Mindreading in Conversation

Evan Westra[1] & Jennifer Nagel[2]

Department of Philosophy, York University[1]

Department of Philosophy, University of Toronto[2]

This research was supported by Social Sciences and Humanities Research Council Postdoctoral Fellowship #756-2018-0012, Social Sciences and Humanities Research Council Insight Grant #435-2017-1041, and the Schwartz Reisman Institute for Technology and Society at the University of Toronto.

Correspondence concerning this article should be addressed to Evan Westra, Department of Philosophy, York University, S448 Ross Building, 4700 Keele Street, Toronto, Ontario, Canada, M3J 1P3. Contact: ewestra@yorku.ca

Abstract

How is human social intelligence engaged in the course of ordinary conversation? Standard models of conversation hold that language production and comprehension are guided by constant, rapid inferences about what other agents have in mind. However, the idea that mindreading is a pervasive feature of conversation is challenged by a large body of evidence suggesting that mental state attribution is slow and taxing, at least when it deals with propositional attitudes such as beliefs. Belief attributions involve contents that are decoupled from our own primary representation of reality; handling these contents has come to be seen as the signature of full-blown human mindreading. However, mindreading in cooperative communication does not necessarily demand decoupling. We argue for a theoretical and empirical turn towards "factive" forms of mentalizing here. In factive mentalizing, we monitor what others do or do not know, without generating decoupled representations. We propose a model of the representational, cognitive, and interactive components of factive mentalizing, a model that aims to explain efficient real-time monitoring of epistemic states in conversation. After laying out this account, we articulate a more limited set of conversational functions for nonfactive forms of mentalizing, including contexts of meta-linguistic repair, deception, and argumentation. We conclude with suggestions for further research into the roles played by factive versus nonfactive forms of mentalizing in conversation.

Keywords: mentalizing; conversation; factivity; knowledge; decoupling.

1. Introduction

The capacity for human beings to thrive and flourish is intimately linked to the capacity for complex forms of coordination and communication (Tomasello, 2019; Vygotsky, 1980). Among different cognitive capacities supporting human interaction, few have occasioned more disagreement than *mentalizing* (or *mindreading*, as we shall also refer to it), our ability to track the underlying psychological causes of behavior (Carruthers & Smith, 1996). Some researchers see mentalizing as a ubiquitous feature of social interaction, underpinning everything from communication (Scott-Phillips, 2014) to moral judgment (Young & Waytz, 2013). Others argue that the role of mindreading has been greatly exaggerated, assigning it a sharply limited function (Andrews, 2012; Gallagher, 2001; Heyes, 2018; Zawidzki, 2013). They credit human social success to various other capacities, from low-level "submentalizing" processes (Heyes, 2014) to more specialized abilities such as trait recognition and stereotyping (Andrews, 2012). Thus, depending on who you ask, mentalizing is either the core of the explanation of how human beings navigate the social world, or else an overblown distraction from the real story.

Nowhere is this dynamic clearer than when cognitive scientists debate the psychological underpinnings of one of our most striking forms of social interaction: conversation. According to a broadly influential framework traceable to the work of H.P. Grice (Grice, 1975), conversation is mentalistic through and through (Clark & Marshall, 1981; D. W. Harris, 2019; Heritage, 2012a; Sperber & Wilson, 2002). In this framework, whether conversational partners are making inferences about speaker meaning, selecting and interpreting referring expressions, or trying to figure something out together, they are constantly engaged in the task of monitoring and updating representations of each other's mental states. Consequently, the Gricean view entails that mentalizing has a number of fairly specific cognitive properties: To match the speed of

ordinary conversation, mentalistic inferences and adjustments must occur in an extremely rapid, online fashion. To handle the wide array of different contents that we must represent in order to communicate effectively, conversational mentalizing must be highly flexible and have access to a wide array of stored information and propositional reasoning processes. To match the phenomenology of everyday conversation, these mentalistic processes must occur unconsciously and without significant cognitive effort.

This picture conflicts with a widespread view within the anti-mentalistic camp: mentalizing is *hard*. The kind of mentalizing that involves propositional contents is thought to be cognitively effortful in ways that make it an unlikely basis for everyday social interaction (Bermudez, 2003; Butterfill & Apperly, 2013; Moore, 2017; Zawidzki, 2013). Researchers on this side of the debate point to a large empirical literature showing that performance on standard developmental measures of mindreading is reliably correlated with measures of executive functioning (Devine & Hughes, 2014). The picture of mindreading that emerges from this body of work shows a slow and taxing form of cognition that is ill-suited to the demands of everyday conversation. Anti-mentalistic models of conversation consequently relegate mindreading to a minor, supporting role, for example, as a device for repairing certain kinds of misunderstandings (Apperly, 2018; Pickering & Garrod, 2004; Shintel & Keysar, 2009). These proposals are challenged, however, by a wide array of evidence indicating that speakers are sensitive to information about speakers' mental states from the very earliest stages of language production and comprehension (Brennan & Hanna, 2009; Hanna et al., 2003; Heller et al., 2012; Rubio-Fernández et al., 2019). It is not obvious how to reconcile this latter research with the evidence suggesting that mindreading is slow and cognitively demanding. It would seem, then, that

mentalizing both cannot and yet somehow does play a significant role in the cognitive underpinnings of conversation.

This puzzling standoff about what is happening in conversation motivates a fresh look at the kind of mentalizing that involves propositional contents. We argue that there are actually two kinds of mentalizing at work here: Factive mentalizing, which is relatively easy and pervasive in conversation, and nonfactive mentalizing, which is relatively difficult and rare. Factive mentalizing is employed when we represent others as either knowing or failing to know about some aspect of reality. Nonfactive mentalizing governs the attribution of beliefs whose contents are decoupled from the attributor's larger representation of what is real, which we will refer to as their *primary representation*. Re-examining empirical work from both sides of the divide, we find that research establishing the difficulty of mental state attribution describes nonfactive mentalizing, where research establishing its pervasiveness focuses on factive mentalizing. The depth of this distinction has been missed because the signature feature of the harder kind of mentalizing—"decoupling"—has been mistaken for the mark of genuine mental state attribution as such. This confusion has arisen in part because tests involving decoupling have had such a prominent role in measures of mentalizing, and in part because as theorists, we always have the option of taking a decoupled perspective. However, the kind of mentalizing that underpins most conversational activity is powerful and pervasive exactly because it forbids decoupling: Typical cooperative conversations are expected to be mutually informative, in a way that simply aligns the task of figuring out what is real with the task of figuring out what the other person has in mind. Seeing others as knowledgeable enables us to take what they say at face value without further reasoning: It is exactly in virtue of having this as our default setting in cooperative conversations that these conversations are an efficient way of pooling knowledge. Mentalizing in

conversation is important because we do need to represent epistemic disparities in order to ask questions that our conversational partners are well-positioned to answer, and to avoid telling them what they already know; however, these basic conversational abilities demand committed rather than decoupled forms of mental state attribution.

Our reexamination of the status of decoupling in mental state attribution and communication is grounded in the distinction between *factive* (knowledge-oriented) and *nonfactive* (belief-oriented) mentalizing. In recent years, a number of theorists have argued that knowledge attribution is ontogenically, phylogenically, and cognitively prior to belief attribution, and should play a more central role in our theories of mindreading (Nagel, 2017; Phillips et al., 2020; Phillips & Norby, 2019). We extend these earlier proposals in several ways. By giving an account of the key mindreading distinction in terms of decoupling, we provide a cognitive mechanism that shows *why* it is that knowledge attribution is prior to belief attribution, and *how* these different types of mental state attribution are represented. We also explain how factive mentalizing enables us to engage in fast and efficient mindreading in cooperative conversational contexts. Finally, we offer a theoretical proposal about conversational contexts that *do* require nonfactive mentalizing, including repair, strategic deception, and argumentation.

From the outset we should be clear that even outside of these special contexts, mindreading involves much more than the detection of knowledge and belief: a full treatment of mindreading in conversation would have to cover attributions emotional and motivational states as well. Explaining these processes falls outside the scope of our current project. In what follows, we focus on attributions of epistemic states; explaining how these states are tracked in conversation is no small matter.

[Box 1]

2. Two kinds of propositional attitude attribution

2.1. The significance of decoupling

It is famously hard to tell whether a response to a situation involving an agent reflects a grasp of that agent's mental states. When shown a video of a man struggling to obtain out-of-reach bananas, or struggling to unlock a cage, Premack and Woodruff's chimpanzee Sarah was able to select a photo showing the appropriate tool for the job; they took this to indicate her sensitivity to the agent's state of mind (Premack & Woodruff, 1978). Critics were swift to point out that Sarah's selections could also be explained by her primary grasp of the environmental problems—distant bananas, a locked door—without reference to the agent. They advocated a test introducing some contrast between the environment and the mental states of the observed agent, such as a false belief task in which the observed agent's representation of the location of some reward is at odds with the current fact of the matter (Dennett, 1978; Harman, 1978).

When there is a clash between the content of the observed agent's belief and the observer's primary representation of reality, the observer who correctly anticipates the deceived agent's action does show some real evidence of grasping that agent's state of mind (Wimmer & Perner, 1983). To anticipate the reasoned behaviour of the observed agent, the observer must represent the content of that agent's mental state in a way that decouples this clashing content from the observer's primary, action-guiding representation of the world. With this content quarantined from the content of observer's own commitments, the observer can use both streams of content to reason separately about what is going on in the world, and about how the agent will act (Leslie 1987). But even if situations involving recognizably mistaken agents, and consequently decoupled contents, serve as particularly good *tests* for propositional mental state

attribution, this is not to say that all propositional mental state attributions must involve decoupled contents.

Confusion over this point dates back to the 1980s. Alan Leslie lists it as a defining characteristic of proposition-embedding mental state terms that they take decoupled content. According to Leslie, "propositions involving mental state terms do not logically imply the truth (or falsehood) of propositions embedded in them. Thus "John believes the cat is white" says nothing about whether or not the cat really is white. Again, one cannot look through the embedded proposition to the world" (Leslie, 1987, p. 416). Leslie is right about the opacity of "believe" (and "expect", the other epistemic state he mentions in this context), but his rule does not cover all mental state terms. One important class of proposition-embedding mental state terms, namely factives (e.g. *know*, *be aware*, *see that*), is distinguished exactly by requiring the speaker's commitment to the truth of their embedded propositions (Kiparsky & Kiparsky, 1970).[1] For reasons of generality, in what follows we will focus primarily on *know*, the most general factive mental state term; all other factive mental states entail knowing (Williamson, 2000). In contrast to parallel sentences involving nonfactives such as *believe*, *think*, and *expect*, a sentence such as "John knows the cat is white" does commit the speaker to the cat's being white. Factive mental state attributions are transparent, in the sense that the attributor looks through them to the world. The attributor may of course be mistaken, taking someone to know something

[1] Philosophers and linguists have slightly different understandings of this requirement, generating an interdisciplinary discrepancy in the meaning of the term "factive": linguists typically define it in terms of presupposition, while philosophers focus on entailment. This terminological difference matters for communicative expressions like "is right that", which entail but do not presuppose their embedded contents, but it seems that all mental state expressions that entail their complements also presuppose them (Anand & Hacquard, 2014), so for present purposes we can adhere to the stricter philosophical understanding of the term. For discussion, see (Nagel 2017).

that it is not actually a fact, but the meaning of an attribution of knowledge crucially includes the attributor's commitment to the embedded content.

Subsequent researchers have followed Leslie's initial line of insisting on a very strong connection between decoupled content and fully mature propositional mental state attribution. On Michael Tomasello's view of "propositional attitude constructions", these always involve "the coordination of three perspectives—yours, mine and the objective perspective" (Tomasello, 2018, p. 8492). In his view, to grasp a propositional attitude ("He believes that the cat is sick") is a matter of recognizing that there is some fact of the matter about the embedded content ("i.e. the cat is or is not objectively sick"), and furthermore to grasp that "this is independent of the attitude about the fact that the speaker expresses in the main clause." Again, this independence condition does indeed apply to attributions of nonfactive propositional attitudes like belief, which can take decoupled content, but it does not apply to factive attitudes. For the speaker who says, "he knows that the cat is sick," there is no independence between the attitude ascribed in the main clause and the truth of the embedded content; the factive verb locks in the speaker's commitment to what is said to be known.

