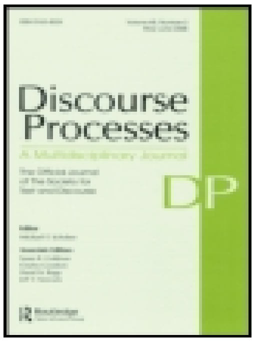

Discourse Processes

ISSN: 0163-853X (Print) 1532-6950 (Online) Journal homepage: https://www.tandfonline.com/loi/hdsp20

Individual Differences in Mentalizing Capacity Predict Indirect Request Comprehension

Sean Trott & Benjamin Bergen

DISCOURSE PROCESSES
https://doi.org/10.1080/0163853X.2018.1548219

Individual Differences in Mentalizing Capacity Predict Indirect Request Comprehension

Sean Trott and Benjamin Bergen

Department of Cognitive Science, University of California, San Diego, California, USA

ABSTRACT

People often speak ambiguously, as in the case of *indirect requests*. Certain indirect requests are conventional and thus straightforward to interpret, such as "Can you turn on the heater?", but others require substantial additional inference, such as "It's cold in here." How do comprehenders make inferences about a speaker's intentions? And what makes a comprehender more or less successful? Here, we explore the hypothesis that comprehenders do so in part by *mentalizing*—encoding what the speaker knows or believes about the world—and that differences in mentalizing ability predict how well comprehenders draw these inferences. In Experiment 1, we find that comprehenders' pragmatic interpretations are significantly influenced by a speaker's inferable knowledge states. In Experiment 2, we find that variability in this effect is explained by individual differences in comprehenders' mentalizing ability. Finally, in Experiment 3, we find that both effects are robust across different dependent measures of inference.

Introduction

People don't always say precisely what they mean. To cite a parade example, "It's cold in here" could be intended not as a reflection on the temperature, but as a request to turn on a heater. *Indirect requests* like this are a pervasive form of pragmatic ambiguity in everyday conversation; in one study, over 80% of requests elicited from participants were indirect in some way (Gibbs, 1981). Comprehenders often successfully infer a speaker's intent, even when it is not explicitly stated, and in some cases, are able to do so very early on (at 3–4 years) in development (Schulze, Grassmann, & Tomasello, 2013). Yet it's unknown what cognitive mechanisms underlie inferences about intent, whether individuals vary in their ability to make these inferences, and why this variability might occur.

Convergent evidence suggests that comprehenders make inferences about speaker intent at least in part by modeling what the speaker knows and believes (Gibbs, 1987). For instance, when trying to determine whether "It's cold in here" is a request to turn on the heater, the comprehender might be swayed by whether the speaker appears to think the heater is working or not. If the speaker knows the heater is broken, then "It's cold in here" probably means something else. In the series of experiments below, we ask whether comprehenders' inferences about the intent of potential indirect requests depend on this ability to encode and represent information about others' knowledge states and desires, known in the literature as Theory of Mind (Baron-Cohen, Leslie, & Frith, 1985; Premack & Woodruff, 1978; Wimmer & Perner, 1983), or *mentalizing* (Frith & Frith, 2006), the process of using Theory of Mind to

CONTACT Sean Trott sttrott@ucsd.edu Department of Cognitive Science, University of California, San Diego, 9500 Gilman Dr., 0515, La Jolla, CA 92093

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hdsp.

Supplemental data for this article, including the results of a Supplementary Experiment, can be accessed here.

reason about the mental states of others. Because mentalizing ability is known to vary across individuals, we also ask whether more adept mentalizers are more successful at inferring the intent of potential indirect requests. This is a specific instance of the more general issue of whether comprehenders resolve pragmatic ambiguity using a capacity to infer what the speaker knows, and whether individual differences in this capacity are responsible for differences in communicative success.

Indirect requests

Some indirect requests are easier to understand than others. Those that follow consistent formal patterns ("Could you …"; "Would you mind …"; etc.), and whose form makes reference to salient obstacles to fulfilling a request, e.g., saying "Do you have X?" to address the potential obstacle in which the listener does not possess or have access to "X" (Gibbs, 1986), are *conventional* indirect requests. These are best characterized as idiomatic constructions with their own idiosyncratic form and often partly compositional meaning (Stefanowitsch, 2003). This makes them relatively straightforward to interpret, especially when produced in the appropriate *situational context*; Gibbs (1986) found that potential indirect requests were read faster when their form referenced the greatest potential obstacle to fulfilling that request. This is consistent with the *obstacle hypothesis* (Francik & Clark, 1985; Gibbs, 1986), which states that speakers producing requests "formulate their utterance to deal with the greatest potential obstacle" (Gibbs, 1986, p. 182); correspondingly, these *conventional* requests are easier for listeners to interpret.

But *non-conventional* indirect requests, like "It's cold in here," are far more challenging for comprehenders to process for several reasons. First, unlike conventional indirect requests, a comprehender cannot use the form of a non-conventional indirect request as a cue to the intended speech act. For instance, there's no cue in the form of "It's cold in here" that indicates it might be anything other than an assertion. And second, the substance of a non-conventional indirect request has to be inferred using information outside of the form of the utterance itself. Whereas a conventional indirect request like "Could you turn on the heater?" explicitly specifies the substance of the request (turning on the heater), a non-conventional indirect request like "It's cold in here" does not. Instead, the comprehender may need to deploy world knowledge (what solutions exist to coldness?), the situational context (is there a heater or blanket at hand?), previous utterances in the discourse ("Is there anything I can do for you?"), and the social relationship between the interlocutors (host or guest, for instance), in order to ascertain the most likely speech act and its substance. Experimental evidence suggests that as a consequence, non-conventional indirect requests are more difficult to process than conventional indirect requests (as measured by slower reading times; Gibbs, 1981; Holtgraves, 1994), and also incur more processing difficulty (as measured by increased pupil dilation) than their literal, non-request counterparts (Tromp, Hagoort, and Meyer, 2016). How, then, do comprehenders understand them at all?

One proposal, the *mutual knowledge hypothesis*, argues that comprehenders understand pragmatically ambiguous utterances by engaging their *mentalizing* capacity—modeling the mental states of an interlocutor, and using this to adjudicate between possible interpretations of an utterance (Gibbs, 1987). This might include which information the speaker has access to, what the speaker believes about the world, and what the speaker wants or desires. Faced with pragmatic ambiguity—such as a potential non-conventional indirect request—a comprehender could use Theory of Mind to infer what the speaker knows, and arrive at an inference about their probable intentions. Below we survey what's known about whether and when comprehenders mentalize to process indirect requests and make pragmatic inferences more generally.

Mentalizing in request interpretation

There is some empirical evidence implicating mentalizing in indirect request comprehension. Recent brain-imaging work has found that overlapping brain regions are recruited for both representing the belief states of others and understanding indirect requests (Van Ackeren, Casasanto, Bekkering, Hagoort, & Rueschemeyer, 2012). Participants read the same sentence (e.g., "It's so hot here") in a request-supporting context (next to a picture of a stuffy room) or not (next to a picture of a desert) while functional magnetic resonance imaging (fMRI) was recorded. Brain regions associated with mentalizing, such as the medial prefrontal cortex (mPFC) and left temporoparietal junction (TPJ), received more blood flow during the request context than the literal-only context, possibly suggesting that the comprehension of indirect requests involves thinking about the mental states and intentions of others. Increased blood flow to regions associated with mentalizing (right TPJ, medial frontal cortex (MFC), and the anterior insula) was also found in an fMRI study investigating neural responses to indirect replies (Bašnáková, Weber, Petersson, van Berkum, & Hagoort, 2013), such as saying "It's hard to give a good presentation" when asked how a colleague's presentation went. This suggests that inferring a speaker's communicative intentions, particularly when these intentions are not explicitly encoded in the linguistic message, might involve mentalizing.

Yet these results leave questions open. For instance, in the Van Ackeren et al. (2012) study, participants were asked to determine on each trial whether the speaker wanted something from them, which could have encouraged mentalizing where it would otherwise be unnecessary. And perhaps more critically, spatial correlation in fMRI does not entail that mentalizing is causally involved in the comprehension of indirect requests (or indirect replies)––how is information about a speaker's mental states used to adjudicate between multiple interpretations of the same utterance?

Additional evidence comes from correlated deficits in Theory of Mind and general pragmatic reasoning abilities across neurodivergent populations. Individuals with schizophrenia have difficulty inferring the nonliteral interpretation of an utterance or proverb (Vygotsky & Kasanin, 1934); this difficulty has separately been found to correlate with impaired performance on Theory of Mind tasks (Brüne & Bodenstein, 2005; Champagne-Lavau & Stip, 2010). Additionally, patients with right-hemisphere brain damage consistently exhibit difficulty understanding nonliteral interpretations of utterances (Kaplan, Brownell, Jacobs, & Gardner, 1990; Winner & Gardner, 1977), as well as representing the belief states of others (Happé, Brownell, & Winner, 1999). Later work has found that these deficits in pragmatic inference and Theory of Mind correlate among individuals with right-hemisphere brain damage (Winner, Brownell, Happé, Blum, & Pincus, 1998), with some evidence also pointing to an important role for executive function (Champagne-Lavau & Joanette, 2009).

Early research on autism (Baron-Cohen et al., 1985) demonstrated that children with autism have difficulty representing the knowledge states of others, and soon after, Baron-Cohen (1988) argued that this deficit was functionally implicated in difficulties understanding pragmatically ambiguous language. Children with autism are less successful at recognizing violations of Gricean maxims (Surian, 1996), and more likely to interpret an utterance literally, rather than inferring the nonliteral "conveyed" meaning (Mitchell, Saltmarsh, & Russell, 1997). However, children with autism do not suffer from a universal pragmatic impairment. One recent study (Deliens, Papastamou, Ruytenbeek, Geelhand, & Kissine, 2018) demonstrated that although they had difficulty understanding ironic statements, they were able to understand both conventional and non-conventional indirect requests, suggesting that comprehending indirect requests may recruit distinct cognitive resources from understanding irony. Other work (Kissine et al., 2015) has also found that children with autism understand and comply with indirect requests, sometimes more reliably than typically developing children. These results paint a more complex picture of the relationship between Theory of Mind and pragmatic reasoning among individuals with autism, particularly when it comes to indirect request comprehension.

Computational models of pragmatic inference also often assume (implicitly or explicitly) that Theory of Mind is functionally involved in the inference process. Indirect requests and other forms of pragmatic ambiguity pose a challenge to Natural Language Understanding (NLU) systems because of their non-compositionality and context-dependence. One approach to solving this problem is using a set of rules about what utterances mean under different contexts (where "context" includes a probability distribution over the speaker's likely beliefs) to adjudicate between competing interpretations of the same utterance (Trott & Bergen, 2017; Williams, Briggs, Oosterveld, & Scheutz, 2015; Williams et al., 2014). For example, the sentence *I need a coffee* could be interpreted as a request for a coffee, or as a statement about the speaker's energy levels; the relative likelihood of each interpretation may depend on whether the comprehender believes that the speaker thinks the recipient of the request is subordinate to the speaker. Other computational models, such as the Rational Speech Act (RSA) framework, capture pragmatic inferences by assuming that a comprehender recursively reasons about a speaker's possible intended interpretations, and chooses the interpretation with the greatest utility, based on a model of both the situational context and the speaker (Frank & Goodman, 2012; Goodman & Frank, 2016). RSA has been expanded to allow for uncertainty in the speaker's knowledge states (Goodman & Stuhlmüller, 2013). This computational work thus serves as a proof of concept in that equipping a machine or algorithm with a model of the speaker's beliefs can improve its inferences about what the speaker meant by what they said.

While previous work has established correlations between mentalizing and pragmatic inference, and also demonstrated the potential utility of incorporating a speaker's probable beliefs into one's pragmatic interpretation, there is less evidence that neurotypical adults routinely follow this strategy. In fact, research on perspective-taking during language comprehension provides evidence that neurotypical adults don't always mentalize during the pragmatic inference process. For example, in a task in which comprehenders are able to assess whether an utterance was intended to be sarcastic or not with high accuracy, they nevertheless make more errors when predicting how a third-party—who does not have access to the same information—will interpret the utterance (Deliens, Antoniou, Clin, & Kissine, 2017; Keysar, 1994). Additionally, during a referential communication task, comprehenders' visual scan patterns include fixations on objects that are known to not be in the speaker's view (Keysar, Barr, Balin, & Brauner, 2000). Imposing time constraints further hinders perspective-taking during both speech production (Horton & Keysar, 1996; Roxβnagel, 2000) and comprehension (Deliens et al., 2017; Epley & Gilovich, 2006; Epley, Keysar, Van Boven, & Gilovich, 2004).

One explanation of these results, called the *egocentric anchoring and adjustment account*, is that comprehenders first interpret an utterance from their own perspective, then adjust perspectives to match an interlocutor's or third-party's as needed (Deliens et al., 2017; Epley et al., 2004). Others (Brown-Schmidt, 2009), however, have found evidence for more immediate perspective-taking when the task is made more interactive. Regardless, it is quite possible that the demands of real-time conversation place a bottleneck on perspective-taking ability, and that comprehenders rely on other information, such as prosodic cues (Deliens et al., 2017), to resolve pragmatic ambiguity. Shintel and Keysar (2009) propose that successful communication is usually achieved not through extensive modeling of an interlocutor's mental states, but rather through more domain-general coordination mechanisms, which in some cases are more akin to procedural routines than declarative knowledge.

