CrossMark

BRIEF REPORT

The director task: A test of Theory-of-Mind use or selective attention?

Paula Rubio-Fernández[1,2] ⓘ

Abstract Over two decades, the director task has increasingly been employed as a test of the use of Theory of Mind in communication, first in psycholinguistics and more recently in social cognition research. A new version of this task was designed to test two independent hypotheses. First, optimal performance in the director task, as established by the standard metrics of interference, is possible by using selective attention alone, and not necessarily Theory of Mind. Second, pragmatic measures of Theory-of-Mind use can reveal that people actively represent the director's mental states, contrary to recent claims that they only use domain-general cognitive processes to perform this task. The results of this study support both hypotheses and provide a new interactive paradigm to reliably test Theory-of-Mind use in referential communication.

Keywords Perspective taking and perspective updating · Selective attention · Mentalizing vs. submentalizing · Task demands · Suspicion · Automatic pragmatic inferences

In the referential communication task designed by Krauss and Glucksberg (1977), pairs of participants sat on opposite sides of a blind and had to verbally coordinate to arrange a series of drawings according to a chart that was available to only one of them ("the director"). Krauss and Glucksberg reported that adults "found the task trivially easy…and made

✉ Paula Rubio-Fernández
   prubio@mit.edu

1 Department of Brain & Cognitive Sciences, Massachusetts Institute
   of Technology, 43 Vassar St., Building 46,
   Cambridge, Massachusetts 02139, USA

2 Centre for the Study of Mind in Nature, University of Oslo,
   Georg Morgenstiernes hus, Blindernveien 31, Oslo 0371, Norway

virtually no errors on their very first try" (p. 101). In the late nineties, Keysar and colleagues (e.g., Keysar, 1997; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003) adapted the referential communication task to investigate language comprehension using eye tracking. It is somewhat ironic, given the good performance initially observed, that since the referential communication task was adapted for eye tracking, it has become a test of miscommunication.

In the eye-tracking version, commonly known as the "director task" (DT), 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, for example, she asks the participant to "move the small candle," 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 (e.g., Barr, 2008; Begeer, Malle, Nieuwland, & Keysar, 2010; Converse, Lin, Keysar, & Epley, 2008; Keysar et al., 2000, 2003; Lin, Keysar, & Epley, 2010).

Participants' performance in the DT was interpreted by Keysar and colleagues as evidence for an "egocentric bias" in communication, according to which listeners initially comprehend language egocentrically. This view has sparked a long-standing debate with other researchers, who have argued that listeners use common-ground information from the earliest stages of language interpretation (e.g., Brown-Schmidt, 2012; Hanna & Tanenhaus, 2004; Heller, Grodner, & Tanenhaus, 2008; Nadig & Sedivy, 2002). In the present study, the aim was not to contribute to this psycholinguistic debate, but rather to challenge Keysar et al.'s view that the DT is a reliable test of Theory-of-Mind use in communication.

_Springer

The director task in social cognition research

Psycholinguistic studies have always ensured that participants believed that the confederate director was a naive participant (e.g., Barr, 2008; Keysar et al., 2003). However, it has been shown that using confederates, instead of pairs of naive participants, may compromise the reliability of experimental pragmatics studies (Kuhlen & Brennan, 2010; Lockridge & Brennan, 2002). Recent social-cognition studies have disregarded these earlier studies by using computer versions of the DT, in which participants have to pretend that a static human figure depicted behind a grid is the director (e.g., Apperly et al., 2010; Dumontheil, Apperly, & Blakemore, 2010; Dumontheil, Küster, Apperly, & Blakemore, 2010; Santiesteban, Shah, White, Bird, & Heyes, 2015; Symeonidou, Dumontheil, Chow, & Breheny, 2016; Wang, Ali, Frisson, & Apperly, 2016). Apperly et al. admitted that this setup may not be a naturalistic test, but they nonetheless concluded that their results support Keysar et al.'s view that adults are rather poor at using Theory-of-Mind inferences in language interpretation.

