How do you know that? Automatic belief inferences in passing conversation

Paula Rubio-Fernández [1, 2]
Francis Mollica [3]
Michelle Oraa Ali [1]
Edward Gibson [1]

[1] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology. [2] Department of Philosophy, University of Oslo. [3] Department of Brain and Cognitive Sciences, University of Rochester.

There is an ongoing debate, both in philosophy and psychology, as to whether people are able to automatically infer what others may know, or they can only derive belief inferences by deploying cognitive resources. Evidence from laboratory tasks, often involving false beliefs or visual-perspective taking, has suggested that belief inferences are cognitively costly, controlled processes. Here we argue that in everyday communication belief reasoning is pervasive and likely automatic. To test our hypothesis, participants in a well-powered, pre-registered self-paced reading experiment (N=91) slowed down when a stranger commented 'That greasy food is bad for your ulcer' relative to conditions where a stranger commented on his own ulcer or a friend made either comment – none of which violated participants' common-ground expectations. We conclude that Theory of Mind models need to start investigating belief reasoning in conversation if they want to account for the bulk of the data from everyday social interaction.

Keywords: Theory of Mind; false-belief tasks; automatic and controlled processes; pragmatics; belief inferences.

A simple communicative act such as deciding whom to ask a question requires estimating other people's knowledge. For example, if you got lost in a new city, you may ask a passerby for directions, but if you could not remember your mother's birthday, you would not ask a random passerby on the street (who is unlikely to know your mom). Estimating another person's knowledge (also known as *belief reasoning*) is a key component of human Theory of Mind: our capacity to interpret and predict other people's behavior by reference to their mental states. The large majority of Theory of Mind studies in the last 30 years have investigated the development of belief reasoning in laboratory tasks where a protagonist holds a false belief (e.g., about the location of an object) and the child has to predict the protagonist's course of action, without defaulting to their own knowledge of the situation (Wimmer & Perner, 1983; Scott & Baillargeon, 2017). Rather than using a false-belief task, the present study was the first to investigate belief reasoning in conversation.

Experimental pragmatics studies have long acknowledged the role of belief reasoning in communication (e.g., Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986; Brennan et al., 2010; Heller et al., 2012). Social cognition research, on the other hand, has not investigated belief reasoning in dialogue. We will frame the present study from a Theory of Mind perspective in order to show how investigating belief reasoning in conversation can advance theoretical debates in social cognition research.

By relying on false-belief tasks to investigate humans' capacity for belief reasoning, theoretical models of Theory of Mind fall short of explaining the data observed in everyday communication. Thus, current accounts aim to explain infants' and children's performance in false-belief tasks (e.g., Butterfill & Apperly, 2013; Helming et al., 2014; Heyes & Frith, 2014; Ruffman, 2014; Scott & Baillargeon, 2017) but do not explain the Theory of Mind development reported in the pragmatics literature, which does not always parallel false-belief performance: communication studies often suggest that toddlers have an immature understanding of other people's knowledge (e.g., Dunham et al., 2000; Matthews et al., 2006; Moll et al., 2011). This methodological bias is also problematic for theoretical models of adult social cognition that fail to account for belief reasoning in conversation (e.g., Apperly & Butterfill, 2009; Heyes, 2014) and as a result leave out of their scope the bulk of the data observed in everyday social interaction. Here we argue that belief reasoning is pervasive in communication and therefore needs to be investigated and characterized in conversational settings, and not only in false-belief tasks.

Does communication involve belief reasoning?

A theoretical view that has been gaining momentum in recent years, both in philosophy and psychology, is the hypothesis that attributing mental states to others (not only beliefs, but also intentions and desires) is too cognitively demanding to be the basis for real-time social interaction (e.g., Gallagher, 2001; Bermudez, 2003; Gauker, 2003; Keysar et al., 2003; Pickering & Garrod, 2004; Millikan, 2005; Apperly, 2011; Butterfill & Apperly, 2013; Heyes, 2014; cf. Borg, 2018). Geurts and Rubio-Fernández (2015) have argued that this hypothesis is based mainly on introspection since there is no solid empirical evidence that attributing mental states to others is rare and effortful. More importantly, none of these theoretical accounts offers an explanation as to how adult conversation works the way it does if it does not rely on belief reasoning (cf. Grice, 1989; Sperber & Wilson, 1995). For example, they fail to explain how we make informative contributions to ongoing conversations, remind each other of upcoming events or pre-empt a misunderstanding (all of which require some form of mindreading).