A related model, known as *interactive alignment* (Pickering & Garrod, 2004), proposes that interaction facilitates the alignment of interlocutors' situation models, via the bottom-up entraining of phonetic, lexical, and syntactic representations. This largely removes the need for cognitively expensive mentalizing processes, as the chance for divergent situation models should decrease throughout an interaction. Of course, such divergences can still occur, but Pickering and Garrod (2004) argue that the resulting miscommunications are usually addressed through an interactive repair process. Notably, Pickering and Garrod (2004) raise the possibility that interlocutors *do* mentalize in extenuating circumstances, such as deception or deliberate misalignment, but that these cases are "clearly costly" and "may involve complex (and probably conscious) reasoning, and there may be great differences between people's abilities (e.g. between those with and without an adequate 'theory of mind')" (Pickering & Garrod, 2004, p. 180).

Where does that leave non-conventional indirect requests? Because they are non-compositional and context-dependent, they are strong candidates to be resolved through mentalizing. But to date, no studies we are aware of have directly tested whether comprehenders use inferred speaker knowledge to resolve the ambiguity of indirect requests. Moreover, given the evidence that comprehenders may not always mentalize (Deliens et al., 2017), and that indirect requests can be comprehended even by individuals from populations who typically suffer from impairments in mentalizing ability (Deliens et al., 2018; Kissine et al., 2015), it is possible but currently undemonstrated that neurotypical adults vary in the cognitive resources they recruit to comprehend indirect requests. Do individual differences in a comprehender's ability or propensity to mentalize in general predict differences in their reliance on a speaker's inferable mental states for pragmatic inference? Characterizing these differences—which people are most likely to mentalize, and when?—should yield a clearer picture of the mechanisms involved in the comprehension of non-conventional indirect requests, and in particular the role of mentalizing in pragmatic disambiguation.

Current work

In three studies, we presented pragmatically ambiguous utterances, such as "It's cold in here" spoken by characters in a narrative. We asked participants to make inferences about what the speaker intended. We manipulated whether the comprehender should infer that the speaker was or was not aware of some obstacle to fulfilling a possible request (like a heater being broken), using a False-Belief Task–inspired design. The speaker was described as either present or not present when the information about the obstacle was presented. In Experiment 1, this allowed us to determine whether comprehenders changed their interpretation as a function of inferable speaker belief. In Experiments 2 and 3, we asked whether individual differences in the capacity to mentalize predicted differences in a participant's reliance on a speaker's inferable beliefs for assessing the speaker's intentions.

Experiment 1

In Experiment 1, we asked whether participants changed their pragmatic interpretation of an ambiguous utterance as a function of a speaker's inferable knowledge state. Participants read eight short passages (see Appendix), each describing a situated interaction with another character, and each ending with a potential indirect request (e.g., "It's cold in here"). In each passage, there was also an obstacle to fulfilling the request implicit in the ambiguous utterance, such as a broken heater. The only manipulation was whether or not the speaker could be inferred to be aware of this obstacle; the participant was always aware of the obstacle.

To evaluate participants' pragmatic interpretations, we asked them to make a paraphrase judgment of the speaker's intentions. They selected one of two paraphrases; one was always an explicit *request*, and the other was always a *literal* non-request interpretation (such as a complaint).

If comprehenders use a speaker's knowledge states to interpret an ambiguous utterance—if they mentalize—participants should be more likely to select a *request* paraphrase when the speaker can be inferred to be unaware of an obstacle to fulfilling the request. (That is, when the speaker doesn't know the heater is broken, the participant should be more likely to indicate that "It's cold in here" is intended as a request.) Similarly, they should be more likely to select a *literal* paraphrase when the speaker can be inferred to be aware of the obstacle. (That is, when the speaker knows the heater is broken, then "It's cold in here" is less likely to be a request.) Evidence that they do not mentalize in this context would come in the form of no modulation of request versus literal interpretations in the face of speaker knowledge or lack of knowledge about the obstacle.

If individuals do recruit their mentalizing capacity to understand indirect requests, there are multiple subprocesses that must be performed. As discussed in Trott and Bergen (2017), one of these is to *sample* information about a speaker's mental states and beliefs. This could be achieved by encoding explicit cues about mental states (e.g., a speaker verbalizing their beliefs) or by using implicit cues to make mental state inferences (e.g., noting that a character was not present when certain information was revealed). Furthermore, this sampling process could occur either *online* (i.e., when the cues initially manifest), or retroactively as needed (e.g., to disambiguate the meaning of an utterance). The second subprocess is actually *deploying* this information for a downstream task, such as interpreting a potential indirect request to determine what kind of speech act it is, and what its substance is (e.g., what the speaker might be requesting).

In this experiment, we are interested in both subprocesses—whether individuals can both sample information about a character's mental states using implicit cues, and deploy this information to adjudicate between competing interpretations of an utterance.

Methods

Participants

Fourty-two participants were recruited through Amazon Mechanical Turk, all right-handed, native English speakers. We aimed to recruit 40 participants, but Amazon Mechanical Turk oversampled to 42. Each participant was paid $1.20, and the experiment took on average 8.8 minutes to complete. There were 24 males and 18 females. The average age of participants was 37 (SD = 12.3), with ages ranging from 23 to 71.

Materials

The stimuli consisted of eight pairs of short (5–9 sentence) narrative passages, each of which ended with an ambiguous sentence that could be interpreted either as an indirect request or as another type of direct speech act. Each pair of passages described the same scenario, in which the participant (addressed via the 2nd-person) interacts with a character in the story, who ultimately utters the potential request. Also, in each situation, the participant learns about an obstacle to fulfilling the request, or a reason as to why the utterance might not be a request; for example, a broken heater in a car would prevent the participant from fulfilling the request to turn on the heater that could be implicit in, "It's really cold in this car."

The manipulation across each pair of passages was whether or not the speaker of the potential indirect request knew about this obstacle. The passages did not explicitly inform the participant about this knowledge state. Instead, it had to be inferred. For instance, the participant might learn about the obstacle while the speaker was either present or not present (see Appendix for an example). The two conditions were called *Speaker Aware* and *Speaker Unaware*.

The experiment was implemented using jsPsych (de Leeuw, 2015). The code for the implementation can be found on GitHub: https://github.com/seantrott/mentalizing_experimental_materials.

Procedure

Participants were instructed that they would read a series of short passages describing possible interactions. Participants were randomly assigned to one of two lists, counterbalanced for which stimuli were in the Speaker Aware and Speaker Unaware condition. Each participant read eight passages in random order-–four in the Speaker Aware condition, and four in the Speaker Unaware condition.

After the participant finished reading a passage, indicating this by button press, a new page appeared, which asked them to choose the sentence that best paraphrased the speaker's *intention* in the final utterance. They were presented with two options. One of these options was phrased as a *conventional indirect request*, using the "Could I …" or "Could you …" construction. The other option was the compositional interpretation of the sentence, phrased in a literal form, such as "how are you planning on getting to that party?" In order to avoid participants misconstruing this literal option as a potential indirect request, it always also included information about the passage's potential obstacle to fulfilling the request, e.g., "Since your car is in the shop …" (see Appendix). The paraphrase judgment options were presented in random order.

Aside from recording participants' responses to the paraphrase judgments, we also collected their reported gender, age, as well as what they thought the experiment was about and whether or not they were native-English speakers. The experiment software automatically recorded how long they spent on each screen of the experiment. No other measures were collected.

Results

As shown in Figure 1 below, the Speaker Unaware (M = 0.76, SD = 0.43) trials elicited a higher proportion of *request* paraphrases than the Speaker Aware trials (M = 0.3, SD = 0.46).

To evaluate the difference between conditions, we ran a generalized linear mixed-effects model in R (R Core Team, 2017) using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). We began with *paraphrase* as the dependent variable, condition (Speaker Aware vs. Speaker Unaware) as a fixed effect, and subjects and items as random intercepts (along with by-subject random slopes for the fixed effect of condition). We then evaluated this full model against a reduced model that removed the fixed effect of condition but preserved the same random effects of subjects and items.

The full model including condition predicted the choice of paraphrase significantly better than the reduced model [$X^2(1) = 35.094$, $p = 3.14*10^{-9}$], demonstrating that the speaker's awareness of an obstacle to fulfilling the request was a strong predictor of a participant's interpretation of their intentions. In other words, participants were more likely to interpret an ambiguous utterance as a request when the speaker could be inferred to be *unaware* of an obstacle, and more likely to interpret the utterance as a literal statement when the speaker could be inferred to be *aware* of an obstacle.

One potential explanation for this effect is that participants were learning what the experiment was looking for over the course of their session. That is, by the final item, they may have guessed that they were supposed to track the speaker's knowledge, and thus chose a consistent paraphrase. If this is the case, the effect should be very weak or nonexistent on the first trial, then gradually become very strong. To test this, we added an interaction between order and speaker knowledge to the model, as well as random intercepts for subjects and items. We found that the addition of order to

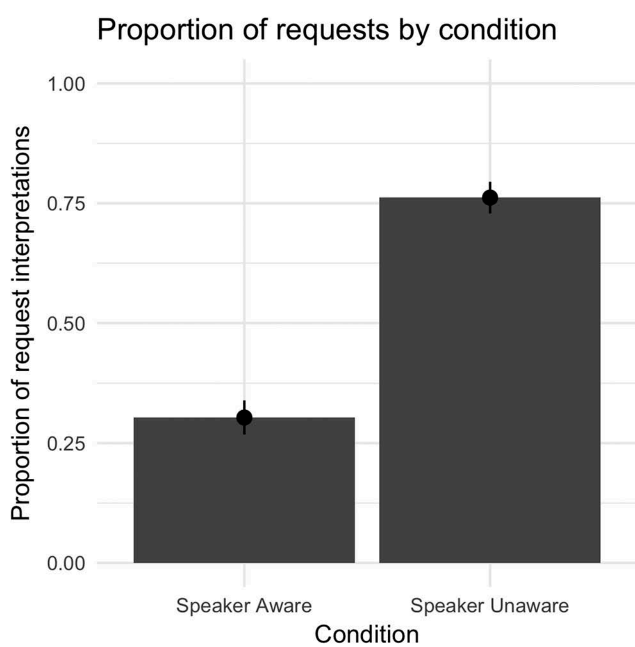

Figure 1. Participants were more likely to select a request paraphrase when the speaker could be inferred to be unaware of an obstacle (M = 0.76, SD = 0.43), than when the speaker was aware (M = 0.3, SD = 0.46). Error bars represent 1 standard error from the mean.

the model did not explain significantly more variance than the sole fixed effects of speaker knowledge and order [$X^2(1) = 1$, $p = 0.32$].

We also tested whether participants were aware of the experimental question by asking them what they thought the experiment was about after they were finished; 86% of participants either said they were "not sure," or repeated some variant of the initial instructions, such as "inferring intentions"; 14% of participants mentioned something about tracking knowledge states in a story, such as: "How people determine intention based on the speaker's knowledge in the text." To ensure that the fixed effect was not due to demand characteristics, we filtered out the 14% of participants who understood the experiment, then re-ran the primary analysis described above. The effect remained robust [$X^2(1) = 11.4$, $p = 0.0007$], demonstrating that even naive participants were affected by the manipulation of speaker knowledge.

Individual differences

Despite a main effect of speaker knowledge overall, there were still significant differences across participants in the degree to which they showed the effect. We operationalized effect size as the difference in the proportion of request paraphrases across conditions; an effect size of 1 would be a categorical effect, meaning that participants always interpreted utterances as requests in the Speaker Unaware condition, and never interpreted utterances as requests in the Speaker Aware condition. This operationalization correlated almost perfectly with the by-subject slope coefficients extracted from the *glmer* model ($r = 0.99$, $p < 2*10^{-16}$).

As shown in Figure 2 below, effect size varied considerably across participants; 81% of participants showed a positive effect, meaning they had more request paraphrases in the Speaker Unaware condition than the Speaker Aware condition. Only 16.7% of participants showed a categorical effect of speaker knowledge in the expected direction; 19% showed an effect of 0 or less, meaning that speaker knowledge either did not influence their judgment of speaker intent at all, or it actually influenced their judgment in the direction *opposite* from expected.

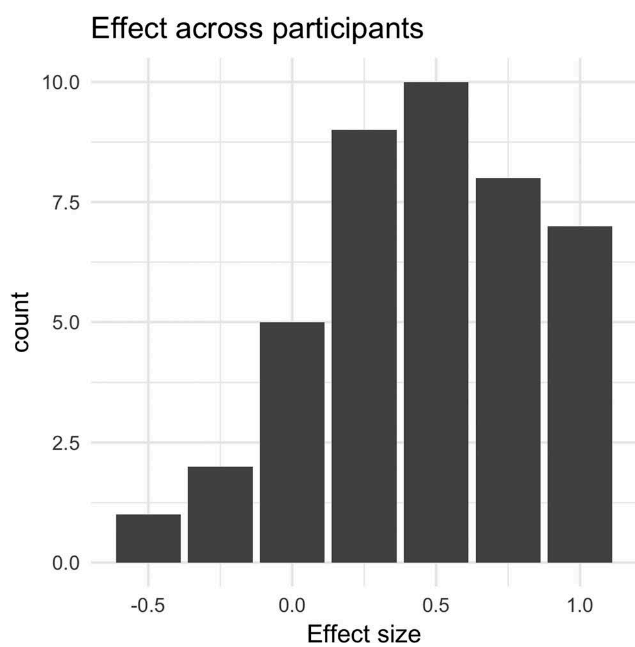

Figure 2. Individual differences in effect size across participants (N = 42). Here, "effect size" was operationalized as the difference in the proportion of request paraphrases across conditions.