Apperly et al. (2010, Exp. 3; Dumontheil, Apperly, & Blakemore, 2010; Symeonidou et al., 2016) observed that participants suffered more interference from their privileged perspective when following the instructions of an avatar director than when applying an arbitrary rule to ignore the objects in the occluded/dark-background cells. Apperly et al. argued that only the former condition requires perspective taking, and they interpreted their results as evidence that Theory-of-Mind inferences are costly.

Santiesteban, Shah, White, Bird, and Heyes (2015) also compared two conditions, one with a human figure as the director and another with a picture of a camera instead of a director. Through this study they aimed to test Heyes's (2014) claim that people often "submentalize" and use domain-general cognitive mechanisms in social situations, rather than representing other people's mental states (i.e., mentalizing). Santiesteban et al. argued that participants in the DT use object-centered spatial coding rather than perspective taking, and explained the better performance observed by Apperly and colleagues in the arbitrary-rule condition as the result of participants' not requiring object–spatial coding to perform the task (rather than not requiring perspective taking).

As they had predicted, Santiesteban et al. (2015) observed similar performance with both the director and the camera, in both neurotypical adults and adults with autism spectrum disorders, and interpreted their results as supporting evidence that "adults use mentalizing sparingly in psychological experiments and in everyday life" (p. 844). Before reaching these conclusions, Santiesteban et al. argued against the possibility that participants may have mentally represented that both the director and the camera were able to see, and therefore treated them equally for that reason. In my view, the real challenge to

Santiesteban et al.'s conclusions comes from the reverse argument: The reason why participants may have treated the human figure and the camera alike is because neither had mental states that they could represent without pretense.

Although it is convenient for psychological experimentation, the use of avatar directors in studies of referential communication is analogous to testing the physical abilities of a tennis player against a wall. Heyes (2014) has recently pointed out that various experimental paradigms that use avatars and have allegedly shown that adults represent other people's mental states automatically need to use inanimate controls to rule out nonmentalistic interpretations of their results. Playing again on reverse arguments, social-cognition studies using the DT with avatars need to use *human* controls before they can conclude that pragmatic inferences are effortful, or that interlocutors only sparingly represent each other's mental states in communication. The aim of the present study was to provide such a control.

Aims and scope of the DT

To determine whether interlocutors take each other's perspectives in communication, experimental studies must ensure that the speaker's and hearer's perspectives are different; otherwise, interlocutors may be using their own perspective as a default (Keysar, 1997). In this respect, the DT is a suitable test of perspective taking in communication. However, by keeping the director and the participant's perspectives apart, the DT also makes demands on participants' executive control, potentially taxing their performance independently of their theory-of-mind use. This argument has previously been used to explain young children's failure in false-belief tasks (e.g., Baillargeon, Scott, & He, 2010). By contrast, adults' poor performance in the DT has generally been interpreted as a failure to use their Theory of Mind in communication, rather than as a possible effect of the high executive control demands of the task.

Lin, Keysar, and Epley (2010) showed that performance on the DT does rely on participants' attentional resources (see also Brown-Schmidt, 2009; Symeonidou et al., 2016). However, rather than challenging the general assumption that poor performance in the DT must reveal limited use of Theory of Mind (vs. heavy demands on executive control), Lin et al. concluded that using theory of mind in communication is generally effortful. This broad conclusion is clearly not warranted by Lin et al.'s results, since they only used the DT in their study and did not include any other pragmatic tasks that might be more naturalistic and less dependent on attentional resources.

Building on the executive control argument, I want to further argue that the design of the DT prevents participants from relying on a universal assumption in human communication: namely, that people know more than what they can see and are

therefore able to refer to entities outside their visual field. Participants engaged in the DT are expected to suppress this basic truth and assume that the director only knows about the objects she can see on the grid. Therefore, although the DT may seem "uncomplicated" (Apperly & Butterfill, 2009:959), it is in fact a highly artificial test of referential communication.