Tomasello expresses no doubts about the mainstream view that factive verbs such as "know" govern committed rather than decoupled content. Having drawn a tight connection between perspective-taking and decoupled content, what Tomasello denies is that knowledge attributions involve genuine perspective-taking. In his view, states of "seeing, hearing, knowing" are identified as "nonperspectival intentional states" (Tomasello, 2019, p. 52). There is something odd about this classification: seeing, for example, might be thought of as the perspectival state *par excellence*, in that what we see is a function of spatial position and orientation, and appropriate attributions of states of seeing must track those relationships.

Tomasello's theory classifies factive states (or "epistemic states" in his terminology) as nonperspectival because it employs a special sense of "perspectival" in which recognition of another's perspective always demands a three-way comparison between self, other, and reality, with the attributor attending separately to each (2018, 2019, p. 64). In this picture, the task of monitoring reality is always quarantined from the task of monitoring the (always potentially mistaken) perspectival states of others, and the perspectival states of others are to be understood in contrast with one's own.

Indeed, Tomasello insists that even one's own mental states are properly decoupled from reality, given a mature human understanding of perspective. He makes this point in connection with the unwitnessed transfer task, in which an object is moved from a drawer to a cabinet when the observed agent's back is turned. The mature mindreader understands not only that the observed agent is mistaken about the location of the target object; in Tomasello's view, "a fully adult-like understanding would include the proviso that the [subject] herself might potentially be wrong: perhaps the cabinet has a false bottom or someone has tricked her" (2018, p. 8492). Mature tracking of perspective, on this view, requires chronic representation of the possibility of slippage between what is real and what is represented, both by others and by oneself. In a more naïve take on the transfer task, the observed agent has a false belief about the location of the object, while the observing child knows where it is; with Tomasello's proviso in place, the child is no longer simply characterized as knowing. Rather, "the child's best guess is that the objective situation matches her own belief (because she has good evidence, and the agent has misleading evidence)" (2018, p. 8493). Where the naïve attribute knowledge, the sophisticated attribute at most a best guess.

There are costs to demoting knowledge attribution to this second-class status. Even if in theory, adults have the option of decoupling or raising doubts about any particular judgment, we are in practice selective about exercising that option. Corpus data suggest that adults attribute knowledge frequently, with "know" appearing as one of the most common verbs in spoken language, outstripping "think" and "believe" (Davies, 2008). One might take knowledge attribution to be an immature or restricted form of evaluation that adults employ when they lack the time or resources to consider multiple ways of seeing something. Our alternative proposal is that both factive and nonfactive mental state attributions are fully mature, but different in function: our selectivity with decoupling serves a valuable purpose. The next section aims to explain when adults decouple and distance themselves from the contents of attributed propositional attitudes, and when they commit.

2.2. Decoupling and committing

In classic false belief tasks, the observer is looking down on the deceived agent from a position of epistemic superiority, and simply anticipating or reporting that agent's misconception (as opposed to correcting it, the more natural response in cooperative interaction). But participants in ordinary cooperative conversations are as often looking up as looking down, as they switch between the roles of asking and telling. Whenever the two people in a conversational interaction differ in whether they know some proposition, there is what we might call an epistemic gradient between them, to use John Heritage's (2012a) terminology, and mentalizing is useful whether one is at the top or the bottom of that gradient. Indeed, mentalistic models of conversation insist that representing a conversational partner as in the more or less knowledgeable position, relative to oneself, is crucial to the structure of conversation, determining whether one is in a position to ask or tell (Heritage, 2012a, 2012b). In the simple asymmetry of interaction, every moment of

epistemic superiority for one side in an interaction is a moment of inferiority for the other, so it is worth examining the kind of mentalizing we can do in both positions.

Epistemic subordinates take themselves to know less on the relevant point than their conversational partners do. Given that factive mental state attributions involve a commitment on the attributor's side to the content of the other person's state of mind, it might sound odd to say that subordinates can be committed to the content of the superior's state of mind. However, by representing you as possessing propositional knowledge which I lack ("what time is the meeting tomorrow?"), I am motivated to ask you, using your state of mind to serve as my guide to reality. In mentalistic models of conversation, question-asking behavior is guided by just this type of mentalizing; section 4 below explores attributions of knowledge-*wh* in more detail.

When our questions are answered, ordinarily we update our models of the world directly, consistent with our initial factive mental state attributions representing the addressee as knowing the answer to the question. This is our default conversational stance of trust, in which we represent no possibility of divergence between what is said and how things are: You are assumed to know, to be expressing the kind of state of mind that one has only towards truths. Of course, it is possible to take a more cautious approach: I could suspect that you might be mistaken, or might be attempting to ensure that I arrive late to the meeting, but note that this type of stance would make it puzzling why I asked you in the first place. Decoupling also leaves the question unresolved, so further research would be needed to settle my own epistemic state. We cannot directly gain propositional knowledge testimonially unless we represent our conversational informants as knowing (McDowell, 1994).

In practice, we can switch from trusting to decoupling when the content of a conversational partner's answer lies outside the range of what we would take on trust, prompting

a re-evaluation of the attributed mental state or its expression: If you tell me that the meeting is at 5am, I will have second thoughts about whether you know, or hear you as joking. But one of the difficulties with theorizing about decoupling is that it is always possible in principle, at least from the perspective of the theoretician who is not immediately charged with the task of figuring out when the meeting is happening: Looking at virtually any situation involving an agent, we can consider hypothetical or frankly counterfactual possibilities of divergence between the agent's state of mind and the environment. One reason why decoupling is always possible is explained by the philosophical orthodoxy that knowing entails believing: If it is true that Bill knows that the meeting is at noon, then it is also true that Bill believes that the meeting is at noon (Ichikawa & Steup, 2018). The fact that all states of knowledge are states of belief means that when we detect a state of knowledge in another person, this state could also accurately, if less informatively, be classified as a state of belief. We could choose to regard Bill's mental state as a condition local to him, in abstraction from what is going on in the larger world. If we see Bill just as believing that the meeting is at noon, then we are not accepting a falsehood, but we are doing something like overfitting our data, reducing the predictive power of our mental state attribution, and refusing the informational gain available from a knowledgeable conversational partner. Recognizing others as knowing more than we do enables us to interpret their statements and behavior as reflective of the larger reality.

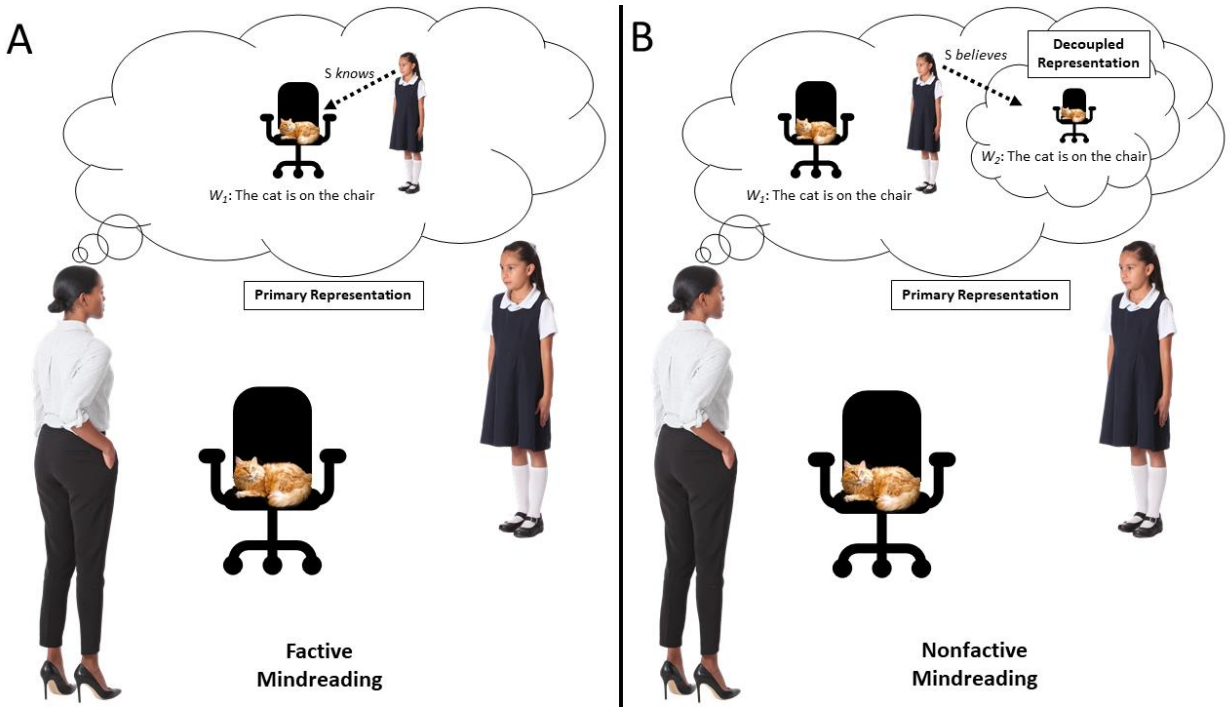
Factive mental state attribution is useful not only when others know more than we do, but also when they know less. If I can see that your view of an event is obstructed, I do not need to reason about what you might believe is happening in order to see you as failing to know. While it is possible for me to think about what decoupled content you might be entertaining, it is not necessary, and indeed mentalistic models of conversation take typical acts of telling to be

motivated just by a sense that the other party lacks knowledge on a given point (Heritage 2012a, 2012b).

3. Coupled and decoupled mentalizing

Nonfactive forms of mindreading are not just functionally redundant in conversational contexts: We suggest that they are also cognitively costly in ways that factive mindreading is not. This cost is a direct consequence of the tight relationship between nonfactive mindreading and representational decoupling; indeed, coordinating between two decoupled sets of representations is precisely why many researchers believe that mindreading is hard (Fizke et al., 2014). Decoupled mentalizing seems to draw upon our limited executive resources in multiple ways (Devine & Hughes, 2014): Maintaining mental representations of another's beliefs alongside one's own is thought to place demands on working memory (Carlson et al., 2002); suppressing our primary representations while making use of the decoupled ones requires inhibitory control (Benson & Sabbagh, 2009); and switching back and forth between one set of representations and another requires cognitive flexibility (Jacques & Zelazo, 2005). All of this makes nonfactive forms of mentalizing an especially effortful mode of social cognition.

Figure 1


A
S knows
W_1: The cat is on the chair
Primary Representation
Factive
Mindreading
B
S believes
Decoupled
Representation
W_2: The cat is on the chair
W_1: The cat is on the chair
Primary Representation
Nonfactive
Mindreading

While decoupled mindreading is required for belief attribution, it is not required for knowledge attribution. Consider a paradigmatic case of knowledge attribution (Figure 1A). You are facing a cat on a chair, and opposite to you stands another agent S, whose line of sight is focused on the cat. The other agent's direction of gaze naturally adds salience to the fact that the cat is on the chair. Because the cat is in your visual field, your grasp of this situation—call it your *primary representation*—includes the fact that the cat is on the chair (we label this element of your primary representation as W_1). Making a spontaneous and unconscious inference from S's visible line-of-sight cues, you attribute to S the mental state of knowing that the cat is on the chair (*S knows that* W_1). The propositional structure of W_1 makes it available for inferences about the world, including mental state inferences. Because factive mental states like knowing can only link agents with features of reality, your representations of knowing are simply structured to connect agents to elements included in your primary representation of reality. The simplicity of this type of attribution carries over to cases of ignorance: if another agent's view of the chair is blocked, you can represent this agent R simply as failing to know W_1, without specifying any independent decoupled content W_n to which R might be committed. You do not need to represent where R might believe the cat is now located in order to see R as not knowing that the cat is in the chair.