Discussion

Our main question going into Experiment 1 was whether or not comprehenders encode what their interlocutors know, using this information to infer the speaker's intentions. More specifically, we asked whether a speaker's awareness of some obstacle to fulfilling a potential request would affect a comprehender's interpretation of that utterance. We found a significant effect of speaker knowledge, suggesting that: 1) overall, participants paid attention to what the characters in the stories knew or didn't know; and 2) when faced with the problem of disambiguating intent, their interpretation factored in not only their *own* knowledge, but also what they thought the speaker knew.

Speaker knowledge strongly predicted utterance interpretation, meaning that participants must have been both sampling and encoding that information while reading the passage. An important feature of the experiment's design was that the paraphrase questions did not appear on the same screen as the story. This means that people did not have the option to return to the passage and determine whether or not the speaker knew about the obstacle. Furthermore, there was no significant effect of order, suggesting that participants were not simply learning how to do the task over the course of a session; speaker knowledge predicted the paraphrase judgment from the very first trial. Even after filtering out participants who appeared to see through the manipulation, the main effect was robust.

In combination, these results suggest that people *naturally* pay attention to what characters in a story know. Participants were not told to attend to the speaker's knowledge states, and initially had no way of knowing that speaker knowledge would be an important variable, yet they reliably encoded this information nonetheless.

On the other hand, there were significant individual differences in the size of the effect of speaker knowledge (see Figure 2). Most (81%) participants showed *some* effect of speaker knowledge, but about 20% showed no effect or a negative effect. There are at least two possible explanations for these individual differences.

First, it is possible that some participants did not take the task seriously and simply answered at random. In this case, we would expect a bimodal distribution, in which one set of participants displayed a strong effect of the manipulation, and the other set were distributed normally around chance (e.g., a null effect of speaker knowledge). However, the distribution of effects does not show signs of bimodality, which makes it unlikely that the full range of individual differences can be explained by low attention.

Second and alternatively, effect size differences could be due to individual differences in *mentalizing*—the ability (and propensity) to model the beliefs and desires of others. The effect of speaker knowledge observed in Experiment 1 is predicated on participants attending to and encoding speaker knowledge. If certain people are less able, or less likely, to encode what others know, they might also be less likely to choose a paraphrase consistent with the speaker's knowledge state. Understanding individual variability casts light on underlying mechanisms; in this case, if the individual variability in Experiment 1 correlated with variability in mentalizing, this would serve as further evidence that mentalizing plays a role in this kind of pragmatic inference. Experiment 2 addresses this hypothesis.

Experiment 2

In Experiment 1, participants were more likely to interpret an ambiguous utterance as a request if the speaker was *unaware* of an obstacle to fulfilling that request—suggesting that speaker knowledge is an important variable for pragmatic interpretation. However, there was substantial variability in the extent to which comprehenders relied on a speaker's knowledge states when inferring intent. Since the task required attending to and encoding the implicit beliefs of the other characters in each passage, one potential explanation for this variability is that participants varied in their underlying *mentalizing* capacity—their ability and propensity to attend to the beliefs and desires of others.

Until recently, most research on mentalizing has focused on its developmental trajectory in young children (Astington & Jenkins, 1999; Wimmer & Perner, 1983), its presence (or absence) in the cognitive repertoire of non-human primates (Call & Tomasello, 2008; Krupenye, Kano, Hirata, Call, & Tomasello, 2016; Penn & Povinelli, 2007; Premack & Woodruff, 1978), and mentalizing deficits among people with schizophrenia (Bora, Yucel, & Pantelis, 2009), autism (Baron-Cohen et al., 1985), and right-hemisphere brain damage (Winner et al., 1998). Assessing individual differences in mentalizing ability among neurotypical adults is difficult, however; many traditional tasks, such as the *false-belief task*, are too easy and thus show ceiling effects (Dodell-Feder, Lincoln, Coulson, & Hooker, 2013). But as Turner and Felisberti (2017) point out, studying mentalizing differences among neurotypical adults—challenging though it might be—is important. Individual differences can lead to better theoretical models of cognitive phenomena. In our case, a metric for individual variability in mentalizing is a critical prerequisite for determining whether mentalizing differences explain the variability in performance observed in Experiment 1.

One solution is to make mentalizing tasks more difficult by asking participants to consider higher-order beliefs (e.g., "John thinks that Jim thinks that John knows …"). However, Dodell-Feder et al. (2013) argue that this introduces additional, "non-social" challenges related to working memory and executive function, and thus may not isolate mentalizing-specific differences. Dodell-Feder et al. (2013) also point out that another problem facing many traditional mentalizing tasks, like the false-belief task, is that they usually only require participants to pay attention to a single character's higher-order beliefs, instead of requiring that participants track the knowledge states (and affective states) of multiple characters throughout the course of a social interaction.

To address these problems, Dodell-Feder et al. (2013) developed the *Short Story Task* (SST), a novel methodology for assessing individual differences in mentalizing ability. In it, participants read a short story and then answered questions about the mental lives of characters in it. Their responses were scored according to the degree to which they made explicit reference to the characters' mental states. The results of the task were validated as a measure of mentalizing by correlating them with scores on tasks known to measure mentalizing—the Reading the Mind in the Eyes task, and the Interpersonal Reactivity Index (IRI), as well as IQ. Reading the Mind in the Eyes involves making judgments about a person's emotional states, based on the eye region of an actor's face, and is used as an assessment of affective empathy. The IRI is a questionnaire designed to capture differences along several dimensions, including *fantasy* (how well a person immerses themselves in a fictional character's mental life), *perspective-taking* (the tendency to adopt the mental states of others), *empathic concern* (the tendency to consider the emotional states of others), and *personal distress* (the tendency to experience negative affect in response to negative events affecting somebody else). Dodell-Feder et al. (2013) found that explicit mental state reasoning scores were significantly correlated with IQ (r = 0.24, p = 0.047), the *fantasy* subscale of the IRI (r = 0.37, p = 0.012), and the Reading the Mind in the Eyes task (r= 0.49, p < 0.0001).

The Short Story Task measures a reader's ability to attend to multiple characters' mental states, which change dynamically over the course of the story; it also measures a reader's ability to make inferences about multiple *kinds* of mental states (e.g., affective state, epistemic state, intentions). Unlike the IRI, the Short Story Task assesses mentalizing capacity through behavior––specifically, the extent to which a participant reasons about the mental states of the characters in a story. And unlike the Reading the Mind in the Eyes Task, which measures the ability to *decode* mental states (Dodell-Feder et al., 2013), the Short Story Task assesses a participant's ability to *reason* about a character's mental states on the basis of what they read in the text. This made the Short Story Task particularly well-suited for our purposes, because we were especially interested in how a participant's ability to reason about the mental states of others––to both *sample* and *deploy* information about their knowledge states––predicted their likelihood of making pragmatic inferences consistent with a speaker's knowledge states. Thus, we used the Short Story Task to investigate whether individual differences in mentalizing predicted variability in the task used in Experiment 1.

Experiment 2 had three goals. First, we wanted to test whether the findings in Experiment 1 would replicate in a new participant sample; specifically, that the speaker's implied knowledge would significantly predict which paraphrase the participant chose as the intended interpretation. Second, we were interested in whether the same pattern of individual variability would arise. And third, if it did, we wanted to determine whether it could be explained by mentalizing ability. If the variability in interpretation is due to differences in the ability to model the mental states of others, then an individual's mentalizing score on the *Short Story Task* should significantly explain their effect size on the paraphrase task from Experiment 1—stronger mentalizers should be more affected by implied speaker knowledge.

Methods

Participants

Eighty-three different participants were recruited through Amazon Mechanical Turk, all right-handed, native English speakers. Our target sample size was 80 participants, but as in Experiment 1, Amazon Mechanical Turk over-sampled to 83. Each participant was paid $2.50, and the experiment took on average 20 minutes and 43 seconds to complete. Before excluding participants on the basis of reaction times, there were 45 males, 37 females, and one participant who identified as non-binary. The average age of participants was 33 (SD = 9.5), ranging from 19 to 66.

Procedure

Participants first completed the same task as in Experiment 1, using the same set of stimuli. They read eight narrative passages, ending with an ambiguous utterance; after each passage, they selected the best paraphrase of the speaker's intentions. As in Experiment 1, the participants were randomly assigned to one of two lists, counterbalanced for which items were in the Speaker Aware condition and which were in the Speaker Unaware condition.

Then participants completed a version of the Short Story task (see Dodell-Feder et al., 2013, and the paragraph below) that we adapted for Web experiments. The main difference was that instead of verbally responding to questions posed by the experimenter, the participant read the questions on the browser page and typed their answers in a text box. The SST is described in more detail in Dodell-Feder et al. (2013), but a short summary of the task follows.

In the SST, the participant read a short story by Ernest Hemingway called *The End of Something*, in which a romantic couple argue while fishing and ultimately split up. Neither the interior lives nor the intentions of the characters are made explicit in the story; instead, to understand the story, the reader must infer what one character thinks of another, why one character said what they did, and how it might have made the other character feel. After finishing the story, the participant answered 14 questions designed to probe their mentalizing abilities, as well as their reading comprehension. Mentalizing ability was further broken down into *spontaneous mental state inference* (whether participants make inferences about a character's mental state without any prompting to do so) and *explicit mental state reasoning* (how well the participants track what different characters think and feel when asked explicitly about them). *Spontaneous mental state inference* was a binary variable indicating whether or not participants spontaneously made reference to a character's mental states when asked what happened in the story (e.g., whether they made mental state inferences, even when unprompted). *Explicit mental state reasoning* was a continuous variable between 0 and 16, based on a participant's answers to eight questions designed to target this ability. For example, one such question was: "What does Nick mean when he says, 'It isn't fun anymore'?" Full points (2) included answers such as "He's tired of the relationship" or "he wants to end the relationship"; 1 point was

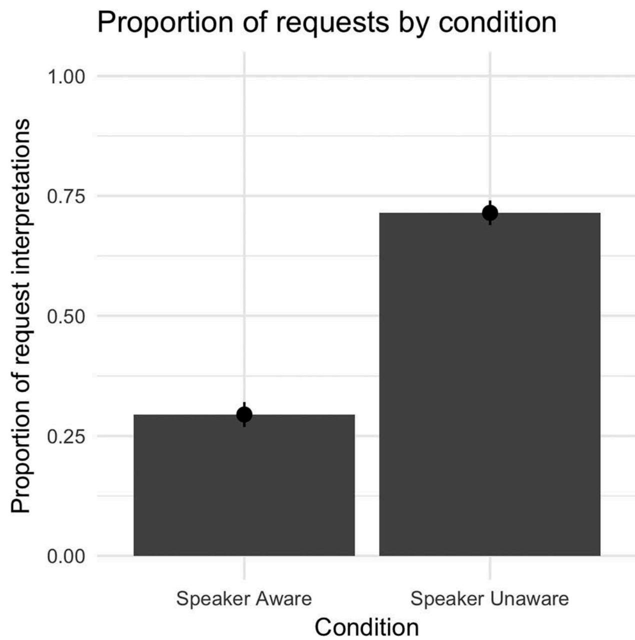

Figure 3. Participants were more likely to interpret an ambiguous utterance as a request if they believed the speaker was unaware of an obstacle to fulfilling that request (M = 0.71, SD = 0.45) than if the speaker was aware of an obstacle (M = 0.29, SD = 0.46).

given to responses that acknowledged he was dissatisfied but not about the relationship specifically; and 0 points were given to answers such as, "He's not having fun fishing."

Unlike the indirect requests task, the Hemingway story remained available to participants to read while they answered the questions, so as to be consistent with the task's implementation in Dodell-Feder et al. (2013). Participants were also asked whether they had read the story before, and if so, how long ago they read it, why they read it, how well they remembered it, whether it was familiar to them, and whether they'd discussed the story with anyone. As mentioned above, the chief difference between the version in Dodell-Feder et al. (2013) and our implementation was that the questions were presented on the browser screen (instead of being asked by the experimenter), and participants typed their responses instead of speaking them aloud.

Results

Before analyzing the data, we attempted to address the potential problem of inattentive subjects. We computed the mean reaction times and their standard deviation per condition, then removed participants who fell three standard deviations above or below the mean on at least twlo trials. This resulted in the removal of two subjects. We also removed three subjects who did not answer all of the questions (either on the indirect requests task, or the Short Story Task). This resulted in a total of 78 participants. No subjects reported having read the Hemingway story before.

Indirect requests task

As in Experiment 1, there were significantly more request paraphrases in the Speaker Unaware condition (M = 0.71, SD = 0.45) than the Speaker Aware condition (M = 0.29, SD = 0.46). This difference is illustrated in Figure 3 above.

To test the main effect of condition predicting the choice of paraphrase, we ran a generalized linear mixed effects model in R (Bates et al., 2015). As before, we began with the maximal model including a fixed effect of speaker knowledge, with random effects for subjects and items, as well as by-subject and by-item random slopes for the main effect of speaker knowledge.

The full model explained significantly more variance than a reduced model that omitted the fixed effect of speaker knowledge but preserved the same random effects structure [$X^2(1) = 14.72$, $p = 0.0001$]. This is a successful replication of the primary finding of Experiment 1; implicit speaker knowledge significantly affected which paraphrase judgment the participant chose. Furthermore, adding an interaction between speaker knowledge and order of item presentation to a model with random intercepts for subjects and items did not significantly improve the model over the model with only fixed effects of speaker knowledge and order, [$X^2(1) = 0.68$, $p = 0.41$]; this demonstrates, as before, that participants did not significantly change their responding behavior over the course of the experiment. To verify that the effect of speaker knowledge was present from the first trial, we ran a generalized linear mixed effects model with a fixed effect of condition and random intercepts for items (as well as by-item random slopes) on only every participant's first trial only. This explained significantly more variance than the null model with only random intercepts for items (as well as by-item random slopes), [$X^2(1) = 4.17$, $p = 0.04$], meaning that most participants were naturally attending to, inferring, encoding, and using the speaker's knowledge states to infer intent.