In line with the general view that Theory of Mind inferences are cognitively costly, Apperly and Butterfill (2009; Butterfill & Apperly, 2013) characterize belief reasoning as a System-2 type of reasoning: namely, slow, controlled, flexible, resource-demanding and effortful. They give the example of anticipating what a group of students might know in preparation for a lecture or working out afterwards how one had misjudged their expertise, both of which require deliberative belief-reasoning (2009:966). To challenge the view that belief reasoning may be the basis for real-time communication, Apperly and Butterfill (2009; Apperly, 2017) discuss Keysar et al.'s (2003; Epley et al., 2004) studies using the *director task*, in which a participant follows the instructions of a confederate to move around various objects in a vertical grid of squares. The confederate sits on the other side of the grid and cannot see all of the objects because some of the cells are occluded on her side. Crucially, the confederate is supposed to be ignorant of the contents of those cells, and when she asks the participant to ‘move the small candle’, for example, the smallest of three candles is visible only to the participant. Over a long series of studies, participants have shown a tendency to consider, and sometimes even reach for, the smallest candle in their privileged view before picking up the medium-sized candle in open view (i.e. the one intended by the confederate).

Keysar and colleagues (2003; Epley et al., 2004) interpreted these results as evidence for an egocentric bias in communication, whereby people suffer interference from their own visual perspective when deriving Theory of Mind inferences about their interlocutors. However, other studies using the same paradigm have challenged the view that language comprehension is initially insensitive to perspective taking (e.g., Hannah et al., 2004; Heller et al., 2008). Furthermore, Rubio-Fernández (2017) provided evidence for the view that the director task is an unnatural task that requires selective attention, and not necessarily Theory of Mind. Thus, whereas in everyday life people can normally refer to entities that they cannot see, participants in the director task must assume that the speaker only knows about the objects in her visual field, which poses artificial constraints on referential communication.

Rubio-Fernández and Jara-Ettinger (2018) recently proposed a computational model of referential communication that jointly infers which object the director is referring to and which objects she can and cannot see in a display given her instructions. Model predictions closely mirrored human data, suggesting that belief inferences can be derived as part of the pragmatic process of reference assignment. In other words: we appreciate what others know as part of the process of understanding what they mean, which suggests that belief reasoning need not be an optional, controlled process.

Can belief inferences be derived automatically in conversation?

Apperly argues that belief inferences may be derived spontaneously, but not automatically: “spontaneous belief inferences require some motivation. In an experiment this might be the frequency of judgements about belief. In real life, I am sure that people are frequently motivated to infer what others are thinking. But in the absence of such motivation there is no evidence at all that beliefs are inferred” (2011:95). Apperly (2011, 2017) also claims that experimental paradigms tapping the derivation of pragmatic inferences have in-built incentives to motivate participants to engage in pragmatic reasoning. For example, McKoon and Ratcliff (1986) asked participants to read sentences such as ‘The woman, desperate to get away, ran to the car and jumped in’ and then later recall whether the word ‘driving’ had appeared in the sentence. Participants often responded positively, suggesting they had inferred that the woman drove away in her car. However, according to Apperly (2011:92), these inferences were triggered by the participants’ awareness of the ensuing memory test: without such motivation, this kind of inferences are not automatically derived.