In what follows, I will challenge the view that adults' difficulties with the DT are evidence of "limited use of Theory of Mind in communication" (Apperly et al., 2010; Keysar et al., 2003), and instead will argue that the design of the DT itself is what limits participants' perspective-taking abilities, by imposing artificial demands on their selective attention. To test the view that the design of the DT taxes normal pragmatic processes, I designed an interactive version of the DT that aimed to challenge the key assumption in this task: that when participants consider the hidden objects in the grid as possible referents for the director's instructions, they are not necessarily failing to use their Theory of Mind.

Pairs of naive participants played this new DT on two computers, on whose screens they saw 2 × 2 grids of objects. One of the four cells in each grid had a gray background and contained an object in the follower's grid, but was empty in the director's (see Table 1). The Critical trials included a subtle manipulation: Unbeknownst to the participants, the position of the gray cell was shifted in the director's grid, so that a figure that appeared on a gray background in the follower's grid now appeared on a white background in the director's. The director would see two fish of different colors, for example, and ask the follower for "the orange fish." However, in the follower's grid the blue fish appeared on a gray background, thus inviting the question: If the director cannot see the blue fish, why does she call the target "the orange fish," and not just "the fish"?

This is a subtle manipulation that need not immediately give away the director's perspective, since it has been extensively documented that speakers tend to use color redundantly in referential communication; that is, people will often refer to "the orange fish" in a display with only one fish (e.g., Rubio-Fernández, 2016; Sedivy, 2003). Therefore, if followers were to suspect that the director sometimes knows what is in the gray cell, that would be evidence that people use their Theory of Mind to actively represent the director's perspective.

This pattern of results would challenge Santiesteban et al.'s (2015) claim that participants engaged in the DT only submentalize. In other words, if participants failed to use their Theory of Mind in this new DT, they would not be able to question what the director knows; therefore, the use of domain-general cognitive mechanisms (along the lines suggested by Santiesteban et al. and by Heyes, 2014) would not produce positive results in this version of the DT.

On the other hand, since the grids contained a single gray cell, followers could afford to adopt a selective-attention strategy and consciously focus on the three white cells. That is

clearly an efficient strategy in the DT, yet also one that would prevent followers from suspecting that the director might know about the contents of the gray cell. This potential selective-attention strategy therefore raises a critical issue: According to the egocentric view defended by Keysar et al. (2003) and Apperly et al. (2010), those followers who managed to block the gray cell from their view would be "model perspective-takers," because they would suffer no interference from their privileged perspective. However, they would actually be *underusing* their Theory of Mind. This would show as a proof of concept—not yet of fact—that optimal performance on the standard DT may only reveal selective attention, and need not require the use of Theory of Mind.

The latter pattern of results would therefore support the hypothesis that optimal performance on the DT, as established by the standard metrics of interference, is possible by selectively focusing on those objects that the director can see in the grid. This would mean that the looking behavior and response times observed in the DT are not reliable indicators of whether or not participants are using their Theory of Mind, since they might also be using (or failing to use) their selective attention. Without further measures of Theory-of-Mind use (such as the ones incorporated in the study below), it would simply not be possible to conclude that adults' poor performance on the DT reveals limited use of Theory of Mind in communication.

Method
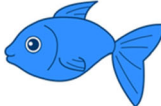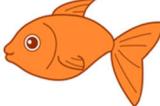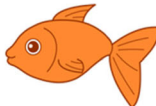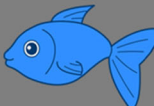
Participants

A total of 42 students from University College London (UCL) participated in the experiment for monetary compensation (£4). They were all proficient speakers of English (as is required by UCL), and 18 of them were native speakers. The participants were recruited in pairs, and knew each other in four instances. On arrival, the participants were randomly assigned the roles of director and follower.

