Journal of Experimental Psychology:
Learning, Memory, and Cognition
2017, Vol. 43, No. 4, 591–610

0278-7393/17/$12.00 http://dx.doi.org/10.1037/xlm0000345

Using Perspective to Resolve Reference: The Impact of Cognitive Load and Motivation

James E. Cane
London South Bank University

Heather J. Ferguson
University of Kent

Ian A. Apperly
University of Birmingham

Research has demonstrated a link between perspective taking and working memory. Here we used eye tracking to examine the time course with which working memory load (WML) influences perspective-taking ability in a referential communication task and how motivation to take another's perspective modulates these effects. In Experiment 1, where there was no reward or time pressure, listeners only showed evidence of incorporating perspective knowledge during integration of the target object but did not anticipate reference to this common ground object during the pretarget-noun period. WML did not affect this perspective use. In Experiment 2, where a reward for speed and accuracy was applied, listeners used perspective cues to disambiguate the target object from the competitor object from the earliest moments of processing (i.e., during the pretarget-noun period), but only under low load. Under high load, responses were comparable with the control condition, where both objects were in common ground. Furthermore, attempts to initiate perspective-relevant responses under high load led to impaired recall on the concurrent WML task, indicating that perspective-relevant responses were drawing on limited cognitive resources. These results show that when there is ambiguity, perspective cues guide rapid referential interpretation when there is sufficient motivation and sufficient cognitive resources.

Keywords: perspective taking, working memory, eye movements, growth curve analysis, motivation

The ability to understand another person's perspective is a cornerstone of many social interactions, and often involves overriding what is known from one's own perspective to consider another's different view. This perspective taking ability is couched within the broader concept of theory of mind (ToM), helping to infer others' mental states (e.g., beliefs, knowledge, attitudes) or approximate another's visual perspective. Perspective taking ability also helps to reduce ambiguity in social interactions and conversations (Hanna, Tanenhaus, & Trueswell, 2003). Although it is typically assumed that perspective taking abilities are fully developed by adulthood (Keysar, Lin, & Barr, 2003; Rubio-Fernández & Geurts, 2013; Wellman, Cross, & Watson, 2001), emerging research has identified a number of factors that give rise to individual differences in this ability; these include mood (Converse, Lin, Keysar, & Epley, 2008), social and cultural relationships (Savitsky, Keysar, Epley, Carter, & Swanson, 2011; Wu & Keysar, 2007), inhibitory control (Brown-Schmidt, 2009b), and more recently working memory load (WML) and working memory capacity (WMC; Lin, Keysar, & Epley, 2010). Moreover, even in healthy adults there are limits on the extent to which perspective taking is used in a given situation, with factors such as communication goals, motivation, and available time likely to modulate perspective use (see Epley, Keysar, Van Boven, & Gilovich, 2004; Ferguson, Apperly, Ahmad, Bindemann, & Cane, 2015; Yoon, Koh, & Brown-Schmidt, 2012). Furthermore, recent research has explored the extent to which perspective taking abilities are deployed over time, examining whether perspective taking abilities are deployed in anticipation of a perspective-relevant response (i.e., at trial onset, prior to language onset), or whether they are deployed only when ambiguities in a discourse need to be resolved by incorporating incoming visual and auditory information, commonly called *integration* (see Barr, 2008b). The present research seeks to build on this previous research in two eye-tracking studies that explore how concurrent cognitive load impacts the deployment of perspective taking abilities over time. In particular, we examine evidence for anticipatory (defined here as the period just prior to disambiguation of the target noun) and integration effects and the impact of cognitive load and WMC on these effects.

One contentious view of perspective taking proposes that communicators are initially egocentric, with incoming information primarily being interpreted according to one's own perspective (Keysar et al., 2003). Consideration of other peoples' perspectives might then be activated at some later point to reduce ambiguity, but this inference is not reliably deployed in social situations.

James E. Cane, Division of Psychology, School of Applied Science, London South Bank University; Heather J. Ferguson, School of Psychology, Keynes College, University of Kent; Ian A. Apperly, School of Psychology, University of Birmingham.

This work was carried out with the support of a grant to Heather J. Ferguson and Ian A. Apperly from the Leverhulme Trust (Ref: RPG-063).

Correspondence concerning this article should be addressed to James E. Cane, Division of Psychology, School of Applied Science, London South Bank University, 103 Borough Road, London SE1 0AA, United Kingdom. E-mail: j.e.cane@lsbu.ac.uk

