*, 974–1008. https://doi.org/10.1111/cogs.12582

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x

Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *Neuroimage*, *184*, 161–170.

Nigro, G., & Neisser, U. (1983). Point of view in personal memories. *Cognitive Psychology*, *15*(4), 467–482.

Noh, E. J. (1998). Echo questions: Metarepresentation and pragmatic enrichment. *Linguistics and Philosophy*, *21*, 603–628.

Nurmsoo, E., & Bloom, P. (2008). Preschoolers' perspective taking in word learning: Do they blindly follow eye gaze? *Psychological Science*, *19*(3), 211–215. https://doi.org/10.1111/j.1467-9280.2008.02069.x

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, *77*, 257–273.

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

Osgood, C. E. (1971). Exploration in semantic space: A personal diary. *Journal of Social Issues*, *27*, 5–64.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*, 89–110.

Pezdek, K., & Prull, M. (1993). Fallacies in memory for conversations: Reflections on Clarence Thomas, Anita Hill, and the Like. *Applied Cognitive Psychology*, *7*, 299–310. https://doi.org/10.1002/acp.2350070404

Phillips, J., & Norby, A. (2019). Factive theory of mind. *Mind & Language*, May, 1–57.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525x04000056

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci. 36*, 329–392.

Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge University Press. https://doi.org/10.1017/9781108610728

Poesio, M., & Viera, R. (1998). A corpus-based investigation of definite description use. *Computational Linguistics*, *24*(2), 183–216.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, *18*, 407–413.

Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science*, *33*, 1468–1482.

Richardson, D. C., Dale, R., & Shockley, K. (2008). Synchrony and swing in conversation: Coordination, temporal dynamics, and communication. In I. Wachsmuth, M. Lenzen, & G. Knoblich (Eds.), *Embodied communication in humans and machines* (pp. 75–93). Oxford: Oxford University Press.

Roberts, C. (1996/2004). Information structure in discourse: Towards an integrated formal theory of pragmatics.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, *37*(3), 322–336. http://dx.doi.org/10.1037/0022-3514.37.3.322

Rubin, R., Brown-Schmidt, S., Duff, M., Tranel, D., & Cohen, N. (2011). How do I remember that I know you know that I know? *Psychological Science*, *22*(12), 1574–1582. https://doi.org/10.1177/0956797611418245

Ryskin, R., Stevenson, S., & Heller, D. (2020). Probabilistic weighting of perspectives in dyadic communication. *Proceedings of the 42nd Cognitive Science Society*.

Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory & Cognition*, *2*(1), 95–100.

Samp, J. A., & Humphreys, L. R. (2007). "I said what?" Partner familiarity, resistance, and the accuracy of conversational recall. *Communication Monographs*, *74*(4), 561–581.

Samson, D., Apperly, I. A., Braithwaite, J., Andrews, B., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266.

Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*, *47*(1), 269–273. https://doi.org/10.1016/j.jesp.2010.09.005

Schiffer, S. R. (1972). *Meaning*. Oxford University Press.

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2011). Eye movements reveal sustained spontaneous processing of others' mental states. *Journal of Experimental Psychology: General*, *141*(3), 433–438.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232. https://doi.org/10.1016/0010-0285(89)90008-X

Searle, J. (1969). *Speech acts*. Cambridge University Press.

Sikos, L., Tomlinson, S. B., Heins, C., & Grodner, D. J. (2019). What do you know? ERP evidence for immediate use of common ground during online reference resolution. *Cognition*, *182*, 275–285.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*(1), 25–38.

Sperber, D., & Wilson, D. (1982). Mutual knowledge and relevance in theories of comprehension. In N. V. Smith (Ed.), *Mutual knowledge* (pp. 61–85). Academic Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. *142*). Cambridge, MA: Harvard University Press.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd edition, pp. 2–9). Oxford/Cambridge: Blackwell Publishers.

Stafford, L., & Daly, J. A. (1984). Conversational memory. *Human Communication Research*, *10*(3), 379–402. https://doi.org/10.1111/j.1468-2958.1984.tb00024.x

Stafford, L., Burggraf, C. S., & Sharkey, W. F. (1988). Conversational memory: The effects of time, recall mode, and memory expectancies on remembrances of natural conversations. *Human Communication Research*, *14*, 203–229.