We also asked participants what they thought the experiment was about. As before, we coded their responses to determine whether they saw through the experimental manipulation. Only 6% of participants mentioned anything about tracking the knowledge states of characters in a story. Just to ensure that the significant findings were not the result of people seeing through the task, we re-ran the generalized linear mixed effects model on the remaining 94% of participants, with speaker knowledge as a fixed effect, random slopes for the effect of speaker knowledge for subjects and items, and random intercepts for subjects and items. The full model explained significantly more variance than the reduced model with only the random effects, [$X^2(1) = 13.4$, $p = 0.0003$].

In Experiment 1, we observed substantial variability among participants in the degree to which speaker knowledge predicted their interpretation of the ambiguous utterance. We observed similar individual variability in Experiment 2 (see Figure 4). Overall, 72.5% of participants showed an effect of speaker knowledge in the expected direction, meaning that when the speaker was aware of an obstacle, these participants were less likely to interpret the utterance as a request. Only 17.5% of participants showed a categorical effect of speaker knowledge in the expected direction, and 27.5% of participants showed a null or negative effect of speaker knowledge.

SST coding and analysis

Participants' responses to the 14 questions were coded by two condition-blind coders according to the rubric provided by Dodell-Feder et al. (2013). These codes were then used to assign each participant three scores: *spontaneous mental state inference, reading comprehension*, and *explicit mental state reasoning*. The first score was binary (whether the participant spontaneously mentalized or not), whereas the second and third were continuous variables. Reading comprehension scores could range from 0 to 10, and explicit mental state reasoning scores could range from 0 to 16.

Inter-rater reliability between the two coders was lower than described in original study (Dodell-Feder et al., 2013), as measured by Cohen's kappa. Inter-rater reliability for explicit mental state reasoning was the highest (.78), followed by reading comprehension (0.67); reliability for spontaneous mental state inference was considerably lower (0.47). Although the participants' score distributions across coders were not identical, there was still a significant correlation for explicit mental state reasoning ($r = 0.87$, $p < 2.2*10^{-16}$), reading comprehension ($r = 0.82$, $p < 2.2*10^{-16}$), and spontaneous mental state inference ($r = 0.52$, $p < 7.9*10^{-7}$). See Figure 5 for a comparison of participants' mental state reasoning scores across the two coders.

To address coder disagreements for the explicit mental state reasoning and reading comprehension scores, we simply took the mean of a participant's score across the two coders. Since disagreement was higher on the spontaneous mental state inference scores, we brought in a third condition-blind coder to act as a tiebreaker.

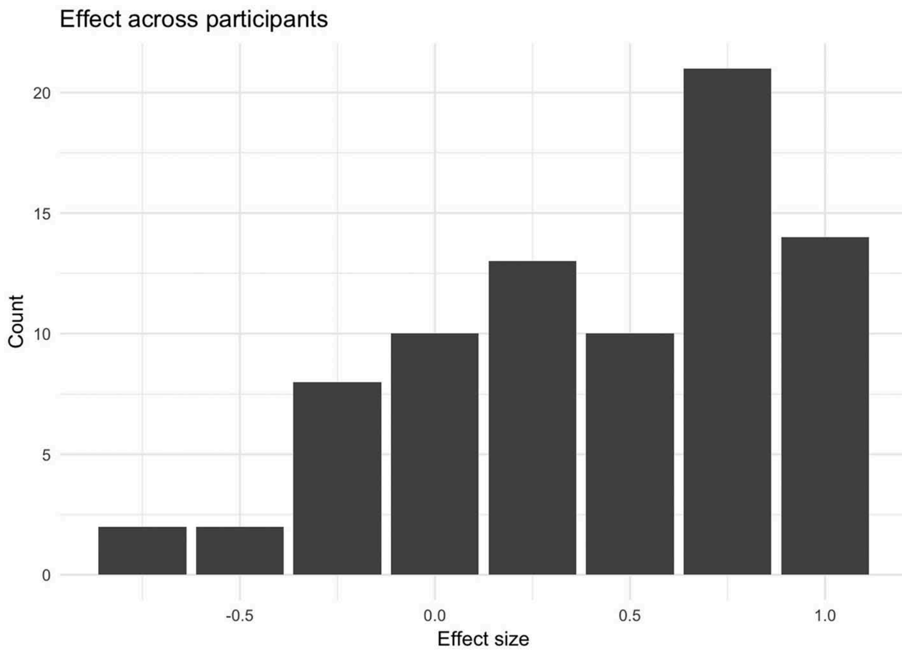

Figure 4. Distribution of the effect of speaker knowledge across participants in Experiment 2. As in Experiment 1, effect size was operationalized as the difference in the proportion of request interpretations across the speaker aware and speaker unaware conditions.

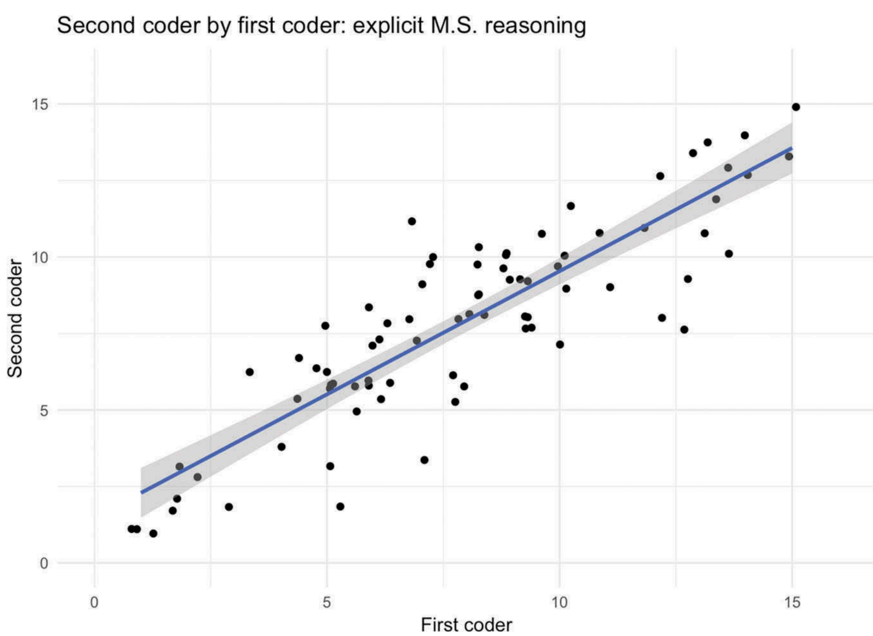

Figure 5. Consistency across first and second coders for explicit mental state (M.S.) reasoning (Cohen's kappa = 0.78, r = 0.87).

Because the SST is a novel instrument for measuring individual variability in mentalizing, we also asked about its internal reliability using Cronbach's alpha. Internal reliability for explicit mental state reasoning was relatively high (*alpha* = 0.78), compared to the internal reliability of 0.54 reported by

Dodell-Feder et al. (2013). Internal reliability for reading comprehension was lower (*alpha* = 0.59), though also higher than the value from the original study, which was 0.31.

Overall, reading comprehension on the SST was quite high (M = 8.9, SD = 1.4). The maximum possible score was 10, and 50% of the participants scored at least a 9.5. The scores ranged from 3.5 to 10. This is consistent with the results of Dodell-Feder et al. (2013), in which participants all scored between 6 and 10 (M = 9, SD = 1.2). Linear regression with heavily skewed data can lead the outliers in a dataset to have a disproportionately large effect. Dodell-Feder et al. (2013) addressed this skew by binning participants into *low* and *high* groups––essentially, those who showed a ceiling effect, and those who did not. To be consistent with the original use of the instrument, we did the same, binning the top 50% of participants into the *high* reading comprehension category, and the bottom 50% into the *low* reading comprehension category.

Explicit mental state reasoning scores were considerably more variable (M = 7.85, SD = 3.3). Out of a maximum possible score of 16, 50% of participants scored above an 8.25, with a range of 1 to 15. This is similar to the findings of Dodell-Feder et al. (2013), in which participants scored between 2 and 14 (M = 8.6, SD = 2.6). Visual inspection of the histogram (Figure 6) suggests the distribution is not as normal as in Dodell-Feder et al. (2013). Critically for our purposes, the presence of substantial variability suggests the task was successful in measuring variation across individuals.

Only 38% of participants made a spontaneous mental state inference (M = 0.38, SD = 0.42). This is less than reported by Dodell-Feder et al. (2013), in which 50% of participants made at least one mental state inference.

We also looked at the relationships between reading comprehension, explicit mental state reasoning, and spontaneous mental state inference. First, we found that *binned reading comprehension* significantly predicted *mental state reasoning scores* [$F(1, 78) = 9.5$, $p = 0.003$], meaning individuals who performed better on the reading comprehension task also had higher explicit mental state reasoning scores. This is in contrast to Dodell-Feder et al. (2013), who found no significant relationship between binned reading comprehension and explicit mental state reasoning. However,

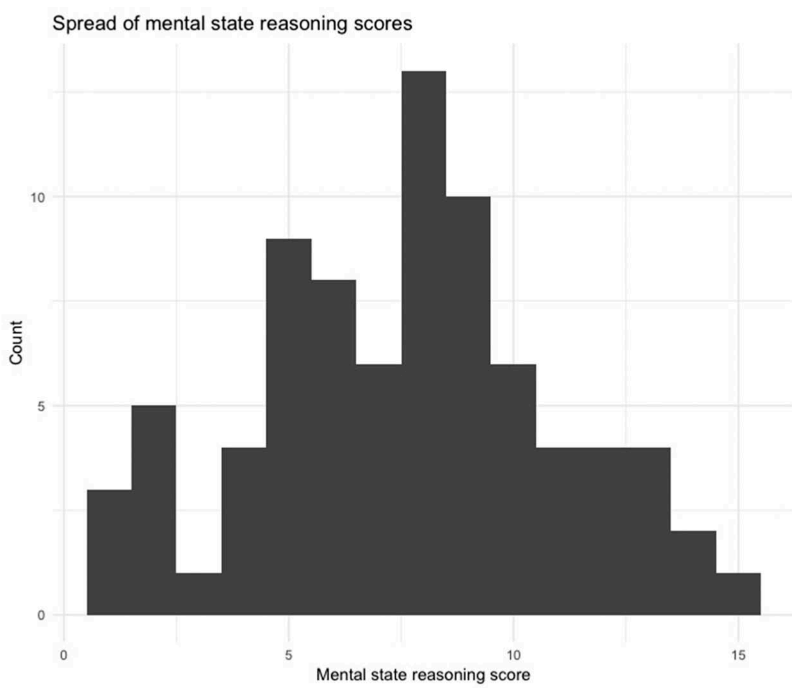

Figure 6. Histogram of mental state reasoning scores across participants (M = 8.6, SD = 2.6).

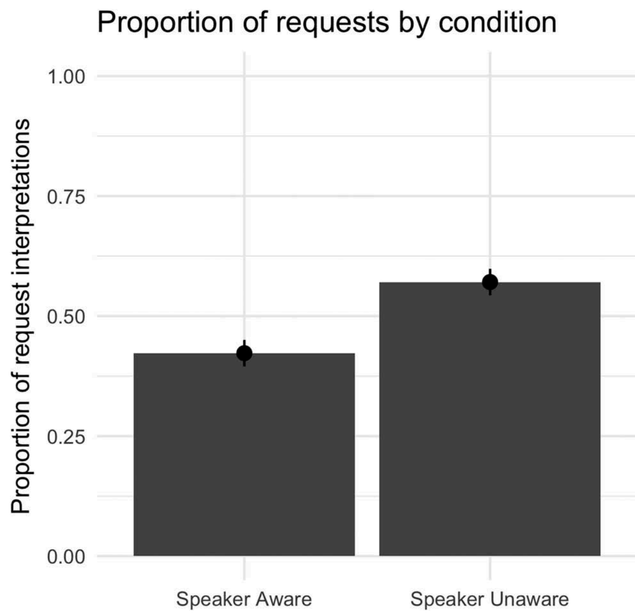

Figure 7. Participants were more likely to interpret ambiguous utterances as requests when speakers were unaware of an obstacle to fulfilling the request (M = 0.57, SD = 0.496) than when speakers were aware (M = 0.42, SD = 0.49).

binned reading comprehension did not significantly predict the likelihood of someone making a *spontaneous mental state inference*, as measured by a generalized linear model with a *logit* link [t (78) = 0.4, p = 0.7]. Lastly, individuals who made *spontaneous mental state inferences* had numerically higher *explicit mental state reasoning scores* (M = 8.8, SD = 2.7) than those who did not (M = 7.3, SD = 3.5), though this difference was only marginally significant [F(1, 78) = 3.7, p= 0.057].