Materials

In all, 32 slides were constructed for the follower role, each including a 2 × 2 grid with four pictures. The cell with a gray background represented the follower's privileged ground, and its position was counterbalanced across trials. The slides were divided into four conditions, depending on the contents of the follower's grid: In the Filler 1 condition (12 slides), the grids included four different objects. In the Filler 2 condition, the grids included two similar objects of different sizes (three slides) or different colors (three slides) on a white background. In the Baseline condition (seven slides), the grids included two identical objects, one on a gray background and one on a white background. In the Critical condition (seven slides),

Table 1   Sample displays from the director's and the follower's perspectives

Condition	Director's perspective	Follower's perspective
Filler 1 (No contrast)		
Filler 2 (Contrast)		
Baseline		
Critical		

The asterisk indicates the target in each trial.

the grids included two similar objects of different colors, one on a gray background and one on a white background (see Table 1).

All of the Baseline and Filler 2 trials, plus three of the Filler 1 trials, were presented before the Critical trials, to test whether in the second half of the task the followers would grow suspicious that the director sometimes could see the contents of the gray cell. The numbers of Baseline and Critical trials were relatively low (seven each) so as not to give participants too many opportunities to suspect that the director could see what was in the gray cell. Unlike the Critical condition, the Baseline condition did not include two similar objects of different colours, to avoid the possibility that when the director started using color adjectives in the Critical condition, followers would simply notice the difference from the earlier Baseline trials. Similarly, the point of the Filler 2 condition was to familiarize the followers with modified descriptions, so that the instructions of the Critical condition would not attract

unnecessary attention. The slides were always presented in the same pseudorandom order.

Procedure

Participants were told that they were going to play a coordination game in which one of them (the director) would give instructions to the other (the follower) to click on a target figure in a display. The experimenter used a color printout of two grids from the Baseline condition to explain that the director's and the follower's grids differed in three ways: first, the follower always saw four objects, while the director only saw three, because her gray-background cell was empty. Second, only the director's grid showed an asterisk next to the target. Third, the position of the objects was scrambled so as to prevent the director from using spatial coordinates. Note that these three differences were maintained in the Critical condition (despite the shifting of the gray cell).

The director and the follower sat across from each other, the former performing on a laptop computer and the latter on an eye-tracking computer. The eye-tracking system was a RED-m by SMI (SensoMotoric Instruments GmbH, Teltow, Germany), which measured eye position at a sampling rate of 120 Hz with a spatial resolution (root-mean square) of 0.1° and an accuracy of 0.5°, and which allowed free head movement during the recording. The experiment lasted approximately 12 min.

Predictions

The aim of this experiment was to test whether followers would be able to take the director's perspective while keeping track of the contents of the gray cell. Doing so in this version of the DT might cause followers to become "suspicious" of the director's perspective. Two measures of suspicion were used, one direct and one indirect. The direct measure was obtained from a posttest questionnaire. The indirect measure of suspicion was the amount of attention that followers paid to the gray cell in the course of the task.

The questionnaire asked followers (a) whether they thought there was anything peculiar in the way the director formulated the instructions; if so, (b) what was peculiar about the instructions, and (c) whether they suspected that the director could see what was in the gray cell (explain why and give examples). If followers responded negatively to Question (a), they were then asked (b') to explain any strategy they may have used, and Question (c).

Two patterns of results were predicted: First, those participants who had not noticed anything peculiar in the director's instructions would not suspect that she had sometimes known about the contents of the gray cell. Presumably, these "unsuspicious" participants, as I will call them, tried to ignore the gray cell. Their eye movements would reveal generally low proportions of fixations on the gray cell, possibly decreasing during the task due to practice.

Second, those participants who kept track of the contents of the gray cell would have grounds to suspect the director's perspective. In the first half of the task, their eye movements should also reveal decreasing attention to the gray cell, due to practice, but in the second half, the eye movements should reveal a growing interest in the contents of the gray cell, due to the suspicion. Overall, the participants' proportions of fixations on the gray cell should be roughly U-shaped across the Baseline and Critical conditions.

Results

Of the 21 followers in the study, only four failed to notice anything peculiar in the director's instructions (19 %). As expected, these unsuspicious participants reported having tried to ignore the gray cell as a potential distractor. The remaining 17 participants noticed that the director included too much information in the instructions. Of these participants, only one did not suspect that the director could see what was in the gray cell, and explained that she thought the director was being helpful by describing the targets in too much detail. The remaining 16 participants (76 %) were suspicious, because in some trials the director would use a word that helped them choose between the target and the object in the gray cell. All of the suspicious participants were able to provide one to three examples of Critical trials.