Much of the evidence for this account comes from research that has used a referential communication task (Keysar, Barr, Balin, & Brauner, 2000). In this task, participants follow the instructions of a confederate 'director' to select ("click on the . . .") or move ("Move/pick up the . . .") target objects (e.g., a ball) in a visual display. The visual display typically consists of a 4 × 4 gridded cupboard; some of the objects in the grid are visible to both the director and the participant, but others are occluded from the director's (but not the participant's) view. On critical trials, the grid contains a range of objects including a target object (e.g., a toy mouse) and a competitor object that is either referentially ambiguous (e.g., a computer mouse) or nonambiguous (e.g., a shoe). To examine perspective taking ability, the competitor object is placed in privileged ground, where it is occluded from the directors' view by a physical barrier. Therefore, to correctly identify the target object from an ambiguous instruction to "move the mouse left," participants must infer the director's limited perspective and restrict attention to the block in 'common ground' (i.e., shared view) while inhibiting access to the competitor in 'privileged ground'. Researchers using this paradigm have reported slower responses to select the target object when the privileged competitor was referentially ambiguous compared to when it was not (Barr & Keysar, 2002; Epley, Morewedge, & Keysar, 2004; Keysar, Barr, Balin, & Brauner, 2000; Keysar et al., 2003). Furthermore, these studies show a surprisingly high number of trials where participants select the perspective-inappropriate object (ranging from 15% to 46%). These findings are often taken as evidence that people are biased to their own egocentric perspective, over the other person's (altercentric) perspective, and only adjust to the other person's perspective at some later point according to need (Epley, Keysar, et al., 2004; Horton & Keysar, 1996; Keysar, 2007). In the case of an incorrect selection this is seen as a clear indication of the failure to take another's perspective into account (Keysar et al., 2003).

However, the 'egocentric-first' proposition has been challenged by findings from a number of studies that have shown early use of perspective, including some showing that people automatically infer other peoples' perspectives, even when doing so is not necessary for the task (e.g., Ferguson et al., 2015; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Schneider, Nott, & Dux, 2014). For instance, Hanna et al. (2003) used a version of the referential communication task involving different colored shapes in a 3 × 3 grid. In this task an object (e.g., a red triangle) was placed in the grid alongside a similar competitor object (e.g., another red triangle) to which the director either referred, so as to indicate it was in common ground, or did not refer, indicating that the addressee had privileged knowledge of the competitor. An instruction was then given to move another object (e.g., a blue triangle) to the target location (e.g., Now put the blue triangle on the red one). They found that participants were always less likely to fixate on the competitor object when it was in privileged ground compared to when it was in common ground. These findings are compatible with findings from other studies that have used linguistic comparators, such as scalar adjectives (e.g., small/big; Heller, Grodner, & Tanenhaus, 2008) and definite or indefinite expressions (e.g., the/one of the; Hanna et al., 2003, Experiment 2) to show immediate effects of perspective when two equally fitting referents are available in common ground but strong constraints in the discourse/visual display narrow down the intended referent. Nevertheless, though there was a general tendency for participants to move attention away from objects in privileged ground in these studies, some attention was still given to these objects, thus indicating that there was not complete inhibition of their egocentric perspective (see Heller, Parisien, & Stevenson (2016) for discussion of this effect). Other studies have examined processing during temporary ambiguities in reference and have shown even earlier use of perspective, without an egocentric bias, when participants are engaged in an interactive dialogue with their partner that explicitly establishes what each speaker does and does not know (e.g., "What's above the cow with a hat?"; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Brown-Schmidt, 2012). Similarly, tasks in which the participant is a passive observer to narrated events in a discourse have shown rapid and accurate prediction of other peoples' actions based on an understanding of their (false) beliefs (e.g., Ferguson & Breheny, 2012; Ferguson, Scheepers, & Sanford, 2010; Rubio-Fernández, 2013), or conflicting desires (Ferguson & Breheny, 2011). These studies, showing early use of perspective, suggest that interpretation of language is driven by multiple probabilistic constraints, one of which is perspective (Brown-Schmidt & Hanna, 2011). Thus, they indicate that in at least some circumstances listeners are able to infer other peoples' perspectives in advance or early on during comprehension, and use this knowledge to distinguish between objects that are perspective-appropriate and objects that are not.

Although there is clear evidence that perspective taking may be undertaken spontaneously during discourse and integrated quite rapidly into online processing, there are also good reasons for thinking that some or all of this requires cognitive effort. For example, studies using dual-task WML manipulations have shown that higher load impedes one's ability to infer another's mental states (Bull, Phillips, & Conway, 2008; McKinnon & Moscovitch, 2007; Schneider, Lam, Bayliss, & Dux, 2012; cf. Qureshi, Apperly, & Samson, 2010). This has been further corroborated by recent eye movement studies showing that individual differences in executive function predict perspective taking ability, with both increased WMC and greater inhibitory control leading to a decrease in the likelihood of suffering interference from one's own perspective (Brown-Schmidt, 2009b; Grodner, Dalini, Pearlstein-Levy, & Ward, 2012; Lin et al., 2010; Nilsen & Graham, 2009). Of particular relevance for the current study is work by Lin et al. (2010), who examined performance on a referential communication task between individuals with high and low WMC (as measured by the operation span [OSPAN] task; La Pointe & Engle, 1990; Turner & Engle, 1989) and between high and low levels of WML in a secondary task (i.e., memorize four or two numbers). Results revealed that participants with higher WMC fixated to a lesser degree on the privileged objects compared with those with low WMC, suggesting that availability of more cognitive resources allowed participants with high WMC to inhibit the perspective-irrelevant object. Similarly, when participants were under low WML, their target preference was greater than when they were under high WML. This finding suggests that when cognitive resources are allocated elsewhere (i.e., another cognitively demanding task), one's ability to use perspective to accurately guide behavior is impaired.