We also asked whether each of the SST variables correlated with demographic variables such as gender and age. Average explicit mental state reasoning for males (M = 7.57, SD = 3.3) was not significantly different from average explicit mental state reasoning for females (M = 7.99, SD = 3.23), [F(2, 77) = 2.6, p= 0.08]. There was also no significant effect of gender in predicting spontaneous mental state inference in males (M = 0.29, SD = 0.46) or females (M = 0.4, SD = 0.5), [F(2, 77) = 0.72, p = 0.49], or binned reading comprehension score [X^2(2) = 1.98, p = 0.4]. Age was not correlated with mental state reasoning ability (r= 0.03, p= 0.79), spontaneous mental state inference (r= 0.09, p= 0.4), or binned reading comprehension (r= 0.15, p= 0.17).

SST and indirect requests

The novel research question of Experiment 2 was whether individual variability in the effect size of speaker knowledge could be predicted by individual differences in mentalizing ability or propensity, when accounting for differences in reading comprehension generally. We operationalized *mentalizing ability* as a participant's explicit mental state reasoning SST score, *mentalizing propensity* as a participant's spontaneous mental state inference SST score, and *reading comprehension* as a participant's reading comprehension score.

Mentalizing ability and indirect requests

Our first question was whether a participant's mentalizing ability predicted the extent to which they incorporated a speaker's knowledge into their interpretation. If mentalizing is an important part of pragmatic inference, then participants who are better at reasoning about the mental states of others should be more likely to factor in a speaker's knowledge state when interpreting their intentions.

Specifically, *explicit mental state reasoning* should interact with *speaker awareness* when predicting a participant's *paraphrase judgment*.

We ran a generalized linear mixed effects model with paraphrase judgment as the dependent variable, an interaction between explicit mental state reasoning and speaker awareness (as well as fixed effects for both), and random slopes for items. We then compared this model to the reduced model with only fixed effects of speaker awareness and explicit mental state reasoning, as well as the random slopes for items and found that the full model explained significantly more variance [$X^2(1) = 16.28$, $p = 5.5*10^{-5}$].

However, it is possible that the improved explanatory power of the model was actually indirectly caused by variability in reading comprehension, since individuals who performed in the top 50% of participants on the reading comprehension task also had higher explicit mental state reasoning scores. In other words, participants may be more likely to integrate a speaker's knowledge states into their interpretation not because of mentalizing in particular, but because they read the passage more closely overall. If this is true, variability in the SST reading comprehension scores should also explain variability in a participant's likelihood to encode and deploy information about the speaker's knowledge, and variability in mentalizing should not provide any *additional* explanatory power beyond variability in binned reading comprehension.

To evaluate this, we tested whether a model including both variability in explicit mental state reasoning and binned reading comprehension explained significantly more variance than a model with only reading comprehension. First, we constructed a generalized linear mixed effects model with paraphrase judgment as the dependent variable, an interaction between binned reading comprehension and speaker awareness (as well as fixed effects for both), and random slopes for items. We found that this model explained significantly more variance than the reduced model with only the fixed effects of speaker awareness and binned reading comprehension [$X^2(1) = 15.302$, $p = 9.2*10^{-5}$]. This suggests that variability in reading comprehension ability partially predicts a comprehender's likelihood of encoding and deploying a speaker's knowledge states.

We then added an interaction between speaker awareness and explicit mental state reasoning (as well as a fixed effect of explicit mental state reasoning), and compared it to a model including an interaction between reading comprehension and speaker awareness and only the fixed effect of explicit mental state reasoning. If variability in participants' performance on the indirect requests task is primarily due to general differences in reading comprehension, the full model including the interaction between mentalizing and speaker awareness should not explain any additional variance. If, however, variability in mentalizing *also* produces differences in the indirect requests task, then the full model with the interaction should explain extra variability beyond variability in reading comprehension.

Indeed, model comparisons revealed that the full model with both measures of individual differences explained significantly more variance than the model including only the interaction with binned reading comprehension [$X^2(1) = 9.2$, $p = 0.002$]. The full model also explained significantly more variance than a model omitting the interaction between reading comprehension and speaker awareness [$X^2(1) = 8.73$, $p = 0.003$]. (Note that the explanatory power of mentalizing did not depend on the dichotomization of the reading comprehension variable; in an additional analysis using an *ordinal rank coding* of reading comprehension, instead of the low/high binning, a full model including the interaction with mentalizing still explained more variance the model with only the fixed effect of mentalizing and the interaction with ranked reading comprehension, [$X^2(1) = 5.6$, $p = 0.02$].)

Although mentalizing ability and reading comprehension are correlated, mentalizing ability explains variance in participants' paraphrase judgments above and beyond that accounted for by reading comprehension differences. In other words, a participant's likelihood of encoding and deploying information about a speaker's knowledge states for their paraphrase judgment is dependent not only on how closely they read the passage overall, but specifically on how skilled they are at inferring and using the mental states of others.

Mentalizing propensity and indirect requests

We also wanted to know whether a participant's propensity to mentalize, as measured by their *spontaneous mental state inference* score, predicted the effect on the first task. We ran a generalized linear mixed effects model with paraphrase judgment as the dependent variable, an interaction between spontaneous mental state inference and speaker awareness (as well as fixed effects for both), and random slopes for items. We then compared this model to the reduced model with only fixed effects of speaker awareness and spontaneous mentalizing, and found that it did not explain significantly more variance, [$X^2(1) = 1.6$, $p = 0.21$]. There was a significant main effect of speaker awareness in the full model [$z = 5.4$, $SE = 0.32$, $p = 7.62*10^{-8}$], no significant main effect of spontaneous mental state inference ($p = 0.45$) and no significant interaction between them ($p= 0.21$). This suggests that a participant's *propensity* to mentalize unprompted did not predict the extent to which they factored in speaker knowledge when interpreting the speaker's intentions.

We also thought it might be possible that spontaneous mental state inference would only cause a difference on the first trial, before participants learned to pay attention to speaker knowledge. To test this, we performed the same analysis as above, but only on the first trial for each participant (and with only random intercepts for items, because there was only one observation for each participant). The full model with an interaction between spontaneous mental state inference and speaker awareness did not explain significantly more variance than the reduced model with only a fixed effect of speaker awareness [$X^2(1) = 0.4$, $p = 0.53$].

Discussion

Our first goal with Experiment 2 was to replicate the findings of Experiment 1. As before, we found that participants were more likely to choose a "request" paraphrase when the speaker was *not* aware of an obstacle to fulfilling the request. This suggests that pragmatic interpretation overall is at least partially modulated by an understanding of what a speaker knows or doesn't know. Furthermore, the fact that this effect was present both on the first trial, and among participants who did not understand what the experiment was about, suggests that participants were naturally attending to and encoding what the characters in a story knew.

As in Experiment 1, there was also considerable individual variability in the degree to which a speaker's knowledge predicted participants' inference. Previous work has suggested that pragmatic inference and mentalizing ability are related. If this is true, then individual variability in mentalizing ability should help explain why certain participants were more likely to factor in a speaker's knowledge when inferring their intentions—above and beyond participants' variability in reading comprehension more generally. Furthermore, we hypothesized at least two mentalizing variables that could predict performance on the task: mentalizing **ability** (how *skilled* someone is at reasoning about the mental states of others) and mentalizing **propensity** (how much someone *tends* to reason about the mental states of others).

We used the SST to capture individual differences in mentalizing ability and propensity via the explicit mental state reasoning and spontaneous mental state inference measurements, respectively. As in the original paper, we found that most participants performed quite well on reading comprehension, suggesting that overall, participants were paying attention to the events in the story, despite the adaptation to a Web-based design. There was considerably more variability in mentalizing ability and propensity, suggesting that this measurement captured something beyond differences in how closely participants read the story overall, but rather how closely they attended to information about the characters' mental states.

Ultimately, we found that while differences in unprompted mentalizing propensity did not predict performance on the first task, differences in mentalizing *ability* did—adding a term for mentalizing ability improved the model beyond the model that just accounted for variability in reading comprehension. This dissociation from reading comprehension shows that the individual participant differences are not just due to how closely they read the stories; even more variability is explained by how skilled participants were at explicitly reasoning about the mental states of others.

These results have several implications for models of pragmatic inference. The overall effect of a speaker's knowledge on participants' judgments of their intentions shows that for the most part, comprehenders are sensitive to changes in speaker knowledge states and they use these changes in speaker knowledge states to adjudicate between the possible interpretations of an utterance. Furthermore, individual differences in mentalizing ability modulated the effect of speaker knowledge on interpretation. This suggests that mentalizing plays a role in the process of inferring speaker intent—whether for encoding information about the speaker's knowledge, actively using that information during inference, or both.

However, there was one concerning limitation to this study. Namely, the paraphrase judgments for both Experiment 1 and Experiment 2 included information about the speaker's knowledge state in the "non-request" option. This means that participants may not have been explicitly adjudicating between "request" and "non-request" interpretations, but rather implicitly adjudicating between speaker knowledge states. Thus, we cannot determine whether mentalizing differences predict differences in how comprehenders incorporate speaker knowledge into their pragmatic interpretations or into their judgments of what people know. To be clear, in either case, the results from Experiments 1 and 2 indicate that comprehenders are making inferences about speaker knowledge states. The question is what they are using that knowledge for. To address this question, we designed Experiment 3.

Experiment 3

In Experiments 1–2, participants' interpretations of an ambiguous utterance was assessed according to which paraphrase judgment they chose. One potential criticism of this methodology is that the non-request paraphrase judgment required specifying the speaker's knowledge state, e.g., "**Since your car is in the shop**, how are you getting to the party?" (See Appendix.) Thus, participants may have been choosing answers based on which knowledge states they had encoded, as opposed to their actual pragmatic interpretation. This is problematic for interpreting how comprehenders are using inferred speaker knowledge for pragmatic interpretation, as well as the individual differences—it could just be that participants with better mentalizing abilities are also better at encoding a speaker's knowledge state in the primary task.

Thus, instead of paraphrase judgments after each passage, participants in Experiment 3 were simply asked: "Is [the speaker] making a request?" If comprehenders use inferred speaker knowledge to make pragmatic interpretations, we should observe a qualitatively similar pattern of results, in which the speaker's awareness of an obstacle to fulfilling a request influences whether a participant answers "Yes" or "No." Participants also completed the Short Story Task, just as in Experiment 2, so that we could assess whether the size of this effect interacted with individual differences in mentalizing ability.

Methods

Participants

Eighty-four new participants were recruited through Amazon Mechanical Turk, all right-handed, native English speakers. We aimed to recruit 80 participants, but Amazon Mechanical Turk over-sampled to 84 participants. Each participant was paid $2.50, and the experiment took on average 22 minutes and 42 seconds minutes to complete. Before excluding participants on the basis of reaction times, there were 46 males, 37 females, and one participant who identified as non-binary. The average age of participants was 34.3 (SD = 9.7), ranging from 20 to 64.

Procedure

Participants first completed a modified version of the task from Experiments 1–2, using the same set of stimuli. Each participant read eight narrative passages, all of which ended with a potential indirect request. After reading each passage, the participant was asked: "Is [the speaker] making a request?" As in Experiments 1–2, the order of the eight passages was randomized. Participants were also randomly assigned to one of two lists, counterbalanced for which items were in which condition.

Participants then completed the modified Web version of the SST (Dodell-Feder et al., 2013); there were no modifications made to the SST between Experiment 2 and Experiment 3. Finally, participants provided demographic information (gender, age, and whether or not they were a native English speaker), and answered what they thought the experiment was about.

Results

Before analyzing the results, we addressed the potential problem of inattentive subjects by removing subjects whose reading times on task 1 deviated from the mean reading time by three standard deviations on more than one trial for both the Speaker Aware and Speaker Unaware conditions. This resulted in the removal of one participant. We also removed one participant who did not answer all of the questions on the first task. No participants reported having read the Hemingway story before.

Indirect requests task

Participants were more likely to interpret an ambiguous utterance as a request when the speaker was unaware of an obstacle (M = 0.57, SD = 0.496) than when the speaker was aware (M = 0.42, SD = 0.49). See Figure 7 for an illustration of this difference.

To test whether speaker knowledge played a significant role in pragmatic interpretation, we constructed a generalized linear mixed effects model with *interpretation* (Yes/No) as the dependent variable, a fixed effect of speaker knowledge, and random intercepts for subjects and items (as well as by-subject and by-item random slopes). We compared this model to a reduced model preserving the same random effects structure but omitting the fixed effect of speaker knowledge, and found that the full model explained significantly more variance [$X^2(1) = 7.8$, $p = 0.005$], suggesting that comprehenders use a speaker's knowledge states to infer their pragmatic intentions. Adding an interaction between knowledge state and order with random intercepts for subjects and items did not explain significantly more variance than a model with only fixed effects for speaker knowledge and item order (with the same random effects) [$X^2(1) = 1.09$, $p = 0.297$]. To verify that the effect was present from the first trial, we constructed a model with a fixed effect of speaker knowledge, as well as random intercepts for items (and by-item random slopes for the effect of speaker knowledge) on only the first trials. This model was only a marginal improvement over a null model with the same random effects for responses on the first trial [$X^2(1) = 2.93$, $p = 0.087$].

We also coded whether or not participants understood what the experiment was about. Only 2.5% of subjects were coded as seeing through the manipulation of speaker knowledge; the other 97.5% gave responses ranging from "I don't know" to "reading and responding to stories," which was the title of the experiment on Amazon Mechanical Turk. To ensure that the primary findings were not a result of people immediately seeing through the experiment, we filtered out participants who understood what the experiment was about and reran the generalized linear mixed effects model with a fixed effect of speaker knowledge, and random intercepts for subjects and items (as well as by-subject random slopes for the effect of speaker knowledge). We compared this model to the reduced model with only the random effects, and found that it explained significantly more variance [$X^2(1) = 14.72$, $p = 0.0001$].