Participants were significantly above chance at getting suspicious of the director's knowledge ($p < .027$, binomial test, two-tailed). Also, importantly, their responses to the posttest questionnaire confirmed that those followers who kept track of the contents of the gray cell were actually taking the director's perspective while doing so.

The mouse-click responses revealed perfectly accurate performance (as would be expected for 2×2 grids). The mouse-click response times were not accurate enough for reliable statistical analysis, for two reasons: First, followers' speeds of response were affected by the location of the mouse cursor at the time of hearing the name of the target object. Second, the instructions were formulated differently across and within directors in each trial (e.g., "Click on the brown dog" vs. "The puppy"), also affecting followers' speed of response across conditions. Note that this problem would not occur in studies using avatar directors or confederates, because the instructions would either be prerecorded or formulated in the same way for all participants.

The proportions of fixations on the gray cell made by the five unsuspicious followers are reported in Table 2, and the eye-tracking data for the suspicious followers are plotted in Fig. 1. The data reported in Table 2 was not sufficient to carry out a reliable statistical analysis, therefore ruling out a comparison with the suspicious participants. However, eyeballing

Table 2 Proportions of fixations on the gray cell (privileged-ground object) in the seven trials of the Baseline condition and the seven trials of the Critical condition (presented in that order) by those followers who were not suspicious of the director's perspective (N = 5)

Follower	Baseline Trials							Critical Trials							Mean
	1st	2nd	3rd	4th	5th	6th	7th	1st	2nd	3rd	4th	5th	6th	7th	
A	.34	.31	.42	.06	.07	.06	.19	.04	.00	.00	.11	.42	.23	.35	.19
B	.25	.09	.36	.08	.08	.53	.10	.15	.07	.00	.11	.06	.26	.18	.17
C	.00	.00	.06	.00	.05	.11	.06	.00	.03	.00	.03	.25	.05	.23	.06
D	.07	.17	.13	.12	.54	.05	.15	.00	.33	.27	.00	.06	.23	.05	.15
E	.18	.06	.24	.00	.07	.28	.00	.06	.00	.04	.10	.09	.00	.19	.09

Only Follower A noticed that the director sometimes used overspecific instructions in the second half of the task. Followers B–E reported having tried to block the gray cell from their view during the task

the looks at the gray cell made by the five unsuspicious followers does not seem to suggest that these data fit a U-shaped distribution. This may be because four of these participants reported having tried to focus their attention on the three white cells during the task. Although I acknowledge that the eye-tracking data for these participants are not strong enough to confirm this selective-attention strategy, I interpret their responses to the posttest questionnaire as evidence that they were adopting such a strategy.

The eye-tracking data of the suspicious followers were compared in the first three trials (Block 1) and the last three trials (Block 2) of both the Baseline condition (first half of the experiment) and the Critical condition (second half of the experiment). It follows from the predicted U-shaped distribution of the data that there should be a Trial Block × Condition interaction, with a higher proportion of fixations on the gray cell being expected in Block 1 in the Baseline condition and in Block 2 in the Critical condition.

Statistical analyses were conducted using the lmer function from the lme4 package (Bates, Mächler, Bolker, &

Walker, 2014) in the R statistical computing language (R Development Core Team, 2014). A linear mixed-effects model was implemented on the proportions of fixations ($N = 16$; 192 observations, not collapsed across condition), positing fixed effects of Trial Block and Condition, plus an interaction between the two. The maximal random-effects structure that converged included random effects of participant and item, plus random slopes of Trial Block and Condition over participants (Trial Block/Condition: model estimates = .1235/.1465; standard errors = .0460/.0465; t values = 2.689/3.153). Comparing models with and without the interaction term between Trial Block and Condition (holding random effects constant) revealed a significant interaction between the two factors [$\chi^2(1) = 11.97$, $p = .0005$].