It is at least a theoretical option for you to represent the original situation using a decoupled representation instead, following Tomasello's advice of keeping the attributed mental content quarantined from the attributor's primary representation. Instead of seeing S as knowing that the cat is on the chair, you might instead see her as having a belief that the cat is on the chair, where this belief happens to agree with your own belief about the matter (Figure 1B). You could use the observed agent's line of sight as a cue to what she has in mind, imagining how the

world looks from her perspective, tokening a decoupled representation (W_2) with the same content as your primary representation (W_1). Note that this type of attribution obliges you to represent "the cat is on the chair" *twice* (see Figure A), and to keep track of the truth of the agent's state of mind by matching it to the corresponding element in your primary representation. While this duplication of contents might seem redundant and inefficient, a defender of decoupling might argue that it would not place any demands on working memory or inhibitory control: there need be no additional processing cost to generating an additional, decoupled representation W_2 as long as its content does not clash with your primary representation of the world. Decoupling, according to this objection, only becomes effortful in contexts like the false-belief task, where a mindreader must simultaneously maintain two conflicting representations.

However, there is evidence that decoupled *true* belief attribution is also cognitively demanding. Consider a situation in which a state of knowledge lapses into ignorance (for example, when an object originally witnessed by a mindreader and target agent as going into the green box is removed from that location when the agent's back is turned). If the object is then randomly returned to the green box before the agent returns, we have a situation in which there is no clash between the mindreader's primary representation of reality (the object is in the green box) and the content of the decoupled representation properly attributed to the agent (the agent thinks the object is in the green box). In true belief cases like this, known as "Gettier cases" in the philosophical literature (Gettier 1963), the observed agent's mental state is only coincidentally aligned with reality; it no longer constitutes knowledge because it is not formed in a way which can only bind an agent to a truth. Five-year-old children and nonhuman primates seem to treat such cases only as cases of ignorance, failing to make the more specific prediction that the agent will nevertheless perform an accurate reach (Fabricius et al., 2010; Horschler et al.,

2019). Here there is no clash between the mindreader's primary representation of reality and the decoupled content, but the mental state inference is still difficult, suggesting that decoupling itself bears a cost.

In general, the utility of decoupling is most obvious in situations where the mindreader must be epistemically vigilant, because the connection between the other agent's mind and the world has been disrupted in some way, and the two are at risk of diverging. As standard false-belief task scenarios illustrate, decoupling makes sense when there is a relevant possibility that the content of the mental state being attributed is epistemically suspect; in these cases, decoupling prevents the inconsistent or false beliefs of others from contaminating your primary deliberative processes (Fizke et al., 2014; Schuwerk et al., 2014). In more ordinary cases of knowledge tracking like the one we have described, however, there has been no disruption to the epistemic relation between S and the world. To the mindreader, S's knowledge of the fact that the cat is on the chair is as obvious and well-founded as the fact itself.[2] In this context, there is no cause for epistemic vigilance, and no obvious reason to engage in decoupling. It would of course be possible for our mindreader to generate a decoupled representation of this same situation framed in terms of true beliefs, but it would serve no practical use unless S were motivated by an overabundance of epistemic caution. Otherwise, decoupling is completely otiose. In short: in cases where decoupled true-belief attributions serve a functional purpose, they are likely to be executively demanding. In cases where they are unlikely to be executively demanding, decoupled true-belief attributions appear to serve no functional purpose.

[2] We discuss how these intuitive knowledge inferences are made and tracked in sections 4 and 6.

One distinctive feature of this model is that the coupled contents attributed in this kind of factive mindreading need to be *propositional* in format to guide propositionally structured acts of telling and asking, directed at particular agents. In this regard, what is attributed in the factive mindreading model that we are proposing is much richer than what is attributed in various "minimal" models of mentalizing – for example, when applying Butterfill and Apperly's *registration* construct, or Burge's *sensing*, which are represented as simple polyadic relations between agents and objects in their environments, and which do not permit for any sort of relations between agents and propositions (Burge, 2018; Butterfill & Apperly, 2013). In contrast, the structure and content of the factive mental state attributions we describe here mirror the structure and content of our primary representations, because they both make use of the same representational tokens (i.e. W_1). And insofar as our primary representations are propositionally structured, the mental states attributed through factive mindreading will be as well (c.f. Carruthers, 2016, p. 154).[3]

This way of representing mental states has some clear limitations. On its own, factive mindreading does not capture the fact that a person's knowledge of the world might be represented in a particular way that might differ from our own – what some refer to as "Level 2 perspective-taking" or the "aspectuality" of propositional attitudes (Butterfill & Apperly, 2013; Flavell, 1977). It also does not distinguish between various ways of failing to know, between ignorance that consists in simply lacking a belief on a certain point, having a false belief, or having a true belief that fails to rise to the level of knowledge (Gettier, 1963; Plato, 1980). These

[3] Despite the fact that it involves propositional attitude attribution, factive mentalizing is not, strictly speaking, a form of meta-representation, as the latter is generally thought to involve decoupling (Perner, 1991). However, factive mental state attributions could be correctly described as meta-*cognitive*, in that they do involve a kind of higher-order cognition.

limitations are tied to the fact that the contents of factive mental state attributions are coupled with the contents of our primary representations of the world; representing possible or actual divergences from the contents or modes of presentation of one's primary representations requires decoupling. But this absence of decoupling is also precisely how factive mindreading would gain its efficiency, since it would not require us to maintain, inhibit, or shift between multiple sets of mental representations, as in the case of nonfactive mindreading. All that it would require would be that mindreaders keep track of the presence or absence of factive mental relations.

4. Mechanisms for efficient knowledge attribution

Fortunately, the presence or absence of knowledge can be reliably inferred through the use of a number of cognitively efficient heuristics, which can be deployed in a fast and fluid manner over the course of a conversation. For example, one of the most reliable strategies for tracking factive mental states is via the monitoring of other agents' gaze. What an agent can or cannot see can be tracked without representational decoupling through the use of heuristics like tracing an agent's line-of-sight, imaginatively superimposed onto one's own egocentric representation of the world (Michelon & Zacks, 2006). When the line-of-sight between an agent and an object or event is unobstructed, one can safely infer that the agent has some knowledge of that object or event; when line-of-sight is obstructed, no factive relative obtains.

Sensitivity to gaze is widespread in the animal kingdom, and for good reason: monitoring line-of-sight can be adaptively relevant in many social contexts, from strategic competition among conspecifics to avoiding predators and stalking prey (Shepherd, 2010). Human beings, however, are morphologically unique in that our distinctive white sclera and elongated eyes make the direction of our gaze especially easy to follow – a trait that some researchers have

suggested evolved as an adaptation for visually based communication (Kobayashi & Kohshima, 2001; Tomasello et al., 2007). Along these lines, a substantial body of research indicates that sensitivity to gaze direction is an important part of language comprehension and development. For example, tracking speakers' gaze direction facilitates the interpretation of referential expressions (Jachmann et al., 2019; Sekicki & Staudte, 2018). Competence in gaze tracking in the first year of life has also shown to be predictive of later language development (Brooks & Meltzoff, 2008; Tenenbaum et al., 2015), while children at risk for autism spectrum disorder display abnormal patterns of gaze-following (Thorup et al., 2016). We thus have good reason to believe that line-of-sight is steadily monitored in face-to-face conversation, and that this supports successful communication. According to the factive mindreading model that we are proposing, line-of-sight is tracked in conversation because it is a reliable way of discerning what another agent does or does not know, which can in turn inform the way we pursue the epistemic goals of conversation.

On its own, using line-of-sight to infer factive states only helps us to infer what an agent knows about their immediate surroundings. This might be useful if we are actually talking about entities, events, or states of affairs in the shared environment (e.g. "Look how cute the cat is!"). But if our topic of conversation concerns something outside our immediate environment – say a past event, or an abstract proposition – then the factive states inferred via gaze-tracking will be of little use. If factive mindreading is to support the kinds of epistemic inferences we make in everyday conversation, however, then it must have some efficient mechanism for representing factive relations between agents and absent entities and events.

One strong candidate for such a mechanism emerges from the memory-based processing model of common ground reasoning (Gorman et al., 2013; Horton & Gerrig, 2005). On this

model, the assessment of whether or not a piece of information is mutually known and available for conversational purposes normally takes place implicitly, and is supported by ordinary, domain-general memory processes. When we share experiences with a particular person, the information contained in those experiences is stored as episodic memory traces, which become associated with that individual. When we re-encounter that person on a subsequent occasion, their presence will serve as a contextual memory cue. This memory cue causes the information previously associated with that person to become increasingly accessible for language production and comprehension. For example, if in the past you and a particular conversational partner once encountered or discussed the game of cricket, then in subsequent encounters with that partner, their presence will cause stored information about cricket to become highly accessible for conversational purposes. The fluency with which you recall cricket-information in this particular interpersonal context thus serves as an implicit cue that it is mutually known (Horton, 2007).

Even when certain pieces of information have not been encountered with a person before, the same memory processes can still lead us to treat that information as available for conversational purposes when it is associated with some feature of that person's social identity. The fact that a stranger is a fellow local, for example, might lead us to fluently refer to pieces of information that are highly accessible when speaking with other locals – for example, particular neighborhoods or transit lines (Isaacs & Clark, 1987).

Meanwhile, when certain pieces of information are not strongly associated with a particular individual, then that information will not be antecedently accessible for the purposes of conversation. References to this information will feel more uncertain, reflecting more disfluent processing. This serves as an implicit cue that the information in question is privileged, leading speakers to provide additional contextual information to ensure that their referential intention is

clear – for example, by using a definite description instead of or alongside a name instead of using a name by itself (Gorman et al., 2013).

Note that while this process relies upon low-level, domain-general mechanisms, it nevertheless supports our capacity for tracking mental states and flexibly judging what interlocutors do or do not know. In this respect, the model we are proposing is very different from anti-mentalistic proposals where representations of what one's conversational partner knows plays a much more limited role in comprehension and production (Pickering & Garrod, 2004; Shintel & Keysar, 2009). Cues of fluency, on our account, are not directly mapped onto particular kinds of behavioral responses. Rather, these low-level cues are used to generate intervening representations of factive mental relations: their propositional content is called into play to flexibly inform a variety of different comprehension and production processes (Gorman et al., 2013; Hanna et al., 2003).[4] What a speaker does with a factive mental state attribution will ultimately depend upon its conversational relevance (Roberts, 2012; Wilson & Sperber, 2003).

The memory-based processing model provides a plausible mechanism for explaining how factive relations to absent entities, individuals, and past events can be inferred in a fast and efficient manner over the course of a conversation. Just as cues like line-of-sight allow us to rapidly infer what a person knows about their immediate environment, feelings of fluency and disfluency when recalling particular pieces of information in a conversational context provide us with simple cues as to whether or not that information is mutually known. Unlike line-of-sight heuristics, however, memory-based heuristics like these are a somewhat weaker means of

[4] Here, our proposal is indebted to the notion of "one-bit" common ground representations invoked by Galati and Brennan in their account of audience design (Galati & Brennan, 2010). However, our account differs from theirs in that the factive representations in our account are not to be construed in meta-representational terms (c.f. Horton & Brennan, 2016).

ascertaining what another person knows. Simple associative links between an individual and an event might be a moderately reliable indicator that that person was present for that event, or has some significant relation to the topic, but occasionally this heuristic will fail. For example, Horton and Gerrig (2005) suggest that when we strongly associate several individuals with one another, the presence of one individual can serve as a retrieval cue for information that originally co-occurred with another. This can then lead one to mistakenly infer that information known to the first individual will also be known to the second – effectively, an error in source monitoring.

Importantly, the information retrieved through these processes need not have been antecedently stored as knowledge attributions that were made in the past. Some accounts of common ground and knowledge attribution posit that for each individual we encounter, we maintain an "experiential registry" or "reference diary" what a person knows, which we can then consult in the context of conversation (Clark & Marshall, 1981; Perner & Roessler, 2012). While it is certainly possible that we sometimes store explicit representations of who knows what (for example, when monitoring whether or not a person is privy to a secret), such representations are often unnecessary. Instead, treating conversational partners as contextual memory cues allow us to compute their epistemic states on demand. The only stored representations that are required for this to take place are the ones that support ordinary forms of context-sensitive memory retrieval.