While some forms of belief reasoning may be deliberative (e.g., assessing an audience’s expertise when preparing a talk) and others may be spontaneous (e.g., as when poker players try to guess what the others are thinking), here we want to challenge the view that belief inferences cannot be derived automatically and propose that everyday conversation is the natural arena for testing such a hypothesis. Consider the following example (adapted from Geurts & Rubio-Fernández, 2015; Rubio-Fernández, 2017): imagine that you are eating at a restaurant when a customer at another table tells you ‘That greasy food is terrible for your ulcer’. If this person were a stranger, his comment would immediately strike you as creepy. The reason why you would react with unpleasant surprise is that you would have automatically inferred that this man knew about your health, which is unexpected. That also explains why you would have reacted differently if your best friend had made the same passing remark.

In pragmatics terminology, the stranger's comment would have violated your *common-ground expectations*: we normally assume a certain amount of shared knowledge with our interlocutors (which may range from today's weather to very personal information, depending on how well we know each other), and the stranger's comment would immediately suggest that your common ground was much more extensive than you had first assumed. Our surprise (which is an automatic response) suggests that we monitor common ground by default and a violation of our expectations triggers a belief update. We investigated this hypothesis using a self-paced reading task, but we assume that if a stranger commented on our personal life in a real-life situation, one would also react with surprise.

Experiment

We used self-paced reading to evaluate the automatic vs. spontaneous views of belief reasoning. The automatic view predicts that when strangers refer to the participants' personal life, participants' reading times would be slow relative to conditions where strangers referred to their own lives, or where friends referred to either themselves or the participant. In contrast, the spontaneous view predicts not such difference in the absence of specific motivation to figure out what the speaker is thinking.

The critical argument behind our hypothesis is that belief reasoning can be triggered in conversation without needing to induce an inquisitive mood in the interlocutors, or experimentally motivate them to figure out what another person is thinking (cf. Apperly, 2011). Defendants of the view that belief reasoning requires specific motivation would predict slower reading times when strangers refer to the participant's personal life only if participants were asked to compute the speaker's beliefs (see, e.g., Apperly et al., 2008). Since we did not ask participants to figure out what the speaker knew in each scenario, the spontaneous-process hypothesis would predict comparable reading times in all conditions.

A caveat seems in order at this point. Readers may argue that if we want to defend that Theory of Mind should be investigated in conversational settings, and not only in laboratory tasks, using a self-paced reading task defeats the purpose of our study. However, while we agree that the ideal test case of our proposal would be a study of belief reasoning in naturalistic interaction, testing the difference between the automatic and spontaneous views of belief reasoning requires a controlled experimental setting. We therefore see the use of a self-paced reading task mimicking conversation as a methodological compromise that should allow us to test a very specific research question, while supporting the view that Theory of Mind should be investigated in dialogue.

Methods

Participants

Ninety-two participants were recruited through Amazon Mechanical Turk with the goal of retaining 80 participants. One of the recruited participants did not complete the task and their data were not used. Simulated power analyses based on a pilot study (see Supplemental Material) indicated this sample size would have greater than 80% power to detect our effect, while maintaining a false positive rate less than 5%. The task took approximately 30 minutes and participants were paid $3. Recruitment was limited to participants located in the US territory (according to their IP address) and who had a 95% reliability rate from previous performance on MTurk tasks.

Materials

Materials consisted of short vignettes made of two sentences: (i) a context sentence, which described a scenario in a public space where 'you' (the participant reading) were co-present with another person (the speaker); and (ii) a comment sentence addressed to you by the speaker. There were 24 items in this format in a 2x2 design crossing (1) the participant's relation to the speaker in the vignette (*stranger* vs. *friend*) and the pronoun reference in the target sentence (*my* vs. *your*). Table 1 shows a sample item in the four conditions.

Relation	**Pronoun reference**	**Context**	**Comment**
Friend	Their-life	You are having dinner with your dad at a restaurant when he says:	This greasy food / is terrible / for my ulcer / but it's an / old favorite and / those are hard / to give up.
Stranger	Their-life	You are having dinner at a restaurant when a customer at another table says:	
Friend	Your-life	You are having dinner with your dad at a restaurant when he says:	This greasy food / is terrible / for your ulcer / but it's an / old favorite and / those are hard / to give up.
Stranger	Your-life	You are having dinner at a restaurant when a customer at another table says:	

Table 1: Sample item in the four conditions of the experiment. The context sentence was presented at once, whereas the target sentence was presented in multi-word regions. This way, the critical pronoun (*my*, *your*) was always the second word in Region 4 (e.g., *of my* in the above item). The critical regions for comparison were Regions 5 and on. Regions in the comment are separated by / marks.