Looking at the effect of Trial Block separately in each condition, the maximal random-effects structure that converged included random effects of participant and item, plus a random slope of Trial Block by participant (Baseline condition/Critical condition: model estimates = .1235/.1221; standard errors = .0403/.0540; t values = 3.064/2.262). Model comparisons revealed a significant effect of Trial Block in both the Baseline condition [$\chi^2(1) = 7.168$, $p = .0074$] and the Critical condition [$\chi^2(1) = 4.766$, $p = .0290$].

Looking at the effect of Condition separately in each Trial Block, the same maximal random-effects structure converged (Baseline condition/Critical condition: model estimates = .1465/.0992; standard errors = .0442/.0524; t values = 3.312/1.893). Model comparisons revealed a significant effect of Condition in Trial Block 1 [$\chi^2(1) = 8.022$, $p = .0046$], and a marginally significant effect in Trial Block 2 [$\chi^2(1) = 3.695$, $p = .0546$].

The eye-tracking analyses of the suspicious participants confirmed their responses to the questionnaire: By keeping track of the hidden object, these participants were able to keep track of the director's perspective. These results therefore challenge the assumption that looking at the hidden objects reveals limited use of Theory of Mind.
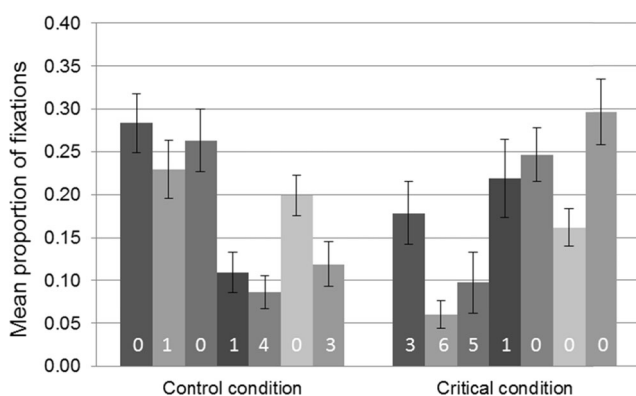
Fig. 1 Mean proportions of fixations on the gray cell in the seven trials of the Baseline condition and the seven trials of the Critical condition (by order of presentation, from left to right) by those followers who were suspicious of the director's perspective (N = 16). The number at the base of each bar indicates how many followers did *not* fixate on the gray cell in that particular trial. Those data points were retained in the statistical analysis

Discussion

The DT makes participants restrict the director's referential domain to those objects in her visual field, thus posing highly artificial demands on participants' pragmatic abilities. However, those researchers defending the view that interlocutors do not always use their Theory of Mind in communication have treated the DT as a representative test of referential communication, drawing conclusions that go well beyond the specific demands of this experimental paradigm and extend to everyday communication (e.g., Apperly & Butterfill, 2009; Apperly et al., 2010; Converse et al., 2008; Keysar et al., 2000; Keysar et al., 2003; Lin et al., 2010; Santiesteban et al., 2015). Moreover, most of these studies failed to account for the large number of studies that have reliably shown perspective taking in communication using the DT and other paradigms (e.g., Brown-Schmidt, 2009, 2012; Hanna & Tanenhaus, 2004; Heller, Gorman, & Tanenhaus, 2012; Heller et al., 2008; Kuhlen & Brennan, 2010; Lockridge & Brennan, 2002; Nadig & Sedivy, 2002).

The results of this study show that optimal performance in the DT, as conceived by the defendants of the egocentric view (Apperly et al., 2010; Keysar et al., 2003), may only reveal selective attention, therefore rendering the standard task unreliable as a test of Theory-of-Mind use in communication: The few participants who tried to block the gray cell and avoided interference from their privileged perspective actually underused their Theory of Mind. In this respect, the present results are consistent with Santiesteban et al.'s (2015) claim that participants may use nonmentalizing strategies in performing the DT.