What is not clear from these studies is at what point executive function impairment exerts its effect on the perspective taking process. Furthermore, recent research has raised questions regarding the robustness of the relationship between executive functions and perspective taking (see Brown-Schmidt, 2012; Ryskin et al., 2014). Studies examining the role of executive function using the referential communication task commonly measure the prevalence of egocentric or altercentric biases using either explicit selection responses or aggregated fixations to the referentially ambiguous competitor. It could be argued that selection measures occur near the end-point of a decision process, and these alongside aggregated fixations therefore do not inform us on the full time-course of perspective taking processes across the decision period. Indeed, a recent time-course examination of processing during the referential communication task has distinguished between anticipatory visual biases, which show a preference for objects in common ground even before verbal instructions to move objects begin, and integration processes, which are susceptible to interference from lexical competitors regardless of whether they are in common or privileged ground (Barr, 2008b; Brennan & Hanna, 2009; Brown-Schmidt & Hanna, 2011). Furthermore, though some studies have shown that reduced executive function can lead to disruption in the ability to take another person's perspective, the majority of participants are still able to choose the perspective-appropriate object. As an example, in the study by Lin et al. (2010) selection errors only occurred on 38% of trials in participants with low WMC (Experiment 1) and on 47% of trials where high WML was applied (Experiment 2). Given that the majority of trials did not end in selection errors indicates that impairment of executive function does not always lead to a complete failure of perspective-taking ability, but rather it impacts one (or more) of the stages in the decision making process. With reference to the egocentric first accounts, the impairment of executive function may interrupt the adjustment phase, thus elongating the time taken to approximate the other person's perspective and make the correct selection, but not influence participants' initial preferential attention to objects that are known to the speaker. The current paper seeks to elucidate these limitations by examining the exact time course of shifts in fixations to the target object and referentially ambiguous competitor objects over the decision-making window by modeling the changes in bias over time. In this way, we aim to identify the specific stages of perspective taking processing that are influenced by executive function impairment (including a comparison of anticipatory vs. integration effects).

The Current Research

On the basis of the issues discussed in the preceding text, the reported experiments examine the impact of cognitive load on the deployment of perspective use over time. Using highly sensitive eye-tracking measures, we examined the temporal deployment of perspective taking and the influence of WML on this ability. Both experiments employed a referential communication task, where participants followed the instructions of an on-screen avatar director to move target objects (e.g., a "teapot with spots on") around a 4 × 4 grid in the presence of temporarily ambiguous competitor objects (e.g., a "teapot with stars on"). We compared performance in three conditions that engage different perspective choices: a listener-privileged condition, where a competitor object was only available to the participant (i.e., it was occluded from the speaker's viewpoint); a common ground condition, where target and competitor objects were available to both participant and speaker; and a no-competitor condition, where no competitor object was available in the grid and one grid square was occluded from the participant's view. The listener-privileged and common ground conditions are comparable to those seen in previous studies (e.g., Hanna et al., 2003). As well as providing a baseline of processing when no competitor is visually available, the no-competitor condition offers a new look at ambiguity in reference assignment by testing whether participants infer the presence of an object behind the occluded grid space, based on the speaker's overinformative contrastive description of the target object (see Engelhardt, Bailey, & Ferreira, 2006; Sedivy, Tanenhaus, Chambers, & Carlson, 1999). WML was manipulated within each condition using a dual-task design (see de Fockert, Rees, Frith, & Lavie, 2001) that required participants to hold a sequence of digits in memory during the referential communication task. Low WML was achieved by presenting these digits in the correct sequential order (0 1 2 3 4), whereas high WML presented the digits in a nonsequential order (e.g., 0 2 1 3 4).

Importantly, instructions to move objects around the grid were carefully constructed so that they allowed sufficient time for participants to build up and maintain expectations about the speaker's perspective. Thus, we used target and competitor objects that were different only according to one salient visual property (e.g., patterns, accessories; Breheny, Ferguson, & Katsos, 2013; Brown-Schmidt, 2009a; Brown-Schmidt et al., 2008; 2012) and temporarily ambiguous verbal object descriptors that described this disambiguating property after the noun (e.g., "move the teapot with the spots on . . ."). In this way, we hoped to reduce the influence of bottom-up lexical integration and allow participants more time to deploy perspective during the anticipation phase. The second experiment reported here pushes this perspective use further by providing explicit motivation (in the form of monetary rewards for fast and accurate responses) for participants to infer the speaker's perspective even before the disambiguating information is available. Previous research has shown that perspective information is not incorporated when producing spoken descriptions of an object under time pressure (Horton & Keysar, 1996; cf. Grodner & Adler, 2013) and speakers are more likely to reveal privileged information when explicitly instructed and rewarded for not doing so (Wardlow Lane & Liersch, 2012), however the impact of external motivation on interpreting perspective-relevant verbal descriptions is not yet known. Furthermore, this manipulation allowed us to further investigate how and when perspective taking is undertaken according to need, and how this might be influenced by cognitive load.

Throughout both experiments we exploit recent advances in fine-grained time-course analysis for eye-movement data using mixed-model and Growth Curve analysis (see Barr, 2008a; Mirman, Dixon, & Magnuson, 2008). This approach allows us to identify specific time-points where perspective and cognitive load may have an impact on fixations to objects within the visual scene and allows us to control for by-participant and by-item variation and test the influence of individual differences in working memory and inhibitory control.