The seventh word of the comment sentence was always the critical pronoun (*my*, *your*). Target sentences were divided into multi-word regions for presentation as follows. Region 1 was the context sentence (*You are having dinner with your dad at a restaurant when he says:,* in Table 1) and was presented all at once. Region 2 was the first three words of the comment sentence (*This greasy food*). Region 3 was the next two words (*is terrible*). Region 4 was the next three words (*for my/your ulcer*), whose second word was always the critical pronoun. This was the first region where we might have been able to detect a critical difference in our conditions. The following regions (5-8 in Table 1 and up to as many as 9 for an item) were all three words each. These regions sometimes made up syntactic constituents (as in Region 2 in Table 1, *This* *greasy food*) but sometimes did not (as in Region 6, *old favorite and*), but importantly this was always the same across conditions.

Critical items were divided in four separate lists using a Latin square design so that each participant would see only one version of each item and six critical items from each condition. The two critical variables (speaker relation and pronoun reference) were therefore manipulated within participants.

We also created 24 filler items, which were used in all four lists, for a total of 48 items per list. Matching the targets, there was also a speaker in the filler trials, half of whom were strangers and half friends. Participants were asked a Yes/No comprehension question at the end of each trial. For half of the critical items, the comprehension question (whether the speaker was a stranger (Yes/No) or someone they knew (Yes/No)), served as a manipulation check. Yes/No responses were counterbalanced across items and conditions. All materials, code, data and our pre-registration can be found at OSF (https://osf.io/zq3dg/).

Procedure

On each trial, participants were presented with sequences of dashes (e.g., '--') marking the word-regions of the vignette. Participants were instructed to press the space bar to read through the vignette region by region following a moving window procedure. At the end of each vignette, participants were presented with a Yes/No comprehension question. They made their response using two keys on their keyboard. Stimulus presentation and response collection was controlled by Ibex Farm, a Web-based experimental platform for self-paced reading (Drummond, 2013).

Participants were randomly assigned to a stimulus list, and the order of presentation of the critical and filler items was randomized individually. Participants completed a total of 48 trials. An extra three practice items were included at the start of the task so that participants could accustom to reading on Ibex Farm. Practice and filler trials were not analyzed.

Results

We predicted that the stranger/your-life condition would violate participants' expectations about what other people know about them. If participants monitor their interlocutor's knowledge states by default, they should incur processing costs for updating their common ground in the stranger/your-life condition and read those items slower than the other three conditions. If participants do not keep track of others' knowledge by default, participants should demonstrate no difference in reading times between conditions.

Informed by our pilot study, we predicted reading time differences to manifest after the disambiguating pronoun in Regions 5 and/or 6. We expect this delay due to (i) spillover-effects, which are commonly observed in self-paced reading tasks where slower responses are often observed a few words after the critical region (e.g., Smith & Levy, 2013); and (ii) as a form of pragmatic reasoning, we expect belief inferences to take a second or two to be derived, rather than being computed instantaneously (e.g., Noveck et al., 2001, 2003).

As per our pre-registration, we removed from further analyses all participants who self-reported as non-native English speaker (n=3) and whose accuracy in the comprehension questions was less than 80% on average (n=2). The remaining 86 participants still provide adequate power to detect our effect. Overall, participants were 94% accurate on the task. Reading times +/- 2 SD away from each participant's mean reading time were removed from further analyses to ensure that our conclusions were not driven by outliers. As a result, 4.4% of the data was not subject to further analyses. We report our analyses following our pre-registration. However, the trends hold even when all data are retained (see Supplementary Material).