SST coding and analysis

As in Experiment 2, we coded participants' responses to the SST according to the rubric provided by Dodell-Feder et al. (2013). This again yielded three scores: *explicit mental state reasoning, reading* *comprehension*, and *spontaneous mental state inference*. Inter-rater reliability as measured by Cohen's kappa was highest for reading comprehension (0.75), followed by explicit mental state reasoning (0.68), with much lower agreement on spontaneous mental state inference (0.41). We addressed disagreement for explicit mental state inference and reading comprehension by taking the mean of the scores from each coder. Spontaneous mental state inference scores had considerably lower reliability, and were also binary, so a condition-blind third coder adjudicated between the disagreements. Internal reliability for explicit mental state reasoning was relatively high (*alpha* = 0.73), with reading comprehension scoring slightly lower (*alpha* = 0.72).

Participants scored well on reading comprehension overall (M = 8.6, SD = 1.9), consistent with Experiment 2 and Dodell-Feder et al. (2013); 50% of participants scored at least a 9.5 out of 10. As in Experiment 2 and Dodell-Feder et al. (2013), we performed a median split for reading comprehension, splitting participants into *high* and *low* bins. There was more variability for explicit mental state reasoning (M = 7.84, SD = 2.95); 50% of participants scored at least an eight out of 16, with scores ranging from 0 to 14. Visual inspection of the histogram (Figure 8) suggests the data is slightly left-skewed, though it also appears to capture substantial variability across participants. Finally, only 33.3% of participants made a spontaneous mental state inference (SD = 0.47).

As in Experiment 2, binned reading comprehension significantly predicted explicit mental state reasoning [F(1, 78) = 5.5, p = 0.02]. Spontaneous mental state inference score did *not* significantly predict explicit mental state reasoning score [F(1, 79) = 0.724, p = 0.4], but was marginally related to a participant's binned reading comprehension [t(79) = 1.8, p = 0.07].

We also examined whether variability along these dimensions was related to demographic variables. Explicit mental state reasoning for males (M = 7.23, SD = 2.98) was slightly lower than for females (M = 8.56, SD = 2.83), but gender did not significantly predict explicit mental state reasoning [F(2, 78) = 2.8, p = 0.13] or spontaneous mental state inference [$X^2(2)$ = 2.7, p = 0.26]. However, gender did significantly predict binned reading comprehension [$X^2(2)$ = 6.15, p = 0.05], with females scoring on average higher (M = 9.12, SD = 1.5) than males (M = 8.03, SD = 2.02). Finally, age was not correlated with mental state reasoning (r= −0.002, p= 0.9), spontaneous mental state inference (r= 0.01, p= 0.9), or binned reading comprehension (r= 0.1, p= 0.4).

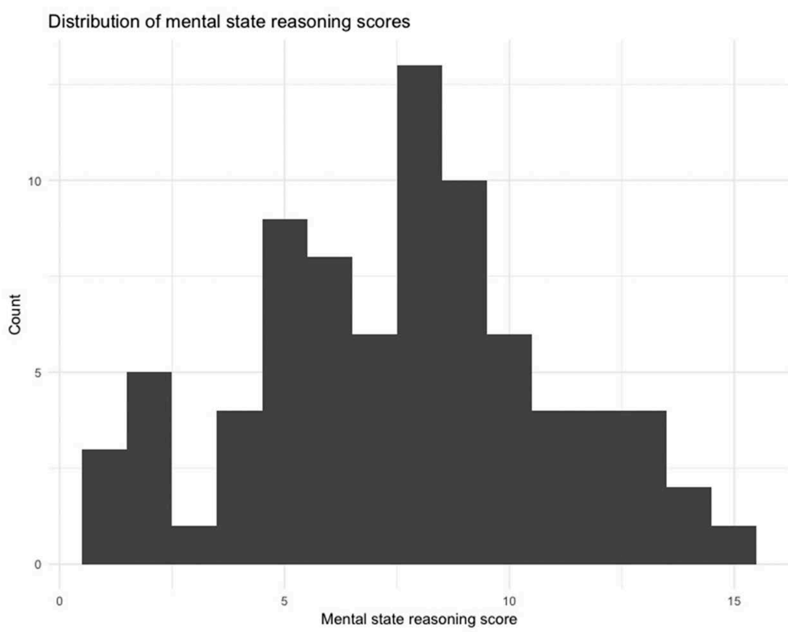

Figure 8. Explicit mental state reasoning scores across participants (M = 7.84, SD = 2.85).

SST and indirect requests

Our other goal with Experiment 3a was to investigate whether the apparent interaction between speaker knowledge and mentalizing ability could be replicated, particularly when the dependent variable was a participant's explicit *pragmatic interpretation*, as opposed to a paraphrase judgment (as in Experiment 2).

Mentalizing ability and indirect request comprehension. We first asked whether a participant's explicit mentalizing ability predicted the extent to which they factored a speaker's knowledge state into their pragmatic interpretations. To test this, we built a generalized linear mixed effects model with pragmatic interpretation as the dependent variable, an interaction between explicit mentalizing ability and speaker knowledge (as well as fixed effects for both), and random slopes for items. We compared this model to a reduced model with only fixed effects of speaker knowledge and explicit mentalizing, as well as random intercepts for subjects and items, and found that the model including the interaction explained significantly more variance [$X^2(1) = 7.64$, $p = 0.006$. As in Experiment 2, adding the interaction term diminished the main effect of speaker knowledge ($p = 0.23$), but there was a significant effect of mentalizing ability [$z = -2.5$, $SE = 0.04$, $p = 0.01$], and a significant interaction between the two [$z = 2.75$, $SE = 0.06$, $p = 0.006$]. In other words, participants with better mentalizing ability were more likely to account for a speaker's knowledge state in their pragmatic interpretations.

Since mentalizing ability was correlated with reading comprehension, we also tested whether explicit mentalizing accounted for more variance than variability in reading comprehension. First, we built a model with an interaction between reading comprehension and speaker knowledge, an interaction between mentalizing and speaker knowledge, fixed effects for mentalizing, speaker knowledge, and reading comprehension, and by-item random slopes for the effect of speaker knowledge. We compared this model to a reduced model which omitted the interaction between mentalizing and speaker knowledge, and found that the full model explained significantly more variance [$X^2(1) = 5.9$, $p = 0.015$]. However, the full model did not explain significantly more variance than a model which omitted only the interaction between reading comprehension and speaker knowledge [$X^2(1) = 0.38$, $p = 0.54$].

Consistent with the findings of Experiment 2, this suggests that the additional variance is best explained by differences in *mentalizing ability* in particular, rather than differences in reading comprehension. For the model with both interactions, the interaction between mentalizing and speaker knowledge was significant [$z = 2.4$, $SE = 0.06$, $p = 0.02$]; there was also a significant main effect of mentalizing ability [$z = -2.7$, $SE = 0.04$, $p = 0.008$], and there was no longer a significant main effect of speaker knowledge ($p > 0.5$). As in Experiment 2, including the interaction with mentalizing significantly improved the model even when reading comprehension was coded as an ordinally ranked variable, [$X^2(1) = 3.8$, $p = 0.05$].

Spontaneous mental state inference and indirect requests

We were also interested once again in whether a participant's spontaneous mental state inference score predicted the extent to which they incorporated speaker knowledge states into their pragmatic interpretation. To operationalize this, we ran a generalized linear mixed effects model with pragmatic interpretation as the dependent variable, an interaction between spontaneous mentalizing and speaker knowledge (and fixed effects for both), and random intercepts for subjects and items. We compared this model to a reduced model with only the fixed effects of speaker knowledge and spontaneous mentalizing, and found that, as in Experiment 2, adding the interaction did not explain significantly more variance in responses [$X^2(1) = 1.96$, $p = 0.16$]. The main effect of speaker knowledge was preserved in the full model [$z = 2.3$, $SE = 0.21$, $p = 0.02$].

Discussion

In Experiment 3, we addressed a limitation of Experiments 1–2, in which the response options after each passage conflated speaker knowledge with pragmatic interpretation. Here, we directly asked participants whether they thought the speaker was making a request, and found that the overall effect of speaker knowledge on pragmatic interpretation was preserved. Furthermore, the size of this effect varied systematically with differences in mentalizing ability—participants with higher mentalizing scores were more likely to factor speaker knowledge into their pragmatic interpretation in the expected direction, above and beyond variability in reading comprehension ability, suggesting that variability on the indirect requests task is well-accounted for by *mentalizing* in particular. Together, these results support the hypothesis that comprehenders use a model of a speaker's knowledge states to infer their intentions, and that the extent to which they do this is related to their mentalizing ability.

One other confound still present in Experiment 3 is that the request interpretation was always the preferred one in the Speaker Unaware condition. Thus, it is possible that comprehenders are simply more likely to interpret ambiguous utterances like "It's cold in here" as requests when the speaker knows less than the comprehender. While this explanation of the results above would still suggest that comprehenders are reliably sampling information about a speaker's knowledge states, it would not demonstrate that they rely on the *content* of a speaker's knowledge states to infer their intent. To address this confound, we designed a follow-up experiment in which the obstacle to fulfilling a request was always common ground, and where speakers were either aware (Speaker Aware) or unaware (Speaker Unaware) of a *solution* to this obstacle. In these passages, the comprehender was always aware of both the obstacle and a solution to the obstacle. As reported in the Supplementary Materials, we found that comprehenders were more likely to interpret ambiguous utterances as requests when the speaker was aware of a solution to the obstacle (M = 0.68, SD = 0.47) than when the speaker was unaware (M = 0.31, SD = 0.47). In other words, the request interpretation was now more likely to be selected when the speaker's knowledge states were aligned with the comprehender's, demonstrating that comprehenders do not just interpret ambiguous utterances as requests when the speaker knows less than them.

Finally, there was a notable change in Experiment 3 (as well as in the experiment reported in the Supplementary Materials) in the *size* of the effect of speaker knowledge. There are several possible explanations for this difference, which will be addressed in the General Discussion.

General discussion

Pragmatic ambiguity pervades everyday conversation. Comprehenders must therefore make frequent inferences about what the speaker intends on the basis of underspecified input. This is especially true in the case of non-conventional indirect requests. We investigated whether comprehenders make these inferences in part on the basis of what they infer the speaker knows and believes about the world, and whether the extent to which they do this depends on their mentalizing ability.

In three experiments, we found that comprehenders' inferences about a speaker's intentions were sensitive to changes in the speaker's implied knowledge states. The same utterance was more likely to be interpreted as a request when the speaker could be inferred to be unaware of an obstacle to fulfilling the request than when the speaker was aware of the obstacle—even though the comprehender critically knew about the obstacle in both situations. Moreover, as reported in the Experiment 3 Discussion, this effect is not simply due to participants being more likely to interpret ambiguous utterances as requests when a speaker has divergent knowledge states. When the speaker's potentially divergent knowledge state concerned a possible *solution* to an obstacle, participants were more likely to interpret the utterance as a request when the speaker was aware of the solution. Furthermore, this effect was robust from the first trial in Experiments 1–2 and in the experiment reported in the Supplementary Materials, and still present (though weaker) in the first trial on Experiment 3. This suggests that at least in these studies, participants spontaneously encoded information about what other characters in a story might know, and used this information for pragmatic disambiguation, despite not having been instructed to pay attention to divergent knowledge states in the stories.

The results of all three experiments point to a role for models of others' mental states in inferring a speaker's intentions. Of further interest, there were considerable individual differences across participants in the extent to which their interpretations could be predicted from the speaker's knowledge states. Some participants performed categorically, always basing their interpretation on what a speaker could be inferred to know, while others seemingly incorporated a speaker's knowledge only some of the time or not at all. In Experiments 2 and 3, we found that these differences could be partially explained by underlying variability in participants' *explicit mentalizing ability*—the ability to encode and represent the belief states of others, as measured by the Short Story Task (SST). To the extent that the SST captures true variability in mentalizing, this suggests that mentalizing ability is an important variable for understanding non-conventional indirect requests—strengthening the link between mentalizing and pragmatic inference. Specifically, better mentalizers are more likely to incorporate information about a speaker's inferable belief states into their pragmatic interpretation.

However, there are still several open questions about the role of mentalizing in pragmatic inference. As noted in the Experiment 3 Discussion, there was a considerable difference in the magnitude of the effect of speaker knowledge between Experiments 1–2 and Experiment 3. The experimental protocols were different: Experiments 1–2 indirectly assessed pragmatic interpretation *and* a participant's ability to identify a speaker's knowledge states, and Experiment 3 directly assessed pragmatic interpretation. This raises the question: Why did participants show a more marked effect in Experiments 1–2 than Experiment 3?

It is unlikely that participants in Experiment 3 were simply worse overall at encoding a speaker's knowledge states, given the similarity in mentalizing distributions across Experiments 2 and 3. Instead, the difference in outcomes is likely due to differences in the tasks. One possible explanation points to distinct states in integrating information about speaker knowledge. As mentioned in the Introduction, engaging mentalizing for pragmatic inference might involve several subprocesses: sampling or encoding speaker knowledge and deploying this information during pragmatic inference. Experiments 1 and 2 could be successfully completed on the basis of the first sub-process alone (knowledge sampling and encoding) because literal paraphrases included descriptions of inferable speaker knowledge. But this was not true for Experiment 3, where responses were high-level speech-act descriptions (e.g. "Is [the speaker] making a request?"). To display an effect of speaker knowledge on Experiment 3, participants had to both encode information about speaker knowledge and later use it for inference. On this reasoning, there must have been some trials in Experiment 3 in which participants successfully made inferences about what the speaker knew, but then did not use this information for inferring intent.