In line with previous research, we predicted that participants would succeed in taking the director's perspective, so that they experience greater interference from a competitor object when it was in shared view, compared with when it was in their privileged view. This effect could result in delayed response times on trials with a common ground compared with those with a listener-privileged competitor as well as weaker and later visual biases to the target object in the common ground condition as participants consider both objects as potential referents. With regard to the WML manipulation, it was predicted that if perspective taking is cognitively effortful and requires special cognitive processing, then we should see detrimental effects in peoples' ability to use perspective to guide processing when those cognitive resources are being used by a cognitively effortful secondary task. Such effects might emerge as delayed target selection on listener-privileged trials under high compared with low WML as well as reduced and delayed visual biases to the target under high WML. It is unknown whether such load effects would emerge during the anticipation or integration stages of processing.

Experiment 1

Method

Participants. Native-English speaking University of Kent students ($N = 39$) were recruited through a university-wide research participation scheme. One participant was excluded from the final analysis because problems with eye-tracking recording, and a further two were excluded as they did not follow the task instructions. Of the remaining 36 participants, 32 were female and 4 were male. The mean age of participants was 19.64 years ($SD = 5.42$).

Stimuli and design.

Referential communication task. We used an avatar version of the referential communication task based on similar tasks used in previous research (see Dumontheil, Apperly, & Blakemore, 2010; Dumontheil, Küster, Apperly, & Blakemore, 2010). The use of an avatar version of the referential communication task allowed us to carry out fine-grained temporal and spatial eye-tracking measures, avoid any within-trial variation in speech that may occur with a confederate, and avoid any nonverbal behavior that may be elicited from a confederate. Studies which have directly compared an avatar with an object which is unlikely to hold a specific self-perspective (e.g., a rectangle) have shown perspective taking effects for the avatar but not for the object (see Samson et al., 2010). These effects are further corroborated by evidence from fMRI that has shown activations in "social" areas of the brain, including the temporoparietal junction and medial prefrontal cortex, in virtual versions of the referential communication task when avatar directors are present compared with when a control object is used (Dumontheil, Küster, Apperly, & Blakemore, 2010; Schurz et al., 2015).

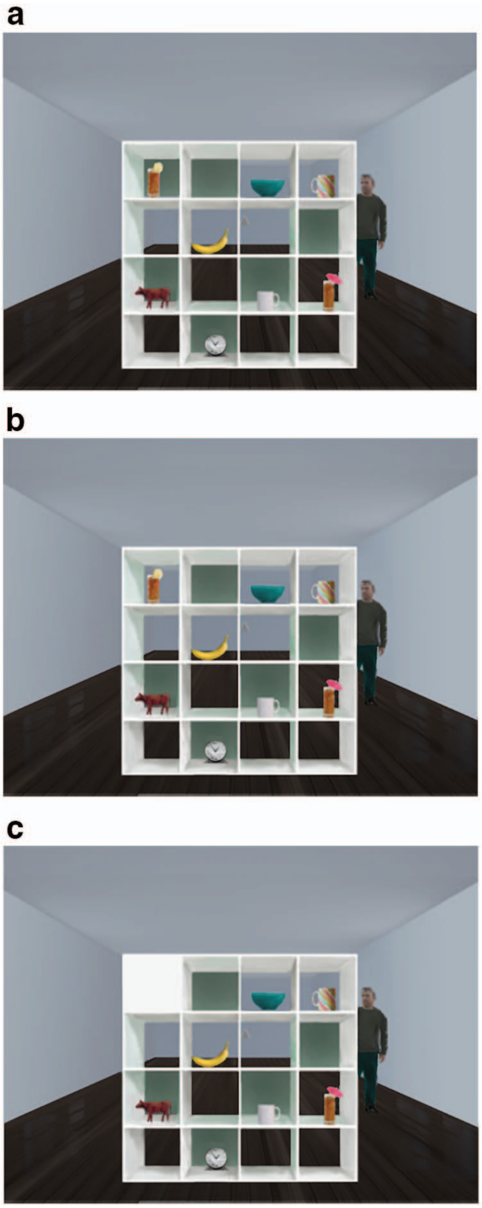

Figure 1. Example stimuli from the referential communication task for each of the competitor conditions. (a) listener-privileged condition; (b) common ground condition; (c) no-competitor condition. See the online article for the color version of this figure.

Each trial of the referential communication task consisted of an image of a room containing a 4 × 4 gridded cupboard with a male avatar standing to the rear right-hand side of the cupboard (see Figure 1). For the listener-privileged and common ground conditions the backs of five spaces within the cupboard were covered so that the contents of these spaces were occluded from the avatar's view. For the no-competitor trials, the backs of four of the spaces were occluded from the avatar's perspective and the front of one space was covered so that the contents of that space was occluded from the participants' perspective. These no-competitor trials ensured that participants would not make assumptions about the consistent appearance of a pair of objects. In total, 39 cupboard configurations were created, three for the practice trials and 36 for the experimental trials (12 listener-privileged, 12 common ground, 12 no-competitor). Twelve sets of seven to eight objects were placed into the cupboard spaces for each competitor condition. For the common ground condition each set included a target object (e.g., a glass with an umbrella in) and a competitor object (e.g., a glass with a lemon in), which were both in visual common ground. For the listener-privileged condition the competitor object was placed in one of the speaker occluded spaces, and for the no-competitor trials no competitor object was shown. Each target/competitor pair (e.g., two glasses) was used once for each of the three trial types, and the specific feature of the target/competitor object (i.e., the straw, the lemon, or the cocktail umbrella in the glass) was changed across the conditions so that the target object could not be predicted on each presentation. We counterbalanced the occurrence of this target feature change across conditions by producing three different versions of the experimental program presented to different participants to ensure that the nature of the target did not influence responses to a particular condition.