In sum, heuristics like gaze-following and memory-based fluency cues show how states of knowledge can be rapidly and efficiently inferred over the course of a conversational interaction. This is by no means an exhaustive list of all the ways that such inferences might occur, but rather a set of illustrative examples that capture the kinds of processes that we suggest undergird factive mindreading in conversation. We also do not deny that knowledge attributions can be based on

explicit, strategic forms of reasoning: as we discuss in section 7, there are a variety of contexts in which our conversational goals might require us to engage in more effortful forms of mindreading. Our main claim here is simply that this kind of reasoning is unlikely to underlie the knowledge attributions that we must regularly make over the course of a conversation. To keep up with the pace with the back and forth of ordinary dialog, our default forms of knowledge attribution must be lean and computationally efficient (Galati & Brennan, 2010; Horton & Brennan, 2016).

5. Representing when others know what you don't

Thus far, this picture of factive mentalizing shows how we can infer when another agent knows about or is ignorant of something that we ourselves know. But this is only part of the story. As we have noted, the epistemic goal of conversation is to extend interlocutors' shared knowledge. Sometimes this means informing others about facts of which they are ignorant, but other times it means seeking out from others knowledge that we ourselves lack. This means representing the fact that someone knows something that we do not, or what Jonathan Phillips and Aaron Norby (2019) call "egocentric ignorance". As it stands, the account of factive mental state representation that we have outlined doesn't provide us with an obvious way to do this. In this section, we show how our factive mindreading framework could be extended to handle these cases.

Our account of tracking egocentric ignorance begins with a peculiar feature of factive mental state verbs: they enable us to express epistemic relations between agents and *questions.* In a positive attribution of knowledge, what is known may be specified directly, with a "that"-clause, as in, "Alice knows that Bill was fired", or indirectly, with an embedded question, as in "Alice knows whether Bill was fired", or "Alice knows who was fired." Attributions of

ignorance (or the lack of knowledge) are similarly capable of taking declarative or interrogative complements ("Charles does not know that Daphne is at home"; "Charles does not know where Daphne is") (Hintikka, 1975). Notably, knowledge-*that* and knowledge-*wh* are not fundamentally different in their content: Both types of attribution are made true by someone's knowledge (or in the negative case, their ignorance) of some fact. In cases of knowledge-*wh*, this fact will be the true answer to the embedded question: if the truth is that Frank bought a car, then to know what Frank bought is to know that Frank bought a car. From an omniscient perspective, all propositional knowledge could be represented as knowledge-*that* or knowledge-*wh*.

From a non-omniscient perspective, however, knowledge-*wh* attributions have a valuable function: They permit us to represent an agent's epistemic state even when we do not know the content of that state ourselves. Positive knowledge-*wh* attributions, for example, enable mindreaders to represent the perspective of an agent whose epistemic position is stronger than their own ("Ella knows what Frank bought, but I still don't"); negative knowledge-*wh* attributions permit mindreaders to represent information that is privately or mutually unknown ("Ella does not know what Frank bought, and I don't either").

In belief attribution, we are only able to directly specify the true or false proposition believed with a that-clause, as in "Greta thinks that Hank has quit." Crucially, belief-attributing verbs do not take embedded questions as complements; we cannot formulate an indirect belief attribution by saying, "Greta believes *whether* Hank has quit", or "Greta thinks *what* Hank did". We can form constructions like "Greta has a belief about whether Hank has quit", but this type of construction fails to capture the content of Greta's state of mind, beyond indicating its topic.[5] By

[5] In particular, note that if Greta believes that Hank has quit while Ingrid believes that Hank has not quit, then they both have beliefs as to whether he has quit, although the contents of these beliefs are diametrically opposed. If they both know whether he has quit, then they must agree.

contrast, indirect knowledge attribution really does capture the relevant content, even if it doesn't specify it overtly: If Greta knows whether Hank has quit, then there is some truth out there in the world about Hank, and whatever it is, this truth is what Greta has in mind, a truth she could express directly if asked. If Greta just has a belief about whether Hank has quit, then we can expect her to give some answer if questioned about the matter, but not necessarily the right answer. However, outside of formulaic greetings and special contexts like pretence, the point of inquiry or questioning is generally not just to come up with some answer or another, or to see what the respondent will say, but to get the right answer. This point does something to explain why constructions involving "knows whether" are more than 100 times more common in corpus data than constructions involving "belief about whether" (Davies, 2008): the simplicity and frequency of the former type of construction are evidence of its greater practical utility.

This difference is a cross-linguistically robust phenomenon: Verbs in the family of "know" embed questions, where verbs in the family of "believe" do not (Karttunen, 1977; Lahiri, 2002). It has been argued that this is because an interrogative complement denotes the true answer to the question that it embeds, so it would be out of place under a cognitive verb like "believe" that can take either true or false complements (Egré, 2008). It may also reflect the fact that the knowing and believing relations themselves have fundamentally different kinds of social and practical significance, which has shaped the evolution and development of our basic mindreading abilities, both linguistic and otherwise.[6] Along these lines, there is even evidence that the capacity to track knowledge-*wh* does not require language at all: At least some nonhuman primates are able to attribute knowledge-*wh*, selectively following the behavior of a

[6] For a recent proposal about the influence of core knowledge structures on the emergence of linguistic universals, see Strickland (2017).

better-positioned agent who knows where a reward is located (Krachun et al., 2009).[7] This is consistent with the idea that tracking knowledge-*wh* is a very basic – and perhaps very ancient – form of mindreading: Agents with this capacity can selectively rely on each other as guides to parts of the world they have not yet explored for themselves.

For some proponents of the factive mindreading framework, how exactly we represent egocentric ignorance has posed something of a puzzle. Phillips and Norby (2019) argue that doing so must be more cognitively demanding than the complementary task of attributing ignorance to another on a question on which one is knowledgeable (they call this *altercentric ignorance*). To represent egocentric ignorance, these authors suggest, one must in effect represent the set of possible knowledge states that the knowledgeable agent *might* have (a "map of maps," in their terminology). If correct, then representing egocentric ignorance would not just be more complex than representing altercentric ignorance, but also more complex than representing *beliefs*: instead of generating a single decoupled representation – as is the case in belief attribution – this way of representing egocentric ignorance would force us to generate multiple decoupled representations at once, each corresponding to a possible knowledge state.

However, this approach is implausibly demanding in conversational contexts. Understanding that others possess information that one lacks is a likely precondition for being able to learn through explicit verbal questioning – that is, requesting information from knowledgeable others. Phillips and Norby's account seems to imply that this very basic form of social learning requires a great deal of cognitive sophistication. But this does not comport with the data on the development of interrogative behavior: by age two, well-formed verbal questions

[7] There is also some evidence that nonhuman primates may be able to track beliefs (Krupenye et al., 2016), but there are doubts about whether this indicates even an implicit representation of false belief, as opposed to some weaker registration of what the agent has recently seen (Tomasello, 2018).

constitute a large proportion of children's speech when interacting with caregivers (Chouinard, 2007), and by roughly 30 months of age, although children can also ask for objects, help, and permission, they are predominantly asking for information (P. L. Harris et al., 2017). During the second year of life, infants also engage in a wide range of selective social learning behaviors, which also suggests a degree of proficiency when monitoring who knows more than they do (Poulin-Dubois & Brosseau-Liard, 2016). It would be quite odd if children at this age – whose basic decoupling abilities are the source of some controversy (Perner, 1991) – are nevertheless able to sustain the kinds of representational activities that Phillips and Norby describe. Rather, it seems that from an early age, children are capable of representing not only the ignorance of others, but also their own ignorance relative to the knowledge of another.

The problem with Phillips and Norby's account, we suggest, is that it implicitly assumes that when we attribute knowledge-*wh* to someone else, we must represent the set of possible *answers* to the embedded question as well. But this is not the right way to construe the representational demands of questions. While there is a sense in which a question semantically denotes the set of propositions constituting its possible answers, where the form of a question constrains what falls into this set (Hamblin, 1976; Roberts, 2012), an agent asking or thinking of a question themselves need not represent these answers as such. Rather, there is good reason to think that questions can function as basic, non-meta-representational representations of the world (Carruthers, 2018). The cognitive content of these "questioning attitudes" is the question itself (e.g. "What's over there?" or "Who's that?"), not its possible answers. Of course, it is possible to also represent the possible answers to a questioning attitude over the course of deliberation. But as Carruthers argues, the basic function of these attitudes is not deliberation as such, but rather the guidance of information-seeking behaviors.

This representational characteristic of questioning attitudes also applies when we represent basic factive relations between agents and questions. Imagine, for instance, that you are seated across the room from another agent, separated by an opaque barrier. A third agent enters and appears to place something behind the barrier, out of your line of sight but within that of the second agent. Curious about the object, you might token a primary representation of the following form: "What is behind the barrier?" This attitude reflects the fact that there is a gap in your knowledge, a patch of the world that has currently become inaccessible to you. Monitoring the second agent's line of sight, you might also think, "S knows *what is behind the barrier*." But attributing knowledge of this question's answer to the second agent does not require you to represent all of its possible answers any more than thinking the question in the first place. Rather, all you need to do is keep track of the fact that "*What is behind the barrier*" is known to the other agent, even if it is not known to you.

As this case illustrates, inferences about basic factive relations between agents and questions could be generated on the basis of the same heuristics described in the previous section. This is most obvious in cases of perceptual knowledge of one's immediate environment, in which line-of-sight cues can tell us whether another agent has visual access to a patch of the world that is currently inaccessible to us. Memory-based processes could likewise account for the attribution of knowledge about absent entities, past events, or abstract topics. Suppose one strongly associates a particular conversational partner with a specific topic – say, the local basketball team. When you encounter this person, the fluency with which that topic comes to mind will serve as a strong cue that they have knowledge of it. When one has a particular question in mind about that topic (e.g. "Who is leading in the Eastern Conference?"), the

presence of this epistemic cue can lead us to quickly judge that this person is likely to have the answer.

In sum, once we have taken into account the fact that questioning attitudes can be a part of our primary representations of the world, tracking when others know what we do not becomes a simple matter of inferring basic factive relations between agents and those questions. Like other forms of basic factive mindreading, this need not involve any kind of meta-representation, nor the storage of special-purpose reference diaries documenting all the things we take a person to know.

6. Epistemic tracking in conversation

6.1. Epistemic territory

The strategies for monitoring each other's factive mental states described in our model are consistent with a notion proposed by researchers in the field of conversation analysis: our conversational turns are often guided by routine assignments of "epistemic territory", or topic areas in which we attribute knowledge by default, following common patterns. For example, in virtually all settings, people are assumed to know more than others about their own feelings, thoughts and recent experiences; individuals are also generally credited with greater knowledge on such close-to-home topics as their own "relatives, friends, pets, jobs, and hobbies" (Heritage, 2012a). These default assignments enable conversation to move forward even between previously unacquainted individuals: for a substantial range of topics, one will have clear expectations about whether the other party will be more or less knowledgeable than oneself.

Initial default mappings of epistemic territory are important, according to Heritage, because as soon as conversation gets past greetings into clausal content, these mappings, which

run on a scale between K+ (clearly possessing knowledge) and K- (clearly lacking knowledge), are needed to solve not only the speaker's problem of whether to ask or tell, but also the addressee's problem of whether they are being told or asked. One might imagine that an addressee solves this last problem just by noticing whether the speaker has used declarative or interrogative syntax; however, the social actions of asking and telling are mutually recognized by the parties involved not simply by grammatical distinctions on their own, but in a way that is subject to prior assignments of epistemic status. Drawing on an extensive body of empirical work (Labov & Fanshel, 1977; Pomerantz, 1980; Stivers, 2010), Heritage argues that even a sentence with declarative syntax can heard as a question if the content is obviously within the addressee's epistemic territory but not the speaker's. For example, when a physician is taking a patient's medical history, her utterance of the declarative sentence, "You're married" will register as a question, even when spoken with falling intonation, because a person's own marital status lies deeply within their own epistemic territory, and typically further from the epistemic territory of a medical professional. By choosing that declarative formulation, the physician takes a mild K- stance, suggesting some familiarity with the patient, but other options are open. Switching to a tag question ("You're married, aren't you?") would constitute more of a K- stance, and full interrogative syntax ("Are you married?") would be a stronger K- stance, appropriate in speaking to a new patient, for example (Heritage, 2012a). Meanwhile, interrogative sentences can function as rhetorical rather than genuine questions, eliciting conversational response patterns virtually identical to the response patterns characteristic of statements (Rohde, 2006). But a rhetorical question is only felicitous if the answer is already obvious to both speaker and addressee, so the possibility of this kind of conversational move again depends on prior calculation of what the parties will (take themselves to) know.