However, the new version of the DT also revealed that adults actively represent their interlocutor's mental states in referential communication, challenging Santiesteban et al.'s (2015) conclusion that adults only use mentalizing sparingly in both psychology experiments and everyday life. In fact, the large majority of participants were able to take the director's perspective while keeping track of the contents of the gray cell. These results are also more reliable than those of Santiesteban et al., who used a static human figure and prerecorded instructions instead of a live director. Therefore, future experimental pragmatics studies should adopt this new eye-tracking paradigm and investigate not only how listeners *take* the speaker's perspective in referential communication, but also how they *update* this perspective during interaction.

Although this paradigm is more naturalistic than those in previous studies, the new interactive DT still uses artificial rules in a verbal coordination game, leaving open the question of how often people mentalize in everyday conversation. A simple example (adapted from Geurts & Rubio-Fernández, 2015) suggests that this is a complex question that requires investigation: Imagine you board a train and sit next to an old lady who breaks the ice by saying: "It's windy today....Just like the day you were born." Your automatic response would probably be to wonder: How does this woman know when I was born? Since the old lady did not use mental state verbs or express a subjective opinion that could require "explicit mentalizing" (Heyes, 2014), we must assume that this automatic inference was triggered by the old lady's making of a statement, which *presupposes* her knowledge.

Future research should investigate how and when automatic pragmatic inferences are derived, because this could provide a critical test of mentalizing—or submentalizing—in everyday communication.

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116,* 953–970. doi:10.1037/a0016923

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology, 63,* 1201–1217. doi:10.1080/17470210903281582

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Science, 14,* 110–118.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109,* 18–40. doi:10.1016/j.cognition.2008.07.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.0-6). Retrieved from http://CRAN.R-project.org/package=lme4

Begeer, S., Malle, B. F., Nieuwland, M. S., & Keysar, B. (2010). Using Theory of Mind to represent and take part in social interactions: Comparing individuals with high-functioning autism and typically developing controls. *The European Journal of Developmental Psychology, 7,* 104–122.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16,* 893–900. doi:10.3758/PBR.16.5.893

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language & Cognitive Processes, 1,* 62–89.

Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: Mood affects Theory-of-Mind use. *Emotion, 8,* 725–730. doi:10.1037/a0013283

Development Core Team, R. (2014). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from www.R-project.org

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science, 13,* 331–338. doi:10.1111/j.1467-7687.2009.00888.x

Dumontheil, I., Küster, O., Apperly, I. A., & Blakemore, S. J. (2010). Taking perspective into account in a communicative task. *NeuroImage, 52,* 1574–1583. doi:10.1016/j.neuroimage.2010.05.056

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

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. doi:10.1016/j.cogsci.2003.10.002

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. doi:10.1111/j.1756-8765.2012.01182.x

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

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science, 9,* 131–143. doi:10.1177/1745691613518076

Keysar, B. (1997). Unconfounding common ground. *Discourse Processes, 24,* 253–270.

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,* 32–38. doi:10.1111/1467-9280.00211

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

Krauss, R. M., & Glucksberg, S. (1977). Social and nonsocial speech. *Scientific American, 236,* 100–105.

Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes, 47,* 567–587. doi:10.1080/01638530903441339

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology, 46,* 551–556.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review, 9,* 550–557. doi:10.3758/BF03196312

Nadig, A., & Sedivy, J. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13,* 329–336.

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology, 7,* 153.

Santiesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic Bulletin & Review, 22,* 844–849. doi:10.3758/s13423-014-0716-0

Sedivy, J. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research, 32,* 3–23.

Symeonidou, I., Dumontheil, I., Chow, W. Y., & Breheny, R. (2016). Development of online use of theory of mind during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology, 149,* 81–97. doi:10.1016/j.jecp.2015.11.007

Wang, J. J., Ali, M., Frisson, S., & Apperly, I. A. (2016). Language complexity modulates 8- and 10-year-olds' success at using their theory of mind abilities in a communication task. *Journal of Experimental Child Psychology, 149,* 62–71. doi:10.1016/j.jecp.2015.09.006