During their introduction to the task, participants were shown example grids from their own and the avatar's perspective (see Figure 2) and were made aware that they had to move objects one space in the grid according to the instructions given by the avatar, and this movement would be either up, down, left, or right but never diagonally. Given that selecting the correct target object in the referential communication task taps Level 1 perspective taking ability (understanding what another person can see) rather than Level 2 perspective taking ability (understanding how another person sees something), we do not expect participants to engage mental rotation while identifying and selecting the target object. Thus, participants were asked to move objects according to their own left–right perspective. Participants moved the objects by clicking on the object with the computer mouse and dragging the object to the new location. Participants received three instructions per trial, comprising two filler instructions and one critical instruction. For critical instructions, the instructions consisted of "Move the . . ." + the target object noun (e.g., ball, shoe, truck) + disambiguating information (e.g., "with a straw in") + a direction (up, down, left, or right). The filler instructions comprised two types of instruction, one containing a noncomparative adjective (e.g., "Move the yellow bucket up") and one not containing an adjective (e.g., "Move the bottle down"). The use of color adjectives to describe some filler objects was based on previous research that has shown that color terms are not typically used or interpreted contrastively (e.g., Davies & Katsos, 2010; Mangold & Pobel, 1988; Sedivy, 2002). The order of filler and critical instructions was counterbalanced across trials and a new instruction was only given once participants had responded to the previous instruction. So that participants would not set up specific expectations that all pairs of objects would be referred to, on some trials (N = 16) included pairs of objects (e.g., two mugs) in addition to the critical target/competitor objects. These items were only referred to occasionally in the filler instructions.

Inhibitory control: Stroop task. The Stroop task (Stroop, 1935) consisted of 40 incongruent trials comprising color words in one of four different ink colors (red, blue, yellow, and green; e.g., the word *red* in the ink color green), 40 congruent trials (e.g., the word *red* in the ink color red), and 40 noncolor word-neutral trials (all animal-related words, e.g., the word *horse* in the ink color red, matched for word length with the color words used). The task was run through E-prime 2.0 software (Psychology Software Tools, Inc. 2012), and responses were recorded using a five-button serial response box with the four extreme buttons (two on the left, two on the right) being used for the ink-color responses (red, green, blue, yellow). Interference scores for the Stroop task were calculated by subtracting response times (RTs) from the neutral trials from incongruent RTs and this was used as the measure of inhibitory control. For the analyses, interference scores were reversed so that greater interference scores could be interpreted as less inhibitory control, and low or no interference is interpreted as greater inhibitory control. Prior to experimental trials, participants received 12 practice trials consisting of country names in the four different colors. Reliability analyses carried out for the Stroop task in Experiment 1 revealed high reliability within all conditions (incongruent: $\alpha = .90$; congruent: $\alpha = .89$; neutral: $\alpha = .91$).[1]

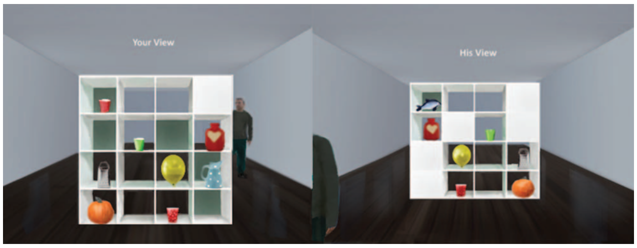

Figure 2. Example screen for the referential communication task presented to participants, showing the cupboard from both participant's (left) and the avatar's view (right). See the online article for the color version of this figure.

WMC: OSPAN task. WMC was measured using the OSPAN task (La Pointe & Engle, 1990; Turner & Engle, 1989). In this task, participants responded to a mathematical equation (e.g., $(4/2)+3 = 5$), stating whether the answer shown was true or false, then read out loud a word that was presented on the screen. After a series of equations/word pairs participants were required to type in the words they had read out in the order they had seen them. This version of the OSPAN task consisted of 12 trials in total, which included 2, 3, 4, or 5 equation/word pairs. The task was run through E-prime 2.0 software, and responses were recorded using the keyboard. Participants pressed Y to indicate a correct equation and N for incorrect equations, and pressed the space bar to proceed after reading aloud the word that followed each equation. WMC scores were calculated by summing the number of words in correctly recalled word sequences; sequences only contributed to the WMC score total where all words were recalled correctly in the right order. WMC scores could range from 0 to 42, with higher scores indicating higher WMC. Reliability analyses revealed high internal reliability for the OSPAN task ($\alpha = .94$).