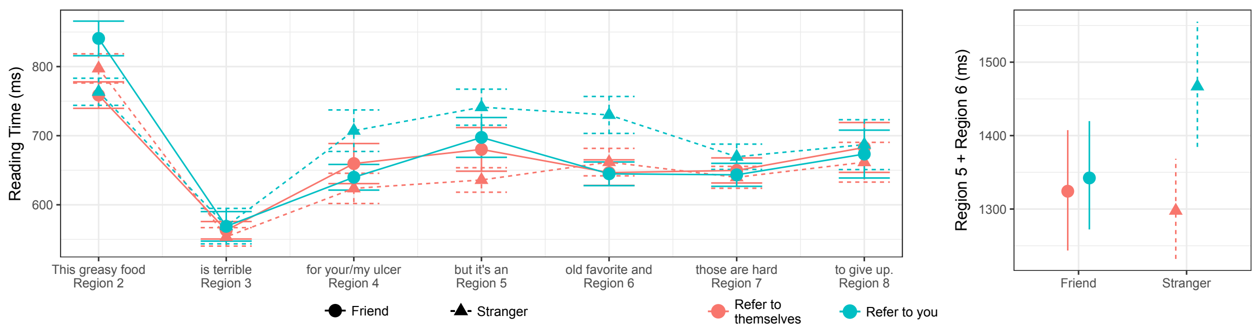

Figure 1. Left plot: Mean Reading Times for critical conditions by region. Region 1 corresponds with the context sentence and is not included in the graph because it was presented at once and reading times were considerably longer. Error bars reflect SEM. Right plot: Summed reading times for Regions 5 and 6 combined. Line ranges reflect 95% bootstrapped confidence interval.

	Estimate	**Std. Error**	***t* Statistic**
5th Region			
Intercept	688.93*	39.94	17.25
Pronoun	61.89*	25.66	2.41
Relation	0.09	27.22	0.00
Pronoun x Relation	88.77	52.66	1.69
6th Region			
Intercept	670.84*	39.56	16.96
Pronoun	34.01	19.85	1.71
Relation	50.79*	21.79	2.33
Pronoun x Relation	70.14*	35.08	2.00
5th + 6th Region			
Intercept	1357.92*	76.22	17.82
Pronoun	94.21*	36.40	2.59
Relation	49.07	36.20	1.36
Pronoun x Relation	151.40*	63.01	2.40

Table 2. Coefficients and *t*-values from linear mixed effect models.

Figure 1 displays the average reading times for each region and the combined reading times for our regions of interest. As can be seen in the rightmost panel, participants were slower to read our regions of interest in the stranger/your-life condition. Following our pre-registration, we conducted three linear mixed effect regression models predicting Reading Time for (1) 5th region,

(2) 6th region and (3) the sum of the 5th and 6th region, with fixed effects for Relation and Pronoun, and their interaction. The maximal random effect structure was utilized — i.e. random intercepts and slopes for Item and Participant. Relation and Pronoun were sum coded. The model was fit using the lme4 package (Bates et al., 2015) in R (R Core Team, 2018).

The parameter estimates can be found in Table 2. As predicted, we find a significant interaction such that reading times for the stranger/your-life condition were significantly longer than reading times to the other three conditions for both Region 6 and Regions 5 and 6 combined, consistent with a default preference to monitor our interlocutor's knowledge states. We do not interpret the main effects as they contradict across regions and are most likely driven by the predicted interaction.

General discussion

The results of our study suggest that people can derive belief inferences automatically in conversation: when someone makes a passing remark about some private matter, we immediately understand that they know about our personal life. This kind of belief inference makes up our common ground with our closest interlocutors, and normally go unnoticed in conversations with friends and family. However, when our participants inferred that a stranger had remarked on their personal lives, the same inference was unexpected, which slowed down their reading times, as we had predicted.