Instinctive mapping of epistemic territory also explains some recent experimental results advanced in support of the hypothesis that mentalizing is pervasive in ordinary conversation (Rubio-Fernández, Mollica et al., 2019). This work investigated the processing of statements which either did or did not violate expectations about what imagined strangers or friends would know. The same comment (about a hidden health condition such as an ulcer, for example) could be unremarkable if an overheard stranger is commenting on his own hidden health condition, or odd if a stranger is commenting on your hidden health condition; we do not expect strangers to have access to this part of our epistemic territory. By contrast, if a friend or family member comments on your ulcer, this is routine again. Rubio-Fernández and colleagues presented scenarios of each type to their participants, framed by contextualizing sentences which invited the participant to imagine themselves as addressees in a social setting with a friend or stranger ("you are having dinner with your dad at a restaurant when he says..."). They found slower reading times for the sentences in which strangers (as opposed to friends or family) were somehow aware of (presumably imagined) facts in the private life of the addressee. They concluded that dialogue is understood in light of "a default preference to monitor our interlocutor's knowledge states." (2019, 6).

Rubio- Fernández and colleagues do not distinguish between factive and nonfactive mentalizing—indeed, in their usage "belief reasoning" is equivalent with "estimating another person's knowledge" (2019, 1)—but we think the task here can be executed entirely with factive mindreading. The anomaly triggering the slowdown was not for example that the content of what was said was specifically processed as questionable or false, therefore triggering the need for belief reasoning: If it were, then readers (presumed not to have ulcers) should have been equally puzzled and slowed by the parallel remarks of the stranger and the friend or family member. The

problem was a violation in what the stranger could be expected to know, and in expecting the stranger to lack knowledge on a topic, it does not matter whether the topical remark is false or even coincidentally true (in the event that the stranger happens to be right). Strangers are not expected to know about our hidden health conditions, and even those we know quite well have limits to their epistemic territory: Rubio-Fernández and colleagues observe that you would find it alarming if your boss were to say something about a nightmare you had last night (2019, 3). We can make different default assignments of epistemic territory to different individuals, differentiated broadly by topic: even strangers are expected to know about the weather, certain friends and I know my health secrets, only I know about the dreams I have just had.

6.2. Cooperatively signaling epistemic status

Beyond our default assignments of epistemic territory, in live conversation we also face the more difficult task of refining and updating our models of what is known in light of what has been said: We ask each other questions, not always sure in advance of whether the other knows the answer, and we expect others to take on board at least some part of what we are sharing with them. Often one's initial knowledge attributions prove to be incorrect, and must be revised. And even if one's default assignments are correct, one must still constantly revise them to register the fact that the epistemic status of one's partner has changed as a result of successful knowledge transmission. Thus, managing the epistemic goals of conversation requires the ability to dynamically update our representations of what our interlocutor knows as the conversation unfolds.

This kind of epistemic tracking is aided by the production of overt signals of epistemic gain. In English, speakers produce the epistemic change-of-state marker "oh" (and, less frequently, "ah") to signal knowledge gain, or the revival of stored knowledge to current

awareness (Aijmer, 2002; Heritage, 1984). Both “oh” and “ah” can be produced to indicate surprise, the most marked form of knowledge gain, but they are also produced in response to unremarkable types of knowledge acquisition. “Oh” commonly functions as a receipt for information received in a response to a question, often coupled with an assessment (“Oh, thanks”; “oh great”; “oh no”). This use of “oh” marks the closing of an information-seeking sequence: the questioner wanted knowledge, the respondent produced it, and the questioner then signals that her epistemic needs are satisfied, and the exchange is complete (Heritage, 1984, 2012b). The production of epistemic gain signals is a cross-linguistically robust phenomenon, with some variation in how it is implemented: the Mandarin “ou” corresponds closely to the responsive English “oh” (Wu, 2017), while other languages, such as Finnish, divide the work of “oh” between several particles (Koivisto, 2016),

“Oh” is heavily used: as a marker taking the initial position in conversational turns, it is second only to “yeah”, which more often acknowledges information already accepted, sometimes signalling agreement and sometimes signalling something more like mere registration of the prior conversational turn (Norrick, 2009). We can clarify the way in which “oh” functions to signal knowledge gain by looking at contexts in which it is inappropriate: In pedagogical questioning, for example, where the teacher already knows the answer to the question she is asking, an expected answer should not elicit an “oh”, but something more like “that’s right, very good” (Heritage, 2018, p. 31). It is also possible to use “oh” deceptively, to pretend that one is just then learning what one already knew, but the availability of this trick is parasitic on awareness of the usual meaning of “oh” to signal knowledge gain. “Oh” can also be produced when one is in receipt of information that only seems to be knowledge (perhaps someone has

answered your question incorrectly), but it is like any other word in this regard ("gold" can also be produced in response to something mistaken for gold).

Importantly, while signals like "oh" are well-suited to signal changes in epistemic status, it is less clear how they might function to signal changes in *belief*. This is because vocalizations like "oh" are unstructured, which means that they have a relatively low informational bandwidth. For the purposes of efficiently communicating information that can be processed by one's partner while they are speaking, this low bandwidth is a feature rather than a bug. And unstructured signals work relatively well when one simply aims to indicate a positive or negative change in one's epistemic status, which can be reduced to a simple, one-dimensional parameter – namely, an increase or decrease in knowledge. Meanwhile, it is harder to imagine how one might convey changes of belief using unstructured signals. In order to convey useful information about a change in belief (e.g. shifting from "I believe it will rain" to "I believe it will snow"), one must be able to express something about their *contents*. But since the contents of a belief can vary in indefinitely many ways, that information is not so easily conveyed in a low bandwidth signaling system. Thus, information about changes in epistemic status is *compressible* in way that information about changes in belief is not, making it possible for us to efficiently signal changes in epistemic status in interactive contexts.

Factive mindreading is supported not only by direct signals like "oh", but also by a variety of other verbal and non-verbal indications of understanding. In recent work on simple "get acquainted" conversations, between participants with little initial common ground, Janet Bavelas and colleagues have studied a wide variety of devices used to signal knowledge transfer. They argue that new information is accepted into the common ground in face-to-face dialogue through a three-step procedure of speaker introduction, addressee response and speaker follow-

up. Here, the addressee response may consist in a verbal signal such as “oh”, or other expressive gestures (eyes widening, nodding); the speaker then demonstrates acceptance of that response with an utterance or gesture that does not introduce new topical content (“yeah”, nodding, smiling). In a study of over 1200 utterances introducing new information, they found that this three-step calibration process was followed 97% of the time (Bavelas et al., 2017). These overt signalling processes provide feedback to participants in cooperative conversations on how they are progressing towards their goal of shared knowledge.

Without the production of overt signals for knowledge gain, each side of a conversation could still begin with some intuitive mapping of the other side’s epistemic position, and update that mapping privately on the hope or expectation that the other side heard and accepted what was said. The production of a knowledge receipt signal (“oh, great”) makes it easier for both sides to engage in factive mindreading as conversation proceeds: the signal is confirmation that the recipient not only heard but accepts what was said, and indeed gained knowledge from it. The burden of representing the epistemic positions of the parties in the conversation is shared between the conversational participants, with each instinctively volunteering indications of progress towards salient shared knowledge, at least in cooperative conversations.

This leads us to a final observation about the conversational role of factive mindreading: it is *interactive*. A common critique of standard, belief-based approaches to mindreading in recent years has been that it is construed as a solitary, observational activity, and that much of the time, our socio-cognitive abilities are deployed in dynamic, interactive social contexts (De Jaegher et al., 2010). While we have some reservations about this “interactionist” approach to social

cognition,[8] we think that it captures an important feature of the kind of mental state attribution we are describing. Listeners, in this model, are not passive interpreters of speakers' utterances. By constantly signaling changes in their epistemic states, listeners actively assist speakers in the task of inferring their respective epistemic territories. Their feedback helps to shape the information that the speaker shares, so that it is appropriately calibrated to the listener's epistemic status. The listener thus actively participates in making their mind more readable, thereby facilitating the achievement of shared knowledge. This interactive mindreading strategy might also be achievable through a belief-tracking approach, but this would likely require a great deal of cognitive effort. Factive mindreading enables us to achieve the joint goals of cooperative conversation in a fluid, seamless, efficient manner.

7. When decoupling is needed

Having laid out our basic account of factive mindreading in conversation, we now turn to its relationship with nonfactive mindreading, and the roles that the latter plays in conversation. In cognitive terms, the distinction we have been drawing between factive and nonfactive mindreading can be thought of in terms of iterative reprocessing theories of dual-process cognition (Cunningham et al., 2007; Van Overwalle & Vandekerckhove, 2013). In these models, intuitive processes provide a quick default way of representing a stimulus or problem, which can either be accepted or modified through the use of slower, more representationally flexible, working memory-based computations. The extent to which the latter kinds of computation are involved in the reprocessing and reconstrual of a stimulus depends upon both the motivations of the agent and the availability of additional cognitive resources. If an intervention occurs, the

[8] Indeed, one of us has argued that concept of interaction often gets misused in these debates (Schönherr & Westra, 2018)

stimulus is reprocessed, this time while retrieving additional information and employing more complex representational formats. Depending on the agent's motivations, reprocessing can be repeated iteratively, generating increasingly nuanced representations of the stimulus with each cycle.

We suggest this is also the way to think about the relationship between factive and nonfactive forms of mentalizing, and in particular the shift towards a heavier reliance on decoupling. In the context of a straightforward cooperative conversation where the shared goal is to extend the speakers' mutual knowledge, coupled, factive mental state representations serve as a default means of mapping out and updating their respective epistemic territories. Nonfactive mindreading resources are recruited as needed when higher level monitoring systems detect that the default construal requires modification.[9] In these cases, more elaborate, fully decoupled models of the target's mind are constructed. This enables the mindreader to represent further properties of the agent's mental states, including their particular mode of presentation and their divergence from reality. Using these representations for the purpose of prediction and interpretation will require agents to inhibit their own perspective, which places a burden on their executive resources.[10]

Below, we outline a few of the contexts where the shift towards is likely to occur: deliberate deception, meta-linguistic repair, and argumentation. This list is intended to be illustrative rather than exhaustive: there are doubtless many other conversational contexts that require nonfactive, decoupled mindreading – for example, pretense, the comprehension of irony, and other cases of

[9] Note that on this model, not all forms of nonfactive mindreading are alike: Depending upon the number of reprocessing cycles involved in a bout of mindreading, our construal of a person's beliefs might be fairly sparse and schematic, or it might be quite rich and detailed.

[10] Although we leave open the possibility that with experience, agents might develop some expertise with this kind of reasoning in particular contexts, allowing them to become faster and more efficient at it over time (Elekes et al., 2016; Westra, 2017b).

semantic or pragmatic ambiguity. Indeed, a variety of communicative goals may regularly require us to depart from our factive mindreading default and engage in bouts of decoupling. There are also some domains, such as judgments of taste, whose status as factual is controversial, understood by some as expressing truths only relative to an assessor (Lasersohn, 2005), and by others as reflecting a combination of facts about the current state of discourse and facts about the world (Barker, 2013).[11] For the sake of brevity, we focus only on the three simpler cases mentioned above.