Apparatus. Eye movements were recorded using an EyeLink 1000 desktop mounted SR Research eye-tracker and were sampled at a frequency of 1,000 Hz. Only the left eye was tracked, and the participant's head was kept immobile with the use of a chin and forehead rest throughout the experiment. A 19-in. TFT monitor screen with a screen resolution of 1,024 × 768 pixels was used to present stimuli at a distance of 60 cm from the participant. Participants' eye movements were calibrated through a nine-point calibration process, which covered all the main central and peripheral aspects of the screen and a drift correction check (central fixation point on the screen) was included prior to each trial. The referential communication task was delivered and controlled using the Eye-Link Experiment Builder Software (Version 1.10.165; SR Research Ltd., 2013). Each box on the 4 × 4 Keysar grid covered an average visual angle of 4.25° on the horizontal plane and an average visual angle of 5.35° on the vertical plane, dependent on

---

[1] Reliability analyses were conducted using the Cronbach function in the R *psy* package (Falissard, 2012). For the Stroop, these analyses used RTs within each condition and accuracy scores were used for the OSPAN task.

the location of the box.[2] The instructions for each trial were delivered to participants through headphones covering both ears; all recordings were delivered in mono sound.

Procedure. Participants sat in front of the monitor and were given instructions on how to complete the referential communication task, using the mouse to move objects one space, left, right, up, or down, according to the instructions given by the avatar. Participants were instructed to take the avatars perspective into account throughout; they were shown a single example of the grid from their perspective and from the avatars perspective to ensure they understood that their perspectives differed. Once the participants had indicated they fully understood the instructions, the participants' eye-movements were calibrated and the headphones for the instructions were placed over the participants' ears. Participants then received three practice trials, one replicating a trial from each of the conditions, before moving onto the main set of 36 experimental trials that were randomly presented for each participant. Each trial began with a 2 s presentation of the sequence of five digits to be remembered, setting up high or low WML conditions, as described earlier. Participants then saw a grid scene for the referential communication task, and responded to three instructions to move objects around this grid. The visual locations of objects in the grid were updated in real time as participants moved them. Each trial ended with a single digit from the initial sequence, and participants were required to recall the next digit that had appeared in the sequence. Halfway through the experimental trials participants were able to take a short break. Once they were ready to continue, eye-movements were recalibrated to ensure accuracy and the remaining trials were delivered. When participants had completed all 39 trials, they completed the Stroop task followed by the OSPAN task. Participants were fully debriefed following all of the tasks.

Analyses.

Accuracy and response time analysis. Accuracy scores for the WML task and accuracy scores and RTs for the referential communication task were analyzed using a series of mixed-effect regression models (see Barr, 2008a). These models allowed us to control for both between participant variation and between item variation (Baayen, Davidson, & Bates, 2008). All models were fitted using the lmer function in the lme4 package (Bates, Maechler, & Bolker, 2012) for the R software interface (R Development Core Team, 2013). Digit recollection accuracy was entered as the dependent variable for the WML task model, and target object selection accuracy and RTs were entered as the dependent variables for the referential communication task models. Object selection RT in the referential communication task was time-locked to the onset of the target noun (e.g., *glass*). For all trials, correct recall of the working memory task number and correct selection of the target object was scored as 1 and incorrect selections as 0, and accuracy scores for each participant were calculated as the mean of these binary scores within each of the competitor and WML conditions. Raw selection RTs were initially assessed for outliers—RTs with standard z scores outside the limits of ±2.5 were treated as outliers and excluded from subsequent analyses. Also RTs from trials where the incorrect object was selected were omitted from the final analyses. Therefore, the RTs used in the analyses relate to correct target selections only.

Each model included fixed effects of competitor type and WML. To accommodate the three levels of competitor type within the mixed-effect model analyses, two deviation coded contrast schemes were applied to the competitor variable: Contrast 1 = (listener-privileged (1/3), common ground (1/3), no-competitor (−2/3); Contrast 2 = listener-privileged (1/2), common ground (1/2), no-competitor (0). The contrast coding of the competitor variable thus allowed us to directly compare the combined listener-privileged condition and common ground competitor condition with the no-competitor condition (Contrast 1) and compare the listener-privileged condition with the common ground competitor condition without the no-competitor condition (Contrast 2). A significant effect of competitor would indicate a significant effect in at least one of these competitor contrasts—the specific coding (and in some cases further post hoc tests) allows us to identify between which competitor conditions these effects lie. The two WML conditions were deviation coded (low [−.5] vs. high [.5]) to ensure high and low working memory conditions could be directly compared. In addition, WMC (OSPAN) scores and Stroop interference scores were entered into each model as fixed effects variables to assess their significance in influencing competitor related responses. These scores were entered as continuous variables, centered around the mean. Significance values for the fixed effects of competitor, WML, WMC, and inhibitory control were calculated using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2015). Significant effects of WMC and inhibitory control are reported where they interact with the competitor variable. Where post hoc analyses were required, models were releveled to incorporate the condition of interest as the reference level (i.e., the intercept). For all tests a significance level of 5% was used and estimates reported are based on least square means.