Like other defendants of the view that belief reasoning is cognitively costly, Apperly (2011) argues that people must be specifically motivated to infer what others are thinking in order to be able to spontaneously infer beliefs. However, our task did not specifically incentivize participants to figure out what the speakers knew or did not know. More generally, our results are not limited to a laboratory setting since people would react with surprise if strangers commented on their personal lives in a public place. Likewise, our results are not limited to conversations with strangers: one would also react with surprise if a close friend made a passing remark that suggested they knew about a secret (e.g., 'How many people are coming to my surprise birthday party?'). One way to accommodate our results with Apperly's view would be to argue that whenever we engage in conversation, we are intrinsically motivated to figure out what our interlocutor is thinking (or trying to communicate), in line with mindreading accounts of communication (Grice, 1989; Sperber & Wilson, 1995). Another way, more in line with Apperly's own proposal, would be to try to determine which aspects of communication may become automatized during communicative development. For example, common ground violations may be automatically detected through associative memory representations (see Gerrig & Horton, 2016).

Overall, our results highlight how Theory of Mind research needs to start including pragmatic measures of belief reasoning in conversation. Otherwise, theoretical models may account for the results of false-belief tasks and other experimental tests, but not for the bulk of the data from everyday social interaction.

Acknowledgements

This research was supported by a *Young Research Talent Grant* from the Research Council of Norway awarded to PRF (230718).

References

Apperly, I. A. (2011). *Mindreaders: The cognitive basis of Theory of Mind*. Hove, UK: Psychology Press.

Apperly, I. A. (2017). Mindreading and psycholinguistic approaches to perspective taking: Establishing common ground. *Topics in Cognitive Science*, 1-7.

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential Theory of Mind task. *Cognition*, *106*, 1093-1108.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*, 953-970.

Bates, D., Machler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48.

Bermúdez, J. L. (2003). The domain of folk psychology. *Royal Institute of Philosophy Supplement*, *53*, 25-48.

Borg, E. (2018). On deflationary accounts of human action understanding. *Review of Philosophy and Psychology*, 1-20.

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialogue: Coordinating speaking and understanding. In *Psychology of Learning and Motivation* (Vol. 53, pp. 301-344). Academic Press.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal Theory of Mind. *Mind & Language*, *28*, 606-637.

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

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Dunham, P., Dunham, F., & O'Keefe, C. (2000). Two-year-olds' sensitivity to a parent's knowledge state: Mind reading or contextual cues? *British Journal of Developmental Psychology*, *18*, 519-532.

Drummond, A. (2013). Ibex farm. Online server: http://spellout. net/ibexfarm

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*, 327-339.

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

Gauker, C. (2003). *Words without meaning*. Cambridge, Massachusetts: MIT Press.

Geurts, B., & Rubio-Fernández, P. (2015). Pragmatics and processing. *Ratio*, *28*, 446-469.

Grice, P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*, 105-115.

Heller, D., Gorman, K. S., & Tanenhaus, M. K. (2012). To name or to describe: Shared knowledge affects referential form. *Topics in Cognitive Science*, *4*, 290-305.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*, 831-836.

Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, *18*, 167-170.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, *9*, 131-143.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, *344*, 1243091-1-6.

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the Memory-Based Processing Approach to Common Ground. Topics in cognitive science, 8(4), 780-795.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on Theory of Mind use in adults. *Cognition, 89*, 25-41.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, *27*, 403-422.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*, 82-91.

Millikan, R. G. (2005). *Language: A biological model*. Oxford University Press.

Moll, H., Carpenter, M., & Tomasello, M. (2011). Social engagement leads 2-year-olds to overestimate others' knowledge. *Infancy*, *16*, 248-265.

Noveck, I. A., Bianco, M., & Castry, A. (2001). The costs and benefits of metaphor. *Metaphor and Symbol*, *16*, 109-121.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*, 203-210.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169-190.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rubio-Fernández, P. (2017). The director task: A test of Theory-of-Mind use or selective attention? *Psychonomic Bulletin & Review*, *24*, 1121-1128.

Rubio-Fernández, P., & Jara-Ettinger, J. (2018). Joint inferences of speakers' beliefs and referents based on how they speak. *Proceedings of the 2018 Meeting of the Cognitive Science Society*, Madison, Wisconsin, USA.

Ruffman, T. (2014). To belief or not belief: Children's Theory of Mind. *Developmental Review*, *34*, 265-293.

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, *21*, 237-249.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*, 302-319.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford, United Kingdom: Blackwell.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103-128.