Stalnaker, R. C. (1978). Assertion. *Syntax and Semantics*, *9*, 315–332.

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43–52.

Surtees, A. D., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology*, *30*(1), 75–86.

Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences*, *109*(21), 8038–8043.

Tanenhaus, M. K., & Carlson, G. N. (1990). Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes*, *5*, 257–280.

Tellings, J. (2019). On the utility of conditional answers. Poster presented at Rational Approaches In Language Sciences (RAILS). Saarbruecken, Germany.

Thornton, M. A., & Tamir, D. I. (2020). People represent mental states in terms of rationality, social impact, and valence: Validating the 3d Mind Model. *Cortex*, *125*, 44–59.

Tullis, J. G., & Benjamin, A. S. (2012). The effectiveness of updating metacognitive knowledge in the elderly: Evidence from metamnemonic judgments of word frequency. *Psychology and Aging*, *27*(3), 683–690. https://doi.org/10.1037/a0025838

Tullis, J. G., & Fraundorf, S. H. (2017). Predicting others' memory performance: The accuracy and bases of social metacognition. *Journal of Memory and Language*, *95*, 124–137.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Ünal, E., & Papafragou, A. (2016). Production–comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language*, *89*, 179–199.

Ünal, E., & Papafragou, A. (2020). Relations between language and cognition: Evidentiality and sources of knowledge. *Topics in Cognitive Science*, *12*, 115–135.

van Boven, L., Kruger, J., Savitsky, K., & Gilovich, T. (2000). When social worlds collide: Overconfidence in the multiple audience problem. *Personality and Social Psychology Bulletin*, *26*(5), 619–628.

Vanlangendonck, F., Willems, R. M., Menenti, L., & Hagoort, P. (2016). An early influence of common ground during speech planning. *Language, Cognition and Neuroscience*, *31*(6), 741–750.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N. J., Fink, G. R., & Zilles, K. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *Neuroimage*, *14*(1), 170–181.

Wachsmuth, I. (2008). 'I, Max'-Communicating with an artificial agent. In I. Wachsmuth & G. Knoblich (Eds.), *Modeling communication with robots and virtual humans* (pp. 279–295). Berlin, Heidelberg: Springer.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody. *Language and Cognitive Processes*, *25*, 905–945.

Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1466–1481.

Westra, E. (2017). Spontaneous mindreading: A problem for the two-systems account. *Synthese*, *194*(11), 4559–4581.

Westra, E., & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, *158*, 165–176. https://doi.org/10.1016/j.cognition.2016.10.021

Westra, E., & Nagel, J. (2021). Mindreading in conversation. *Cognition*, *210*, 104618. https://doi.org/10.1016/j.cognition.2021.104618

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*(2), 183–194. https://doi.org/10.1016/0749-596X(92)90010-U

Willett, T. (1988). A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language*, *12*(1), 51–97.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Wittgenstein, L. (1953). *The philosophical investigations*. Oxford: Blackwell.

Wu, S., & Keysar, B. (2007). The effect of information overlap on communication effectiveness. *Cognitive Science*, *31*(1), 169–181.

Yoon, S. O., Benjamin, A. S., & Brown-Schmidt, S. (2016). The historical context in conversation: Lexical differentiation and memory for the discourse history. *Cognition*, *154*, 102–117.

Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *40*, 919–937.

Yoon, S. O., & Brown-Schmidt, S. (2018). Aim low: Mechanisms of audience design in multiparty conversation. *Discourse Processes*, *55*(7), 566–592.

Yoon, S. O., & Brown-Schmidt, S. (2019). Audience design in multiparty conversation. *Cognitive Science*, *43*(8), e12774.

Yoon, S. O., Duff, M. C., & Brown-Schmidt, S. (2017). Learning and using knowledge about what other people do and don't know despite amnesia. *Cortex*, *94*, 164–175.

Yoon, S. O., & Stine-Morrow, E. A. L. (2019). Evidence of preserved audience design with aging in interactive conversation. *Psychology and Aging*, *34*(4), 613–623. https://doi.org/10.1037/pag0000341

Zipf, G. K. (1935). *The psychobiology of language*. London: Routledge.