7.1.Deliberate Deception

In ordinary communicative contexts, conversational partners have the joint goal of achieving shared knowledge; our factive mindreading framework is in part an experience of how we pursue this end. But in contexts of deception or manipulation, conversation is no longer fully cooperative, and conversational partners are not pursuing a shared epistemic goal.[12] In these cases, factive mindreading is no longer sufficient (c.f. Pickering & Garrod, 2004, p. 180). Deceivers instead aim to unilaterally reduce the epistemic status of their interlocutors by causing them to form false beliefs. To do this effectively, deceivers need to maintain a representation of how the listener is thinking about the world that diverges from their own, and inhibit their own first-personal representations from interfering with their utterances. This is reflected in developmental research on lying, which shows that children's ability to lie effectively emerges in

[11] We are grateful to an anonymous referee for pressing us on the question of what should count as factual, including predicates of taste, modal statements and statements about the future. Insofar as the latter types of statement can already enter into one's primary representation of the world independently of interaction with another agent, they should count as factual for our purposes here. It is a good question whether judgments of taste call for decoupled mindreading; arguably, decoupling would be necessary only when standards are contested, but this is a question for future research.

[12] Not all acts of deception are entirely uncooperative, of course. In the course of otherwise cooperative conversations, we often engage in white lies and other limited acts of deception, which also require decoupling. Because lies of this sort are in fact quite common in our everyday lives (DePaulo et al., 1998), it stands to reason that decoupling is likely to be a fairly routine occurrence.

tandem with their belief-reasoning skills and is closely linked to levels of inhibitory control (Talwar & Lee, 2008; Williams et al., 2016); similar capacities are also implicated in adult deception (Christ et al., 2009). Deceiving others thus forces us to generate explicit representations of the minds of others, while also expending greater cognitive resources.

Something similar may be true in contexts where listeners are attuned to the possibility that a speaker intends to deceive, a disposition sometimes referred to as "epistemic vigilance" (Sperber et al., 2010). In such cases, we recognize that a speaker's goal is not to pursue shared knowledge, but instead to induce the listener to act on the basis of a misrepresentation of the world. Like the capacity to deceive, this ability emerges around age four (Mascaro & Sperber, 2009), around the time that children pass the false-belief task (Wimmer & Perner, 1983). Notably, children around this age are also slightly better at passing the false belief task when the target's false belief is the result of a deliberate deception (Chandler et al., 1989; Hala et al., 1991). Adults are also more likely to refer to an agent's beliefs when describing a silent video when it depicts a false-belief-inducing act of deception (Papafragou et al., 2007). Recognizing that an actor has deceptive intent thus creates a pragmatic context in which belief representation suddenly becomes relevant (Westra, 2017a), prompting a shift from factive to nonfactive forms of mindreading.

7.2. Meta-linguistic repair

Normally, the combination of a shared language, shared environment, and shared socio-cultural experiences will be enough for speakers to successfully communicate new information using only factive mindreading strategies. But occasionally, differences in the way speakers construe the world and differences in the way speakers use language can lead to communication difficulties. One clear example of this occurs after failures in "message formation" (Horton & Gerrig, 2005), when speakers use referring expressions that fail to determine the identity of the

referent. While selecting and interpreting referring expressions might normally operate according to simple associative principles (Horton & Gerrig, 2005; Shintel & Keysar, 2009), occasional breakdowns lead us to engage in decoupled forms of mindreading that allow us to better represent the way a speaker is thinking about the world. For example, when a speaker says, "I spoke to John this week," and the listener associates two different individuals with the name "John," they might initially misinterpret which person the speaker is referring to (Pickering & Garrod, 2004, p. 180). Such misalignments between speakers lead to confusion, perhaps accompanied by a verbal or nonverbal signal of lowered epistemic status (e.g. "huh?"). Resolving this kind of confusion requires interlocutors to explicitly consider the contrast between their different beliefs about the referent of "John," which requires decoupled mindreading.[13] This kind of miscommunication is also likely to prompt speakers to engage in explicit meta-linguistic negotiations (e.g. "Wait, by 'John,' do you mean John *Smith*?") until realignment is achieved.

A parallel form of collaborative meta-linguistic reasoning is conversation is evident in *reference communication tasks* (Brennan & Clark, 1996; Ibarra & Tanenhaus, 2016; Schober & Clark, 1989). In these studies, one participant plays the role of the director, while the other plays the role of a matcher. Separated by an opaque screen, the director must help the matcher select one particular image from a set of abstract, geometric shapes (or tangrams). To do this, the director must find a way of referring to the figure in question that the matcher will understand. Usually, the director offers multiple different possible descriptions of the referent, and the matcher counters with their own descriptions, until they have successfully aligned upon the same referent. In this process, the two partners might cycle through a number of different ways of

[13] Repair following communication breakdowns need not always require decoupled mindreading. Often the source of a communication breakdown is a misjudgment about what a hearer knows. This can be resolved by adjusting one's model of the speaker and listener's respective epistemic territories.

construing the same entity, describing it as “a candle”, “an anchor,” or “a dancer”, until their conversational goal is achieved (Horton & Brennan, 2016). In such tasks, unusual, contrived features of the experimental design – like having to refer to completely novel objects without a shared visual environment – prevent speakers from relying upon ordinary coupled mindreading strategies, forcing them to explicitly consider different possible modes of presentation of the same referent. Under these conditions, speakers must rely upon decoupled mindreading strategies in order to achieve conversational alignment.

The link between decoupled mindreading and repairing linguistic misalignment is supported by another body of evidence: The development of children’s meta-linguistic understanding. Specifically, between three and five years of age, children’s performance on tasks measuring their grasp of synonymy and homonymy is very highly correlated with their performance on traditional false-belief tasks, even after controlling for verbal mental age (Doherty, 2000; Perner et al., 2002). On our approach, this makes a lot of sense: Understanding synonymy and homonymy and learning to pass false-belief tasks both require us to actively consider different ways that an agent can represent the world. The similar representational demands of false-belief reasoning and meta-linguistic reasoning also seems to underlie bilingual children’s advantage on false-belief tasks (Diaz & Farrar, 2018; Schroeder, 2018), since these children are constantly confronted with people thinking about and referring to the same parts of the world in different ways.[14]

In short, failures of alignment at the level of referring expressions and resultant bouts of meta-linguistic reasoning can force speakers and listeners into a decoupled, executively

[14] This activity might also be what promotes the development of bilingual children’s cognitive flexibility and inhibitory control (Bialystok & Viswanathan, 2009), executive abilities that are also implicated in nonfactive mindreading (Farrant et al., 2012; Sabbagh et al., 2006).

demanding mode of mindreading. Resolving these failures requires them to generate decoupled representations of how their conversational partner is thinking about the world, and why they are describing the world in a particular way.[15]

7.3.Argumentation

In contrast to simpler practices of unchallenged telling, argumentative conversations involve the public production of explicit reasons for accepting or rejecting a claim. It is possible to engage in an argumentative conversation with like-minded individuals, sharing various supporting reasons for an already-accepted claim, but the work of producing explicit arguments would most naturally be motivated in contexts where a speaker stands to gain by persuading a disagreeing party; indeed, this application has been seen as the evolutionary basis of the universal human capacity to reason. As evidence that argumentation has the natural function of resolving disagreement, arguments among disagreeing parties are higher in quality, in the sense of showing more rigor in the evaluation of reasons, and generally producing more accurate outcomes (Mercier & Sperber, 2017).

Factive mindreading plays a large part in the successful construction of persuasive arguments. For example, when challenged to justify their decisions or persuade an ignorant party, even preschoolers are sensitive to what their questioners will and will not know (Köymen et al., 2016; Mascaro et al., 2019). Factive mindreading on its own may be sufficient for some simple arguments, and for sharing reasons with others who are like-minded. However, when arguments emerge in their natural role of resolving disagreement, nonfactive mindreading will

[15] As noted by two anonymous reviewers, what explains the role of mindreading in these meta-linguistic cases is not nonfactivity as such, but rather the process of decoupling, which is what enables us to reason about modes of representation (i.e. the "aspectuality" of mental states (Butterfill & Apperly, 2013)). Decoupling also happens to be what explains how we attribute false beliefs and other nonfactive states, and so mindreading in nonfactive and metalinguistic cases reflects the same underlying process.

swiftly become vital to successive argumentation. Recognizing that another person disagrees with you on some question (say, whether *p*) is a matter of belief attribution, not simply a matter of seeing them as failing to know whether *p*. Someone might fail to know whether *p* by having no opinion at all on that question, by having the false belief that not-*p*, or the true belief that *p*, but held for the wrong reasons. To judge one's adversary as disagreeing, one needs to see them as having a view on *p* opposite to one's own, so in ordinary cases of sincere argument, they will appear to their argumentative adversary to have a false belief, and the goal of the argument will be to talk them out of it, keeping track of their adherence to that false belief as a measure of how well the argument is progressing towards that goal.[16]

In particular, the construction of effective counter-arguments requires correctly identifying the content of an adversary's belief, as opposed to simply registering this adversary as failing to know. Research on the development of more advanced argumentative skills in 3-8 year-olds shows a significant correlation between argumentative performance and theory of mind scores on a battery of first-and second-order false belief tasks; indeed, theory of mind is a more significant predictor of argumentative competence than age or verbal ability (Slaughter et al., 2013). The correlation between nonfactive mindreading and argument extends beyond typically developing children: In a recent study also including children with autism spectrum disorder and deaf children of hearing parents, scores on false belief tasks remain stronger predictors of argumentative persuasion than age, verbal ability or disability status (Peterson et al., 2018). Nonfactive mindreading ability is especially predictive of "hetero-oriented" persuasion strategies, which take account of the reasons others are giving for their refusal, while affective

[16] Insincere argument, in which one argues for a conclusion that one personally rejects, is a special case of deliberate deception, and so will additionally involve belief reasoning for the reasons already reviewed in section 7.1.

empathy and verbal ability were not found to correlate with this type of argumentative skill (Lonigro et al., 2017).

8. Conclusion

Having argued that factive mindreading is generally easier than non-factive mindreading, we should also acknowledge explicitly that there are some special cases in which it will be harder. On many questions, epistemic authority is easy to spot; even if I have no idea of your marital status, I can easily take you to know what it is. On other questions, epistemic authority can be hard-won, and its signs harder to detect than the signs of sincere belief. It could be clear to me that an enthusiastic young paleontologist believes that a fossil fragment they have just found is from some obscure species of dinosaur, even if I am in no position to discern whether they actually know this. Similarly, while instinctive attributions of ignorance concern a natural domain of salient truths—I instinctively see you as failing to know that there is a threat looming behind you, for example—reasoned attributions of ignorance can concern almost any proposition, true or false, and the detection of ignorance can involve not only simple sightline calculations, but complex evaluations of the quality of someone's reasoning.

The existence of these harder cases does not speak against the notion that cooperative conversations generally involve participants with more easily recognized forms of knowledge and ignorance, however. Our model predicts that conversation will be easier with cooperative partners whose epistemic state is evident: for example, conversational partners will be swifter to ask questions of or accept information from those who are conspicuously well-positioned to reply, especially when compared to conversational partners for whom the relevant facts appear to lie outside their likely epistemic territory (Rubio-Fernández et al., 2019). One way to explore these predictions might be to exploit the relationship between factive mentalizing and gaze

tracking in order to manipulate default attributions of epistemic territory – for example, by placing participants in tasks where they are reliant upon the advice of two speakers with different levels of visual access, and then measuring how readily they seek out or respond to information from an incongruent epistemic source. A slightly different way to manipulate default attributions of epistemic territory might be to take advantage of listeners' prior assumptions about the differences in the background knowledge between experts and non-experts. In general, we should expect differences in default attributions of epistemic territory to affect both information-seeking behavior and the speed of information uptake.

If conversational partners learn from each other chiefly through factive mentalizing, we should expect to see evidence of this in the development of communicative abilities. At the age of 18 months, infants already show selective social learning from conversational partners, distinguishing between informants who had and had not accurately named familiar objects. In a novel word-learning task, 18-month-olds were more likely to retain new words taught by an informant with a track record of reliability; performance on this task was correlated with success on a knowledge attribution task, but not with measures of false belief attribution or domain-general statistical learning skills (Crivello et al., 2018). Contrasting these results with evidence on the relations between performance on the false-belief task and abilities such as lying (Talwar & Lee, 2008) and epistemic vigilance (Mascaro & Sperber, 2009) yields a further prediction: Competence on factive and non-factive mentalizing tasks should exhibit different developmental relations with cooperative and non-cooperative communicative abilities.