However, the current design does not indicate whether the conversion rate from knowledge to pragmatic inference was consistent across individuals. In principle, it could be that participants varied in how reliably they inferred speaker mental states, but that they integrated that knowledge into their pragmatic inferences at a uniform rate. Better mentalizers might focus more attention on cues that give insight into an interlocutor's mental states, leading them to make inferences about what an interlocutor knows or does not know. Alternatively, participants could vary not just in their likelihood of correctly inferring a speaker's knowledge states from implicit information, but also in their success at deploying this information for pragmatic inference. Mentalizing could affect how likely a comprehender is to *deploy* the information they've encoded about a speaker's knowledge states in the following way: given that two comprehenders have both identified that a speaker is unaware of a broken heater, the better mentalizer might be more likely to find that information relevant (and access it in a timely manner) for inferring what a speaker means when they say, "It's cold in here." Future work should aim to isolate these two stages to better characterize the mechanisms involved in each.

As Trott and Bergen (2017) describe, comprehenders must both: 1) recognize the possible pragmatic interpretations of an utterance; and 2) sample and deploy whatever contextual information helps disambiguate these interpretations. Our results speak to the latter process––information about a *speaker's knowledge state* assists in disambiguation––but do not explicitly speak to *how* or *when* comprenders recognize that an utterance could be a request in the first place. Relevant to the question of *how* is the *obstacle hypothesis* (Gibbs, 1986), which argues that speakers produce requests that address the most salient obstacle in the situational context; these *conventionalized* indirect requests are read faster than less canonical request forms, suggesting that they might be easier to process. None of the requests in our study addressed an obstacle to fulfilling that request, though it is possible that certain utterance-context pairings were more conventionalized than others; this is consistent with the results of Experiment 3, in which some items had a higher rate of request interpretations overall (despite all showing a main effect of speaker knowledge). As of yet, it is unknown whether these differences were due to formal properties of the utterances themselves (e.g., *It's cold in here* might be more recognizable to most comprehenders as a request than *This wine is excellent*), the salience of particular obstacles and solutions in the situational context (e.g., whether a blanket is at hand, even if the heater is broken), or both. In Experiments 1–3, we were specifically interested in the role that a speaker's knowledge states played in pragmatic inference, but future research will investigate the interplay between a model of the speaker and the conventionality of the utterance.

Regarding *when*, it is unknown at what point in the comprehension process participants in our experiments entertained alternative interpretations of a potential request (if at all). In two previous experiments (Coulson & Lovett, 2010; Gibbs, 1979), comprehenders seemed to have *early* access to the request interpretation; this suggests that comprehenders need not always begin with the literal, compositional interpretation of an utterance, then use the *Principle of Relevance* to derive the intended interpretation, as is posited by Relevance Theory (Sperber & Wilson, 1987). Instead, cues in the situational and linguistic context can sometimes provide direct access to the nonliteral interpretation. However, it is unknown whether these findings would extend to cases in which the only disambiguating cue is what the speaker can be inferred to know. Future work should aim to characterize the temporal dynamics of how comprehenders entertain competing interpretations, particularly when the disambiguating feature is information about a speaker's perspective, either by applying time pressure (as described below), or by analyzing the trajectory of mouse movements towards a final interpretation, as has been used in previous work to characterize the time course of certain cognitive processes (Duran, Nicholson, & Dale, 2017; Spivey, Grosjean, & Knoblich, 2005).

Another open question is whether the findings reported here generalize to naturalistic dialogue. The results of Experiments 1–3 demonstrate that reasoning about a speaker's mental states is useful for inferring the intentions of a character in a story, but do not demonstrate that comprehenders rely on mentalizing in actual conversations. There are two reasons why mentalizing might play a reduced role in conversation.

First, conversations are carried out at a very rapid pace; this time pressure could impose constraints in both the production and processing of utterances, leaving interlocutors little time to track divergent knowledge states. As mentioned in the Introduction, previous work (Deliens et al., 2017; Horton & Keysar, 1996) has found that imposing time constraints led to an increase in egocentric biases for both speakers and comprehenders, suggesting that perspective-taking is a cognitively demanding and time-consuming process. In our study, participants were given ample time to both read the passage and interpret the ambiguous utterance—thus, if there are temporally-induced bottlenecks in either the encoding or deployment of speaker knowledge states, our experiment was not designed to discover them. Future work should investigate whether imposing time constraints affects either the encoding of divergent speaker knowledge states, or the deployment of these knowledge states during pragmatic inference. This would also yield a clearer picture of how information about a speaker's mental states is integrated during the process of comprehending non-conventional indirect requests.

Second, the *interactive alignment* account (Pickering & Garrod, 2004) argues that interlocutors in actual conversations do not have to reason extensively about each other's knowledge states, but instead form implicit common ground via iteratively aligning multiple levels of representation—this, in turn, facilitates online predictions about what a speaker will say and mean. The formation of implicit common ground is argued to reduce the chance of misaligned situation models, thereby also reducing the need to track a conversational partner's mental states. In other words, conversation is made *easier* because dynamic interaction facilitates alignment. When misalignment does occur, Pickering and Garrod (2004) argue that it is resolved through an interactive repair process.

The findings reported above are not incompatible with a situated and nuanced version of the interactive alignment account. If our current findings are replicated in naturalistic dialogues, they could be reconciled with the interactive alignment account in the following way. For the most part, low-level mechanisms reduce the chance of misaligned situation models during interaction. In the case of misalignment, at least some comprehenders recruit mentalizing processes to reason about what the speaker may or may not know—particularly when faced with the task of inferring a speaker's intentions from a pragmatically ambiguous utterance. Notably, not all comprehenders do this to the same extent. Thus, when there is both misalignment and insufficient mentalizing on the part of the listener, a *miscommunication* will arise––potentially leading to the need for the interactive repair process described by Pickering and Garrod (2004). Future work should investigate the extent to which listeners in actual conversations track a speaker's knowledge states and use them to differentiate between multiple pragmatic interpretations.

Finally, the results reported here do not speak to whether or how mentalizing is engaged for other kinds of pragmatic inference. Pragmatic ambiguity includes not only other indirect speech acts, such as indirect replies, but also sarcasm, irony, ambiguous reference, veiled threats, and more. Some have argued that understanding ambiguous utterances, such as irony, relies on the attribution of second-order beliefs to the speaker (Winner & Leekam, 1991), but others (Deliens et al., 2017; Keysar, 1994) have found that participants do not always account for differences in mutual knowledge when predicting how a third-party will interpret a potentially ironic utterance. Future work should investigate whether mentalizing and differences in it play a role in understanding other kinds of pragmatic ambiguity. Given the limited bandwidth and ambiguous nature of language, it would hardly be surprising if comprehenders used any and all resources at their disposal to figure out what anyone ever means—and that includes resources evolved for entirely different purposes like inferring the mental states of others.

Acknowledgments

We are grateful to Stefanie Reed, Alex Labranche, and Ankhit Pandurangi for their help in coding the Short Story Task answers.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, *35*(5), 1311. doi:10.1037/0012-1649.35.5.1311

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: Cognitive or affective? *Journal of Autism and Developmental Disorders*, *18*(3), 379–402.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2013). Beyond the language given: The neural correlates of inferring speaker meaning. *Cerebral Cortex*, *24*(10), 2572–2578. doi:10.1093/cercor/bht112

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: Meta-analysis. *Schizophrenia Research*, *109*(1), 1–9. doi:10.1016/j.schres.2008.12.020

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190. doi:10.1016/j.jml.2009.04.003

Brüne, M., & Bodenstein, L. (2005). Proverb comprehension reconsidered—'Theory of mind' and the pragmatic use of language in schizophrenia. *Schizophrenia Research*, *75*(2), 233–239. doi:10.1016/j.schres.2004.11.006

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192. doi:10.1016/j.tics.2008.02.010

Champagne-Lavau, M., & Joanette, Y. (2009). Pragmatics, theory of mind and executive functions after a right-hemisphere lesion: Different patterns of deficits. *Journal of Neurolinguistics*, *22*(5), 413–426. doi:10.1016/j.jneuroling.2009.02.002

Champagne-Lavau, M., & Stip, E. (2010). Pragmatic and executive dysfunction in schizophrenia. *Journal of Neurolinguistics*, *23*(3), 285–296. doi:10.1016/j.jneuroling.2009.08.009

Coulson, S., & Lovett, C. (2010). Comprehension of non-conventional indirect requests: An event-related brain potential study. *Italian Journal of Linguistics*, *22*(1), 107–124.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. doi:10.3758/s13428-014-0458-y

Deliens, G., Antoniou, K., Clin, E., & Kissine, M. (2017). Perspective-taking and frugal strategies: Evidence from sarcasm detection. *Journal of Pragmatics*, *119*, 33–45. doi:10.1016/j.pragma.2017.08.002

Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective pragmatic impairment in autism spectrum disorder: Indirect requests versus irony. *Journal of Autism and Developmental Disorders*, *48*(9), 2938–2952.

Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using fiction to assess mental state understanding: A new task for assessing theory of mind in adults. *PLoS One*, *8*(11), e81279. doi:10.1371/journal.pone.0081279

Duran, N. D., Nicholson, S. P., & Dale, R. (2017). The hidden appeal and aversion to political conspiracies as revealed in the response dynamics of partisans. *Journal of Experimental Social Psychology*, *73*, 268–278. doi:10.1016/j.jesp.2017.07.008

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, *17*(4), 311–318. doi:10.1111/j.1467-9280.2006.01704.x

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

Francik, E., & Clark, H. (1985). How to make requests that overcome obstacles to compliance. *Journal of Memory and Language*, *24*, 560–568. doi:10.1016/0749-596X(85)90046-4

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998. doi:10.1126/science.1218633

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534. doi:10.1016/j.neuron.2006.05.001

Gibbs, Jr, R. W. (1979). Contextual effects in understanding indirect requests. *Discourse Processes*, *2*(1), 1–10.

Gibbs, R. W. (1981). Your wish is my command: Convention and context in interpreting indirect requests. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 431–444. doi:10.1016/S0022-5371(81)90542-9

Gibbs, R. W. (1986). What makes some indirect speech acts conventional? *Journal of Memory and Language*, *25*(2), 181. doi:10.1016/0749-596X(86)90028-8

Gibbs, R. W. (1987). Mutual knowledge and the psychology of conversational inference. *Journal of Pragmatics*, *11*(5), 561–588. doi:10.1016/0378-2166(87)90180-9

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. doi:10.1016/j.tics.2016.08.005

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*(1), 173–184. doi:10.1111/tops.12007

Happé, F., Brownell, H., & Winner, E. (1999). Acquired 'theory of mind' impairments following stroke. *Cognition*, *70*(3), 211–240.

Holtgraves, T. (1994). Communication in context: Effects of speaker status on the comprehension of indirect requests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1205.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117.

Kaplan, J. A., Brownell, H. H., Jacobs, J. R., & Gardner, H. (1990). The effects of right hemisphere damage on the pragmatic interpretation of conversational remarks. *Brain and Language*, *38*(2), 315–333.

Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology*, *26*(2), 165–208. doi:10.1006/cogp.1994.1006

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38. doi:10.1111/1467-9280.00211
Kissine, M., Cano-Chervel, J., Carlier, S., De Brabanter, P., Ducenne, L., Pairon, M. C., … Leybaert, J. (2015). Children with autism understand indirect speech acts: Evidence from a semi-structured act-out task. *PLoS One*, *10*(11), 1–13. doi:10.1371/journal.pone.0142191
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. doi:10.1126/science.aaf8110
Mitchell, P., Saltmarsh, R., & Russell, H. (1997). Overly literal interpretations of speech in autism: Understanding that messages arise from minds. *Journal of Child Psychology and Psychiatry*, *38*(6), 685–691.
Penn, D. C., & Povinelli, D. J. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology*, *58*. doi:10.1146/annurev.psych.58.110405.085555
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27* (2), 169–190.
Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. doi:10.1017/S0140525X00076512
R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/
Roxβnagel, C. (2000). Cognitive load and perspective-taking: Applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, *30*(3), 429–445. doi:10.1002/(SICI)1099-0992(200005/06)30:3<429::AID-EJSP3>3.0.CO;2-V
Schulze, C., Grassmann, S., & Tomasello, M. (2013). 3-Year-old children make relevance inferences in indirect verbal communication. *Child Development*, *84*(6), 2079–2093. doi:10.1111/cdev.12093
Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science*, *1*(2), 260–273. doi:10.1111/j.1756-8765.2009.01018.x
Sperber, D., & Wilson, D. (1987). Précis of relevance: Communication and cognition. *Behavioral and Brain Sciences*, *10*(4), 697–710. doi:10.1017/S0140525X00055345
Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. doi:10.1073/pnas.0503903102
Stefanowitsch, A. (2003). A construction-based approach to indirect speech acts. *Pragmatics and beyond New Series*, *113*, 105–126.
Surian, L. (1996). Are children with autism deaf to Gricean maxims? *Cognitive Neuropsychiatry*, *1*(1), 55–72. doi:10.1080/135468096396703
Tromp, J., Hagoort, P., & Meyer, A. S. (2016). Pupillometry reveals increased pupil size during indirect request comprehension. *The Quarterly Journal of Experimental Psychology*, *69*(6), 1093–1108.
Trott, S., & Bergen, B. (2017). *A theoretical model of indirect request comprehension*. Proceedings of the AAAI Fall Symposium Series on Artificial Intelligence for Human-Robot Interaction (AI-HRI), Arlington, VA.
Turner, R., & Felisberti, F. M. (2017). Measuring mindreading: A review of behavioral approaches to testing cognitive and affective mental state attribution in neurologically typical adults. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.00047
Van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S. A. (2012). Pragmatics in action: Indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive Neuroscience*, *24* (11), 2237–2247. doi:10.1162/jocn_a_00274
Vygotsky, L. S., & Kasanin, J. (1934). Thought in schizophrenia. *Archives of Neurology & Psychiatry*, *31*(5), 1063–1077. doi:10.1001/archneurpsyc.1934.02250050181009
Williams, T., Briggs, G., Oosterveld, B., & Scheutz, M. (2015, January). *Going beyond literal command-based instructions: Extending robotic natural language interaction capabilities*. AAAI, Austin, TX, 1387–1393.
Williams, T., Núñez, R. C., Briggs, G., Scheutz, M., Premaratne, K., & Murthi, M. N. (2014, November). *A dempster-shafer theoretic approach to understanding indirect speech acts*. Ibero-American Conference on Artificial Intelligence, Springer: Cham, pp. 141–153.
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.
Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language*, *62*(1), 89–106. doi:10.1006/brln.1997.1889
Winner, E., & Gardner, H. (1977). The comprehension of metaphor in brain-damaged patients. *Brain*, *100*(4), 717–729.
Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology*, *9*(2), 257–270. doi:10.1111/bjdp.1991.9.issue-2

Appendix

Stimuli for Experiments 1–3.