There is ongoing debate as to how random effects should be included in mixed-effect models (see Barr, Levy, Scheepers, & Tily, 2013; Bates, Kliegl, Vasishth, & Baayen, 2015). In our models, we retained the maximal random effects structure, including random effects for participants and items, and crossed random effects for competitor by participant, WML by participant, competitor by item, and WML by item. Random effects were only removed where they lead to nonconvergence because of overparameterization.

Fixation data processing and time-course analyses. Participants' eye movements around the scene in the referential communication task were tracked and time locked to the target-noun onset in the concurrent auditory instructions from the avatar (e.g., the *g* of *glass* in "move the glass with the umbrella in down"). Regions of interest (ROIs) were specified around all of the objects within the 4 × 4 cupboard. If a participant moved a particular object with the mouse during the task, then the location of the related ROI was also updated. Fixations to all ROIs were recorded from 1 s prior to the target-noun onset to 3 s after the target-noun onset. To analyze the time course of visual biases, this 4-s window was broken down into 20 ms bins and fixations were coded as 1 belonging to a ROI within each bin and 0 if there was no fixations in a particular ROI for a particular bin. These fixations were aggregated across participants and items to calculate a target preference score for each condition (similar to that used in previous research; see Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Brown-Schmidt,

---

[2] This was based on the participant's eyes being 4 cm above the center of the screen and 60 cm from the front of the screen.

Gunlogson, & Tanenhaus, 2008; Ferguson, Scheepers, & Sanford, 2010; Ferguson & Breheny, 2011, 2012; Heller, Grodner, & Tanenhaus, 2008). Specifically, in the listener-privileged and common ground conditions fixations to the target object were compared to fixations to the competitor object. In the no-competitor condition, because no competitor object was available, fixations to the target object were compared with one of the four to five unmentioned (i.e., not mentioned in the filler instructions) distractor objects in the grid. A different (in terms of grid location and identity) unmentioned distractor was selected for each trial.

This calculation produces a single value that takes into account the proportion of fixations on both the target and competitor/distractor object, and measures the bias toward each critical object in each condition across the time course. Scores above zero indicate a greater bias to fixate the target object and scores below zero indicate a greater bias to fixate the competitor/distractor object. The target-preference score was calculated as in Ferguson and Breheny (2011, 2012): log(Target/Competitor) $= ln(P_{(\text{Target})}/P_{(\text{Competitor})})$, where $P_{(\text{Target})}$ is the sum of fixations to the target object divided by the total fixations to all ROIs on that trial, and $P_{(\text{Competitor})}$ is the sum of fixations to the competitor object divided by the total fixations to all ROIs on that trial. As in the behavioral response data, analyses compared effects in the listener-privileged, common ground, and no-competitor conditions.

Appropriate time-windows for analysis were assessed by plotting the grand mean for the log-transformed target bias score, collapsed across all competitor and WML conditions from 1s prior to noun onset until 2 s after noun onset. This grand-mean time window selection procedure has been highlighted as a semiprincipled way to select the key time-window(s), allowing for the main patterns in fixation data to be captured without being influenced by condition differences (Barr, 2008a). Where it was deemed necessary to capture the rise and fall in fixations to target versus competitor in the time windows over time, we used growth curve analysis using the lmer function in the lme4 R package (Bates et al., 2012; see Mirman et al., 2008). Visual examination of the grand-mean plot for Experiment 1 identified two separate time periods: from $-1{,}000$ ms to target-noun onset (preonset analysis time window) and from target-noun onset (0 s) to 2,000 ms (postonset analysis time window).[3] The time course of target fixations in the postonset period was modeled by averaging over 100 ms time bins, using second-order orthogonal power polynomials (see Mirman et al., 2008) incorporating intercept, linear, and quadratic components. The intercept effects represent condition effects irrespective of the time-course of fixations, the linear component represents the unidirectional increase or decrease in fixations over time, and the quadratic component indicates the rate of change (acceleration or deceleration) in fixation shifts over time (see Mirman et al., 2008). Given the lack of change in fixation bias over time within the preonset period, mixed model analyses were conducted excluding any effects over time prior to the target-noun onset (i.e., intercept only, no polynomials).

In relation to the time course of effects within the conditions, evidence that participants were, in any way, sensitive to perspective cues in the listener-privileged condition would come from a significantly greater target bias in the intercept in the listener-privileged condition compared with the common ground condition (where both target and competitor are mutually available). We were interested in whether such effects might be apparent at the first opportunity participants had to show sensitivity to the speaker's perspective (i.e., before target-noun onset) or whether they only occurred once this information became potentially relevant to interpret the speaker's message (i.e., after target-noun onset). Posttarget-noun onset, we would expect to observe increasing bias toward the target described by the noun in all conditions, as the unfolding sentence gradually enabled participants to identify a single referent. Critically, if participants in the listener-privileged condition were using the speaker's perspective to assist with identification of the target, then the increasing bias toward the target should develop earlier in this condition than in the common ground condition. Thus, the key effect demonstrating integration of the speaker's perspective with his message would be shown in differences on the linear or quadratic components in the common ground condition compared to the listener-privileged condition. Regarding effects in the no-competitor condition, we can predict that if participants consider that the avatar has access to a relevant competitor then the effects should be similar to those in the common ground condition (i.e., delayed target bias but a steep rise postdisambiguation). In contrast, if participants ignore whether the avatar might hold privileged information (which is likely because the participant cannot act on an object in the occluded space), then responses should be as if only the single target object is available. This would therefore lead to similar effects to those predicted for the listener-privileged condition, but these effects should be more pronounced given that no competitor object is being considered at all.