In many conversational tasks, it can be unclear whether factive or nonfactive mentalizing is called into play, and researchers may want to re-examine tasks that have been assumed to demonstrate facility with decoupled content. For example, the Diverse Beliefs task is typically

interpreted as testing whether children are able to represent beliefs opposed to their own, in a situation in which the truth is an open question. However, factive mentalizing would afford another way to pass this test. The task is conversational, in that children are told about the mental state of the protagonist, rather than needing to infer it from nonverbal behavior. To begin, the child is shown a picture of an agent (Linda) and asked to guess which of two locations contains a target object (a cat). For example, the child is shown a picture of some bushes and a garage, and is asked where the cat is hiding. Whatever answer the child gives, he is told that Linda thinks the cat is in the other location. The child is then asked where Linda will look for the cat, and is counted as passing the task if he gives an answer contrary to his own initial guess. Most mentalistic interpretations of this task see it as demonstrating a capacity to represent a belief contrary to one's own. However, it is also possible that the child takes the experimenter to be knowledgeable, and hears the contrary suggestion as a tip about the location of the cat, prompting an update of the primary representation of the scenario (Westra & Carruthers, 2017). To test this hypothesis, instead of asking children about Linda's likely action, the experimenter could instead give them a chance to "search for the cat" themselves at this juncture (perhaps by lifting a flap). If they search in the place associated with Linda, then their success in the task does not demonstrate a capacity to hold contrary contents in mind at the same time: The report about Linda has changed the child's mind, and at the moment of test the child no longer has a primary representation contrary to the one associated with Linda. If this prediction holds up, then nonfactive mentalizing is not needed to pass this task. To establish that factive mentalizing is needed, one might check whether children respond differently to tips from experimenters who are reliable and those who are unreliable, building on the novel word-learning paradigm.

Beyond these specific predictions, our model also yields one very major prescription: Researchers interested in the role of mentalizing in communication should pay closer attention to the distinction between factive and nonfactive mentalizing. Although this distinction is often elided in practice, whether we frame a claim about mentalizing in terms of knowledge attribution or belief attribution has significant implications for the underlying cognitive processes involved. Thus, when researchers describe a sociocognitive phenomenon as "true belief attribution" when they really mean "knowledge attribution", they are inadvertently invoking a more complex and executively demanding set of cognitive mechanisms with a distinct set of conversational functions, thereby making their underlying claims less plausible. By the same token, researchers who refuse to characterize a communicative process in terms of mentalizing on the grounds that it need not involve decoupling risk oversimplifying the sophistication and representational richness of our efficient epistemic tracking abilities. Even without decoupling, speakers and listeners are still in a position to grasp the propositional contents of each other's mental states – a fact that has significant implications for how we understand each other as epistemic agents.[17]

[17] Thanks to Peter Carruthers, John Michael, and Timothy Williamson for their comments on earlier versions of this paper.

References

Aijmer, K. (2002). English discourse particles. In *English discourse particles: Evidence from a corpus*. John Benjamins.

Anand, P., & Hacquard, V. (2014). Factivity, belief and discourse. In L. Crnic & U. Sauerland (Eds.), *The art and craft of semantics: A festschrift for Irene Heim, Vol. 1* (pp. 69–90). MIT: CreateSpace Independent Publishing Platform.

Andrews, K. (2012). *Do apes read minds?: Toward a new folk psychology*. MIT Press.

Apperly, I. (2018). Mindreading and Psycholinguistic Approaches to Perspective Taking: Establishing Common Ground. In *Topics in Cognitive Science* (Vol. 10, Issue 1, pp. 133–139). https://doi.org/10.1111/tops.12308

Barker, C. (2013). Negotiating Taste. *Inquiry*, *56*(2–3), 240–257.

Bavelas, J. B., Gerwing, J., & Healing, S. (2017). Doing mutual understanding: calibrationg with micro-sequences in face-to-face dialogue. *Journal of Pragmatics*, *121*, 91–112.

Benson, J. E., & Sabbagh, M. A. (2009). Theory of Mind and Executive Functioning: A Developmental Neuropsychological Approach. In P. D. Zelazo, M. Chandler, & E. Crone (Eds.), *The Developing Infant Mind: Integrating Biology and Experience* (pp. 63–80). Guilford Press.

Bermudez, J. L. (2003). The Domain of Folk Psychology. *Royal Institute of Philosophy Supplement*, *53*, 25–48. http://journals.cambridge.org/action/displayFulltext?type=1&fid=6247456&jid=PHS&volumeId=53&issueId=-1&aid=6247448&bodyId=&membershipNumber=&societyETOCSession=

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, *112*(3), 494–500. https://doi.org/10.1016/j.cognition.2009.06.014

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., & Hanna, J. E. (2009). Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science*, *1*(2), 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(1), 207–220. https://doi.org/10.1017/S030500090700829X

Burge, T. (2018). Do infants and nonhuman animals attribute mental states? *Psychological Review*, *125*(3), 409–434. https://doi.org/10.1037/rev0000091

Butterfill, S., & Apperly, I. (2013). How to Construct a Minimal Theory of Mind. *Mind and Language*, *28*(5), 606–637.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive

function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*(2), 73–92. https://doi.org/10.1002/icd.298

Carruthers, P. (2016). Two Systems for Mindreading? *Review of Philosophy and Psychology*, *7*(1), 141–162. https://doi.org/10.1007/s13164-015-0259-y

Carruthers, P. (2018). Basic questions. *Mind and Language*, *33*(2), 130–147. https://doi.org/10.1111/mila.12167

Carruthers, P., & Smith, P. K. (1996). *Theories of theories of mind*. Cambridge University Press.

Chandler, M., Fritz, A. S., Hala, S., Chandler, M., Fritz, A. S., & Hala, S. (1989). Articles Small-Scale Deceit : Deception as a Marker of Theories of Mind. *Child Development*, *60*(6), 1263–1277.

Chouinard, M. M. (2007). Children's questions: A mechanism for cognitive development. In *Monographs of the Society for Research in Child Development* (Vol. 72, Issue 1, pp. 1–121). Wiley. https://doi.org/10.1111/j.1540-5834.2007.00413.x

Christ, S. E., Essen, D. C. Van, Watson, J. M., Brubaker, L. E., & Mcdermott, K. B. (2009). The Contributions of Prefrontal Cortex and Executive Control to Deception : Evidence from Activation Likelihood Estimate Meta- analyses. *Cerebral Cortex*, *19*, 1557–1566. https://doi.org/10.1093/cercor/bhn189

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge University Press.

Crivello, C., Phillips, S., & Poulin-Dubois, D. (2018). Selective social learning in infancy: looking for mechanisms. *Developmental Science*, *21*(3), 1–10. https://doi.org/10.1111/desc.12592

Cunningham, W. a., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, *25*(5), 736–760. https://doi.org/10.1521/soco.2007.25.5.736

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): One billion million words, 1990-2019.* https://www.english-corpora.org/coca/

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447.

Dennett, D. C. (1978). Beliefs about beliefs [P&W, SR&B]. *Behavioral and Brain Sciences*, *1*(04), 568. https://doi.org/10.1017/S0140525X00076664

DePaulo, B. M., Kashy, D. A., Dahne, H., Epstein, J., Hairfield, J., Kirkendol, S., Prescott, C., Reeves, S., & Smith, C. (1998). Everyday lies in close and casual relationships. In *Interpersonal relations and group processes* (Vol. 74, Issue 1, pp. 63–79). http://smg.media.mit.edu/library/DePaulo.Kashy.EverydayLies.pdf

Devine, R. T., & Hughes, C. (2014). Relations Between False Belief Understanding and Executive Function in Early Childhood: A Meta-Analysis. *Child Development*, *85*(5),

1777–1794. https://doi.org/10.1111/cdev.12237

Diaz, V., & Farrar, M. J. (2018). The missing explanation of the false-belief advantage in bilingual children: a longitudinal study. *Developmental Science*, *21*(4), 1–13. https://doi.org/10.1111/desc.12594

Doherty, M. J. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language*, *27*(2), 367–392. https://doi.org/10.1017/S0305000900004153

Egré, P. (2008). Question-embedding and factivity. *Grazer Philosophische Studien*, *77*, 85–125.

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition*, *41*, 93–103. https://doi.org/10.1016/j.concog.2016.02.010

Fabricius, W. V, Boyer, T. W., Weimer, A. A., & Carroll, K. (2010). True or false: do 5-year-olds understand belief? *Developmental Psychology*, *46*(6), 1402–1416. https://doi.org/10.1037/a0017648

Farrant, B. M., Maybery, M. T., & Fletcher, J. (2012). Language, cognitive flexibility, and explicit false belief understanding: Longitudinal analysis in typical development and specific language impairment. *Child Development*, *83*(1), 223–235. https://doi.org/10.1111/j.1467-8624.2011.01681.x

Fizke, E., Barthel, D., Peters, T., & Rakoczy, H. (2014). Executive function plays a role in coordinating different perspectives, particularly when one's own perspective is involved. *Cognition*, *130*(3), 315–334. https://doi.org/10.1016/j.cognition.2013.11.017

Flavell, J. H. (1977). The development of knowledge about visual perception. *Nebraska Symposium on Motivation*, *25*, 43–76.

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. https://doi.org/10.1016/j.jml.2009.09.002

Gallagher, S. (2001). The practice of mind. Theory, simulation or primary interaction? *Journal of Consciousness Studies*, *8*(5–7), 83–108.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123.

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/10.1038/jid.2014.371

Grice, H. P. (1975). Logic and Conversation. In D. Davidson & G. Harman (Eds.), *The Logic of Grammar* (pp. 64–75). Dickenson.

Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledgling Theories of Mind: Deception as a Marker of Three-Year-Olds' Understanding of False Belief. *Child Development*, *62*(1), 83–97. https://doi.org/10.1111/j.1467-8624.1991.tb01516.x

Hamblin, C. L. (1976). Questions in Montague English. In B. Partee (Ed.), *Montague grammar* (pp. 247–259). Academic Press.

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

Harman, G. (1978). Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences*, *1*(04), 576. https://doi.org/10.1017/S0140525X00076743

Harris, D. W. (2019). We talk to people, not contexts. *Philosophical Studies*. https://doi.org/10.1007/s11098-019-01335-8

Harris, P. L., Ronfard, S., & Bartz, D. (2017). Young children's developing conception of knowledge and ignorance: work in progress. *European Journal of Developmental Psychology*. https://doi.org/10.1080/17405629.2016.1190267

Heller, D., Gorman, K. S., & Tanenhaus, M. K. (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.x

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. *Structures of Social Action: Studies in Conversation Analysis*, 299–345.

Heritage, J. (2012a). Epistemics in Action: Action Formation and Territories of Knowledge. *Research on Language and Social Interaction*, *45*(1), 1–29. https://doi.org/10.1080/08351813.2012.646684

Heritage, J. (2012b). The Epistemic Engine: Sequence Organization and Territories of Knowledge. *Research on Language and Social Interaction*, *45*(1), 30–52. https://doi.org/10.1080/08351813.2012.646685

Heritage, J. (2018). The ubiquity of epistemics: A rebuttal to the 'epistemics of epistemics' group. *Discourse Studies*, *20*(1), 14–56. https://doi.org/10.1177/1461445617734342

Heyes, C. (2014). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *9*(2), 131–143. https://doi.org/10.1177/1745691613518076

Heyes, C. (2018). *Cognitive gadgets: the cultural evolution of thinking*. Harvard University Press.

Hintikka, J. (1975). Different Constructions in Terms of the Basic Epistemological Verbs: A Survey of Some Problems and Proposals. In *The Intensions of Intentionality and Other New Models for Modalities*. Springer.