(1)

Speaker Aware

You and your friend Jonathan are taking a road trip. You began in California, and are now passing through Michigan. It's almost winter, so it's very cold outside—especially for Southern California dwellers like you and Jonathan. You see that you're almost out of gas, so you stop at a gas station in a small town.

You fill up the tank, and then the two of you go inside the gas station to buy some water and snacks. When you return to the car and start up the engine, you and Jonathan notice with some dismay a blinking light, which indicates that the car's heating system is broken. You both bundle up.

As you leave the station, Jonathan shivers in his seat. He turns to you and says, "Man, it's really cold in here."

Speaker Unaware

You and your friend Jonathan are taking a road trip. You began in California, and are now passing through Michigan. It's almost winter, so it's very cold outside—especially for Southern California dwellers like you and Jonathan. You see that you're almost out of gas, so you stop at a gas station in a small town.

While you fill up the tank, Jonathan goes inside to buy some water and snacks. As you're checking the meter, you notice with some dismay a blinking light, which indicates that the car's heating system is broken. You finish filling up the gas and wait for Jonathan.

Jonathan returns with some snacks, and you both set off. As you leave the station, he shivers in his seat. He turns to you and says, "Man, it's really cold in here."

Paraphrase option (Experiments 1–2):

Could you turn on the heater?

I'm really cold; it's too bad the heater is broken.

Experiment 3 prompt:

Do you think he is making a request?

(2)

Speaker Aware

You've been renting a house with your two roommates, Lisa and Brian, for three months now. Things are going well, though both you and Lisa are not very responsible about cleaning; consequently, the house has a tendency to get dirty quickly—unwashed dishes in the sink, cluttered counters, the whole shebang.

Even worse, you and Lisa are directly responsible for the latest mess, which is the product of a recent dinner party. Brian's something of a neat freak, so he's pretty anxious about the mess.

This morning, during a house meeting with you and Brian, Lisa volunteered to take the lead on cleaning the kitchen when she gets home from work this evening.

In the afternoon, while Lisa is still at work, you and Brian are standing in the kitchen doorway, gazing in. Brian turns to you and says, "You guys really did a number on the kitchen."

Speaker Unaware

You've been renting a house with your two roommates, Lisa and Brian, for three months now. Things are going well, though both you and Lisa are not very responsible about cleaning; consequently, the house has a tendency to get dirty quickly—unwashed dishes in the sink, cluttered counters, the whole shebang.

Even worse, you and Lisa are directly responsible for the latest mess, which is the product of a recent dinner party. Brian's something of a neat freak, so he's pretty anxious about the mess.

This morning, while Brian was still sleeping, Lisa volunteered to take the lead on cleaning the kitchen when she gets home from work this evening.

In the afternoon, while Lisa is still at work, you and Brian are standing in the kitchen doorway, gazing in. Brian turns to you and says, "You guys really did a number on the kitchen."

Paraphrase option (Experiments 1–2):

Could you clean up the kitchen when you get a chance?

This kitchen is really dirty; good thing Lisa will take care of it soon.

Experiment 3 prompt:

Do you think he is making a request?

(3)

Speaker Aware

You accompany your teenage daughter, Sarah, to the doctor so that she can be tested for allergies. She's quite reluctant to go, but you promise her that after the appointment, you can go out to her favorite hamburger joint. The appointment goes well, and is very informative; as it turns out, Sarah does have several major food allergies. The doctor gives you a pamphlet containing more detailed information. Then, as promised, you take Sarah to a nearby diner to buy some hamburgers. You order a hamburger and some fries for yourself, a cheeseburger for Sarah, then sit down.

When Sarah sits down, you flip through the pamphlet together; you're both surprised to learn that she's allergic to members of the "nightshade" family, which includes eggplant and potatoes.

About five minutes later, your food is served. Sarah eyes your plate of fries and says, "Man, those fries look delicious…"

Speaker Unaware

You accompany your teenage daughter, Sarah, to the doctor so that she can be tested for allergies. She's quite reluctant to go, but you promise her that after the appointment, you can go out to her favorite hamburger joint. The appointment goes well, and is very informative; as it turns out, Sarah does have several major food allergies. The doctor gives you a pamphlet containing more detailed information. Then, as promised, you take Sarah to a nearby diner to buy some hamburgers. You order a hamburger and some fries for yourself, a cheeseburger for Sarah, then sit down.

While Sarah is washing her hands in the restroom, you flip through the pamphlet; you're surprised to learn that she's allergic to members of the "nightshade" family, which includes eggplant and potatoes.
About five minutes later, your food is served. Sarah eyes your plate of fries and says, "Man, those fries look delicious…"

Paraphrase option (Experiments 1–2):

Could I have one of your fries?

I wish I could have a fry; too bad I'm allergic.

Experiment 3 prompt:

Do you think she is making a request?

(4)

Speaker Aware

After a long week, you decide to treat yourself to a dinner for one at the fanciest restaurant in town. Just for the occasion, you break out your black and white tuxedo. Your spirits are high, but when you arrive at the restaurant, you're somewhat embarrassed to realize that you're wearing the exact same outfit as the restaurant's waitstaff.

While you're waiting in line, you encounter another guest, who initially mistakes you for a waiter. You laugh about the misunderstanding, and then an actual waiter shows you to your table.

Later, as you walk to the kitchen to pay your compliments to the chef, you pass by the same guest's table. He is staring down at his meal with irritation.

As you pass, he looks up at you, recognizes you from the line, and says, "This steak is so overdone!"

Speaker Unaware

After a long week, you decide to treat yourself to a dinner for one at the fanciest restaurant in town. Just for the occasion, you break out your black and white tuxedo. Your spirits are high, but when you arrive at the restaurant, you're somewhat embarrassed to realize that you're wearing the exact same outfit as the restaurant's waitstaff.

While you're waiting in line, you encounter another guest, who initially mistakes you for a waiter. You laugh about the misunderstanding, and then an actual waiter shows you to your table.

Later, as you walk to the kitchen to pay your compliments to the chef, you pass by another guest's table. He is staring down at his meal with irritation.

As you pass, he looks up at you, catches your eye, and says, "This steak is so overdone!"

Paraphrase option (Experiments 1–2):

Could you get me another steak?

Too bad you don't work here, since I really need another steak.

Experiment 3 prompt:

Do you think he is making a request?

(5)

Speaker Aware

After your first year of college, you move back home for the summer. Your parents are glad to have you home, but your dad insists that you do some weekly chores to "earn your keep". You'd prefer to just relax, but it does keep you occupied, which isn't all bad. Besides, the house is in need of some repairs; your parents are both busy people, so some chores have fallen by the wayside while you've been away. Most notably, the lawn has gotten overgrown, and the paint on the backyard fence is beginning to peel. During the first couple of weeks, you mow the lawn, rake the driveway, and water the garden, but you keep putting off painting the fence.

One day, your mom decides to take matter into her own hands.

With your dad's encouragement, she calls a painting company to come and paint the fence; they schedule a time two days from now.

Later in the day, you're reading in the living room. Your dad walks in and peers out the window with a slightly displeased expression on his face.

He turns to you, his hands on his hips, and says, "Man, that fence is looking worse than ever; a new paint-job couldn't come soon enough."

Speaker Unaware

After your first year of college, you move back home for the summer. Your parents are glad to have you home, but your dad insists that you do some weekly chores to "earn your keep". You'd prefer to just relax, but it does keep you occupied, which isn't all bad. Besides, the house is in need of some repairs; your parents are both busy people, so some chores have fallen by the wayside while you've been away. Most notably, the lawn has gotten overgrown, and the paint on the backyard fence is beginning to peel. During the first couple of weeks, you mow the lawn, rake the driveway, and water the garden, but you keep putting off painting the fence.

One day, your mom decides to take matter into her own hands.

While your dad's away at work, she calls a painting company to come and paint the fence; they schedule a time two days from now.

Later in the day, you're reading in the living room when your dad returns from work. He walks in and peers out the window with a slightly displeased expression on his face.

He turns to you, his hands on his hips, and says, "Man, that fence is looking worse than ever; a new paint-job couldn't come soon enough."

Paraphrase option (Experiments 1–2):

When you get a chance, could you paint the fence?

I'm glad someone is coming soon to paint the fence.

Experiment 3 prompt:

Do you think he is making a request?

(6)

Speaker Aware

You and your friend are exhausted and sweaty from playing tennis. After playing, you invite him back to your apartment to hang out.

When you return to your apartment, you both see a notice taped to your apartment door: unfortunately, the water in your apartment building will be shut off for the next 24 hours.

Back in the kitchen, your friend turns to you and says: "I'm so thirsty…"

Speaker Unaware

You and your friend are exhausted and sweaty from playing tennis. After playing, you invite him back to your apartment to hang out.

When you return to your apartment, **you go to your room**, and privately check your email. You see that your landlord has sent you a message: unfortunately, the water in your apartment building will be shut off for the next 24 hours.

You return to the kitchen. Your friend turns to you and says: "I'm so thirsty…"

Paraphrase option (Experiments 1–2):

Could you get me a glass of water?

I am really thirsty, too bad your water is not working.

Experiment 3 prompt:

Do you think he is making a request?

(7)

Speaker Aware

After some deliberation, you invite a new acquaintance over for a dinner date. You prepare a feast, which you serve alongside some red wine. There's only a small amount of wine left in the bottle, but you assure your date you have another bottle stashed away.

About halfway through dinner, you excuse yourself to go to the bathroom. On the way back, you stop in the kitchen to grab the other bottle of wine. You're surprised and irritated to find that there is no other bottle; you must have been mistaken.

Sadly, you call out from the kitchen that the wine is all gone.

When you return to the dining room, your date sips the last of their glass of wine. They look up at you and say, "This wine is excellent."

Speaker Unaware

After some deliberation, you invite a new acquaintance over for a dinner date. You prepare a feast, which you serve alongside some red wine. There's only a small amount of wine left in the bottle, but you assure your date you have another bottle stashed away.

About halfway through dinner, you excuse yourself to go to the bathroom. On the way back, you stop in the kitchen to grab the other bottle of wine. You're surprised and irritated to find that there is no other bottle; you must have been mistaken.

You walk back towards the dining room, unsure about what to tell your date.
When you return to the dining room, your date sips the last of her glass of wine. She looks up at you and say, "This wine is excellent."

Paraphrase option (Experiments 1–2):

Could I have another glass of wine?

This is really good wine; too bad there's no more.

Experiment 3 prompt:

Do you think she is making a request?

(8)

Speaker Aware

You and your friend Rob get along well; your only complaint is that since he doesn't have a car, you almost always have to pick him up and give him rides. Whenever you're both going somewhere, it's generally expected that you'll drive him—but he treats you to coffee sometimes, so it all evens out.

One day, after classes get out, you walk together to get some coffee at nearby coffeeshop.

While you and Rob are sitting talking, you're surprised to see your mechanic at the coffee shop. He walks up and informs you that your car, which you recently took in for a tune-up, will be under repair for at least another week. You rely on that car, so this is very frustrating.

Once the mechanic leaves, you and Rob get to talking. Rob realizes you've both been invited to the same party later on in the week.

Rob says, "So… how are you planning on getting there?"

Speaker Unaware

You and your friend Rob get along well; your only complaint is that since he doesn't have a car, you almost always have to pick him up and give him rides. Whenever you're both going somewhere, it's generally expected that you'll drive him - but he treats you to coffee sometimes, so it all evens out.

One day, after classes get out, you walk together to get some coffee at a nearby coffeeshop.

Before you sit down, Rob steps outside to take a phone call. While he's outside, you're surprised to see your mechanic at the coffee shop. He walks up and informs you that your car, which you recently took in for a tune-up, will be under repair for at least another week. You rely on that car, so this is very frustrating.

Anyway, Rob finishes his phone call and returns to the table, and you both get to talking. Rob realizes you've both been invited to the same party later on in the week.

Rob says, "So… how are you planning on getting there?"

Paraphrase option (Experiments 1–2):

Could I get a ride to the party?

Since your car is in the shop, how are you planning on getting to the party?

Experiment 3 prompt:

Do you think he is making a request?