Horschler, D. J., Santos, L. R., & MacLean, E. L. (2019). Do non-human primates really represent others' ignorance? A test of the awareness relations hypothesis. *Cognition*, *190*(April), 72–80. https://doi.org/10.1016/j.cognition.2019.04.012

Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, *22*(7),

1114–1139. https://doi.org/10.1080/01690960701402933

Horton, W. S., & Brennan, S. E. (2016). The role of metarepresentation in the production and resolution of referring expressions. *Frontiers in Psychology*, *7*(JUL), 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

Ibarra, A., & Tanenhaus, M. K. (2016). The flexibility of conceptual pacts: Referring expressions dynamically shift to accommodate new conceptualizations. *Frontiers in Psychology*, *7*(APR), 1–13. https://doi.org/10.3389/fpsyg.2016.00561

Ichikawa, J. J., & Steup, M. (2018). *The analysis of knowledge* (E. N. Zalta (ed.)). The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, *116*(1), 26–37.

Jachmann, T. K., Drenhaus, H., Staudte, M., & Crocker, M. W. (2019). Influence of speakers' gaze on situated language comprehension: Evidence from Event-Related Potentials. *Brain and Cognition*, *135*(June), 103571. https://doi.org/10.1016/j.bandc.2019.05.009

Jacques, S., & Zelazo, P. D. (2005). Language and the Development of Cognitive Flexibility: Implications for Theory of Mind. In *Why Language Matters for Theory of Mind* (pp. 144–162). Oxford University Press.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, *1*(1), 3–44. https://doi.org/10.1007/BF00351935

Kiparsky, P., & Kiparsky, C. (1970). Fact. In M. Bierwich & K. E. Heidolph (Eds.), *Progress in linguistics* (pp. 143–147). Mouton.

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, *40*(5), 419–435. https://doi.org/10.1006/jhev.2001.0468

Koivisto, A. (2016). Receipting information as newsworthy vs. responding to redirection: Finnish news particles aijaa and aha (a). *Journal of Pragmatics*, *104*, 163–179.

Köymen, B., Mammen, M., & Tomasello, M. (2016). Preschoolers use common ground in their justificatory reasoning with peers. *Developmental Psychology*, *52*(3), 423.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535. https://doi.org/10.1111/j.1467-7687.2008.00793.x

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110

Labov, W., & Fanshel, D. (1977). *Therapeutic discourse: Psychotherapy as conversation*. Academic Press.

Lahiri, U. (2002). *Questions and answers in embedded contexts* (Vol. 2). Oxford University Press on Demand.

Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy*, *28*(6), 643–686. https://doi.org/10.1007/s10988-005-0596-x

Leslie, A. M. (1987). Pretense and Representation: The Origins of "Theory of Mind." *Psychological Review*, *94*(4), 412–426. http://eugeniesantiago.com/resources/Pretense_and_representation-the_origins_of_the_theory_of_mind.pdf

Lonigro, A., Baiocco, R., Baumgartner, E., & Laghi, F. (2017). Theory of mind, affective empathy, and persuasive strategies in school-aged children. *Infant and Child Development*, *26*(6), e2022.

Mascaro, O., Aguirre, M., Brun, M., Couderc, A., & Mercier, H. (2019). Nonverbal rhetoric: 2-to 4-year-old children select relevant evidence when trying to influence others. *Developmental Psychology*.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, *112*(3), 367–380. https://doi.org/10.1016/j.cognition.2009.05.012

McDowell, J. (1994). Knowledge by hearsay. In *Knowing from words* (pp. 195–224). Springer.

Mercier, H., & Sperber, D. (2017). *The Enigma of Reason*. Harvard University Press.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics*, *68*(2), 327–337. https://doi.org/10.3758/BF03193680

Moore, R. (2017). Gricean communication and cognitive development. In *Philosophical Quarterly* (Vol. 67, Issue 267, pp. 303–326). https://doi.org/10.1093/pq/pqw049

Nagel, J. (2017). Factive and non-factive mental state attribution. *Mind & Language*, *32*, 525–544.

Norrick, N. R. (2009). Interjections as pragmatic markers. *Journal of Pragmatics*, *41*(5), 866–891.

Papafragou, A., Cassidy, K., & Gleitman, L. (2007). When we think about thinking : The acquisition of belief verbs. *Cognition*, *105*(1), 125–165. https://doi.org/10.1016/j.cognition.2006.09.008.When

Perner, J. (1991). *Understanding the representational mind. Learning, development, and conceptual change.* MIT Press.

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences*, *16*(10), 519–525. https://doi.org/10.1016/j.tics.2012.08.004

Perner, J., Stummer, S., Sprung, M., & Doherty, M. J. (2002). Theory of mind finds its Piagetian

perspective: Why alternative naming comes with understanding belief. *Cognitive Development*, *17*(3–4), 1451–1472. https://doi.org/10.1016/S0885-2014(02)00127-2

Peterson, C. C., Slaughter, V., & Wellman, H. M. (2018). Nimble negotiators: How theory of mind (ToM) interconnects with persuasion skills in children with and without ToM delay. *Developmental Psychology*, *54*(3), 494.

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L., & Knobe, J. (2020). Knowledge before belief. *Behavioral and Brain Sciences*, 1–37.

Phillips, J., & Norby, A. (2019). Factive Theory of Mind. *Mind & Language*, *May*, 1–57. http://people.fas.harvard.edu/~phillips01/papers/FactiveToM.pdf%0Apapers3://publication/uuid/6008137D-748A-4291-B250-9CEACCE27203

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190; discussion 190-226. https://doi.org/10.1017/S0140525X04000056

Plato. (1980). *Meno* (G. M. A. Grube (ed.)). Hackett Publishing.

Pomerantz, A. (1980). Telling my side:"Limited access' as a "fishing" device. *Sociological Inquiry*, *50*(3-4), 186–198.

Poulin-Dubois, D., & Brosseau-Liard, P. (2016). The Developmental Origins of Selective Social Learning. *Current Directions in Psychological Science*, *25*(1), 60–64. https://doi.org/10.1177/0963721412436806

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515. https://doi.org/10.1017/S0140525X00076512

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*(6), 1–69. https://doi.org/10.3765/sp.5.6

Rohde, H. (2006). Rhetorical questions as redundant interrogatives. *San Diego Linguistics Papers*, *2*, 134–168.

Rubio-Fernández, P., Mollica, F., Oraa Ali, M., & Gibson, E. (2019). How do you know that? Automatic belief inferences in passing conversation. *Cognition*, *193*, 104011. https://doi.org/10.1016/j.cognition.2019.104011

Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Development*, *77*(4), 1034–1049.

Schober, F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211–232.

Schönherr, J., & Westra, E. (2018). Beyond 'Interaction': How to Understand Social Effects on Social Cognition. *The British Journal for the Philosophy of Science*, *0*, 1–26. https://doi.org/10.1093/bjps/axx041

Schroeder, S. R. (2018). Do Bilinguals Have an Advantage in Theory of Mind? A Meta-Analysis. *Frontiers in Communication*, *3*(August), 1–8.

https://doi.org/10.3389/fcomm.2018.00036

Schuwerk, T., Schecklmann, M., Langguth, B., Döhnel, K., Sodian, B., & Sommer, M. (2014). Inhibiting the posterior medial prefrontal cortex by rTMS decreases the discrepancy between self and other in Theory of Mind reasoning. *Behavioural Brain Research*, *274*, 312–318. https://doi.org/10.1016/j.bbr.2014.08.031

Scott-Phillips, T. (2014). *Speaking Our Minds: Why Human Communication is Different, and how Language Evolved to Make it Special* (Vol. 3). Palgrave Macmillan. https://books.google.com/books?hl=en&lr=&id=B0yjBQAAQBAJ&pgis=1

Sekicki, M., & Staudte, M. (2018). Eye'll Help You Out! How the Gaze Cue Reduces the Cognitive Load Required for Reference Processing. *Cognitive Science*, *42*(8), 2418–2458. https://doi.org/10.1111/cogs.12682

Shepherd, S. V. (2010). Following gaze: Gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience*, *4*(MARCH 2010), 1–13. https://doi.org/10.3389/fnint.2010.00005

Shintel, H., & Keysar, B. (2009). Less Is More: A Minimalist Account of Joint Action in Communication. *Topics in Cognitive Science*, *1*(2), 260–273. https://doi.org/10.1111/j.1756-8765.2009.01018.x

Slaughter, V., Peterson, C. C., & Moore, C. (2013). I can talk you into it: Theory of mind and persuasion behavior in young children. *Developmental Psychology*, *49*(2), 227.

Sperber, D., Fabrice, C., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, *25*(4), 359–393. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0017.2010.01394.x

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind and Language*, *17*(1&2), 3–23. https://doi.org/10.1111/1468-0017.00186

Stivers, T. (2010). An overview of the question–response system in American English conversation. *Journal of Pragmatics*, *42*(10), 2772–2781.

Strickland, B. (2017). Language Reflects "Core" Cognition: A New Theory About the Origin of Cross-Linguistic Regularities. *Cognitive Science*, *41*(1), 70–101. https://doi.org/10.1111/cogs.12332

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development*, *79*(4), 866–881. https://doi.org/10.1111/j.1467-8624.2008.01164.x

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190.

Thorup, E., Nyström, P., Gredebäck, G., Bölte, S., & Falck-Ytter, T. (2016). Altered gaze following during live interaction in infants at risk for autism: An eye tracking study. *Molecular Autism*, *7*(1), 1–10. https://doi.org/10.1186/s13229-016-0069-9

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality

account. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(34), 8491–8498. https://doi.org/10.1073/pnas.1804761115

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Belknap Press.

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal of Human Evolution*, *52*(3), 314–320. https://doi.org/10.1016/j.jhevol.2006.10.001

Van Overwalle, F., & Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: dual processes driven by a shared neural network. *Frontiers in Human Neuroscience*, *7*(September), 560. https://doi.org/10.3389/fnhum.2013.00560

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard university press.

Westra, E. (2017a). Pragmatic Development and the False Belief Task. *Review of Philosophy and Psychology*, *8*(2), 235–257. https://doi.org/10.1007/s13164-016-0320-5

Westra, E. (2017b). Spontaneous mindreading: a problem for the two-systems account. *Synthese*, *194*(11), 4559–4581. https://doi.org/10.1007/s11229-016-1159-0

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

Williams, S., Moore, K., Crossman, A. M., & Talwar, V. (2016). The role of executive functions and theory of mind in children's prosocial lie-telling. *Journal of Experimental Child Psychology*, *141*, 256–266. https://doi.org/10.1016/j.jecp.2015.08.001

Williamson, T. (2000). *Knowledge and Its Limits*. Oxford University Press.

Wilson, D., & Sperber, D. (2003). Relevance Theory. In G. Ward & L. Horn (Eds.), *Handbook of Pragmatics* (pp. 607–32). Blackwell.

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

Wu, R.-J. R. (2017). Particles and epistemics: Convergences and divergences between English and Mandarin. In *Enabling human conduct: Studies of talk-in-interaction in honor of Emanuel A. Schegloff* (pp. 273–298). John Benjamins. https://doi.org/10.1075/pbns.273.14wu

Young, L., & Waytz, A. (2013). Mind attribution is for morality. *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*, 93–103. http://moralitylab.bc.edu/wp-content/uploads/2013/06/06_Baron-Cohen_CH06.pdf

Zawidzki, T. W. (2013). *Mindshaping: A New Framework for Understanding Human Social Cognition*. MIT Press.

Box 1

Key terms

Mindreading – the process of attributing mental states to some target agent; also known as mentalizing

Factive mental state – a state of mind of a type that binds agents only to truths (e.g. knowledge)

Nonfactive mental state – a state of mind of a type that binds agent to true or false propositions (e.g. belief)

Fact - a true proposition

Complement – the propositional content embedded under a mental state verb, either as a declarative (*that*-) clause or an interrogative (*wh*-) clause

Primary representation – an agent's model of reality, which may incorporate other agents and their mental states

Decoupled representation – an agent's representation of the content of another agent's state of mind, quarantined from the agent's primary representation of reality

Mode of presentation – the description under which something is grasped, for example when the planet Venus is seen as either the Morning Star or the Evening Star

Gettier case – a case in which an agent has justified true belief without knowledge, for example because they have a contingently veridical hallucination

Egocentric ignorance – a situation in which the observed agent is taken to know something that the observer does not know

Epistemic territory (of an agent) – a set of topics on which the agent is by default treated as knowledgeable