As with the accuracy and RT analyses, each model included fixed effects of competitor, WML, WMC, and Stroop interference scores, alongside the time polynomials (for postonset period only). Both competitor and WML variables were deviation coded as described previously, and WMC and Stroop interference scores were kept as continuous variables. In all models we sought to include random effects of item and participant on all polynomial time terms, and participant-by-competitor, participant-by-WML, item-by-competitor and item-by-WML random effects on all polynomial time terms. Where models did not converge, the random effects structure was reassessed and the random effects that led to nonconvergence were removed.

Results

WML task accuracy: Manipulation check. Overall accuracy on the WML task was high, with a mean accuracy score of 90.88% (*SD* = 15.39%) collapsed across both high and low WML conditions. Mixed-model analyses revealed a significant effect of WML on digit recall accuracy (Estimate $= -.152$, $SE = .027$, $t = -5.83$, $p < .001$) with higher accuracy in the low WML condition ($M = .98$, $SD = .03$) compared to the high WML

---

[3] Note that some previous research suggests that it takes 200 ms to program and launch an eye movement (see Hallett, 1986), which appears to fit with the time course of shifts to target in the grand mean plots. However, we chose to conduct analyses time-locked to the absolute onset of the target noun rather than from 200 ms because of more recent debates in the variability in estimates of this eye movement delay (see Altmann, 2011) and because there may be slightly earlier shifts in attention to the target when the conditions are examined separately.

condition ($M = .83$, $SD = .13$; see Figure 3). This indicated, as expected, that recall of digits was more difficult in the high WML condition than the low WML condition. The type of competitor did not influence digit recall accuracy, and neither WMC nor inhibitory control influenced recall accuracy (all $ps > .1$).

Referential communication task.

Accuracy and selection response times. The mean target selection accuracy rate over all trials was 99.46% ($SD = 3.36\%$).[4] Given that the accuracy rates were near ceiling no further analyses were carried out on accuracy scores.

All selection RTs were within 2.5 standard deviations of the mean within each participant and within each condition. However, three data points (constituting .002% of total data) were longer than 6 s (6,637 ms; 7,203 ms; and 9,161 ms), these were considered outliers and removed prior to analyses (remaining RTs range: 1,215 ms–5,411 ms). For target selection RTs, analyses revealed a significant effect of WML (Estimate $= -138.89$, $SE = 35.20$, $t = -3.94$, $p < .001$), with faster RTs to select the target object in the low WML condition ($M = 2{,}746$ ms, $SD = 246$) compared with the high WML condition ($M = 2{,}886$ ms, $SD = 316$; see Figure 4). There was also a significant effect of competitor on RTs (Contrast 1: $t = -2.32$; $p < .05$; Contrast 2: $p > .05$) with slower RTs in the common ground condition ($M = 2{,}870$ ms, $SD = 298$) compared with the no-competitor condition ($M = 2{,}763$ ms, $SD = 314$; Estimate $= 108.49$, $SE = 38.47$, $t = -1.36$, $p < .01$). By contrast, there were no significant differences in RTs between the listener-privileged condition ($M = 2{,}813$ ms, $SD = 266$) and either the common ground condition or the no-competitor condition (all $ps > .1$). There was also no significant interaction between competitor and WML on target selection RTs ($ps > .1$).

In relation to individual differences of inhibitory control and WMC, only inhibitory control showed an effect on RTs to select the target object, with faster RTs related to higher levels of inhibitory control (Estimate $= -95.77$, $SE = 40.03$, $t = -2.39$, $p < .05$). There were no interactions between either inhibitory control or WMC with competitor condition or WML (all $ps > .1$). The mean WML accuracy, target-object selection RTs, and selection accuracy for each condition are displayed in Table 1.

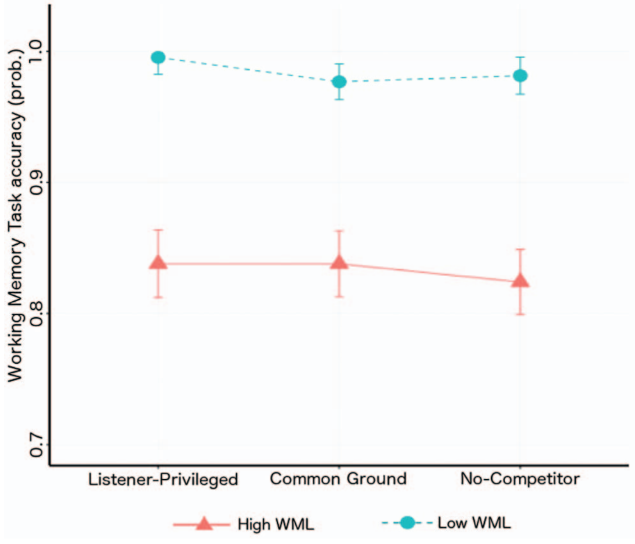

Figure 3. Working memory load task recall accuracy by competitor condition and working memory load condition for Experiment 1. See the online article for the color version of this figure.

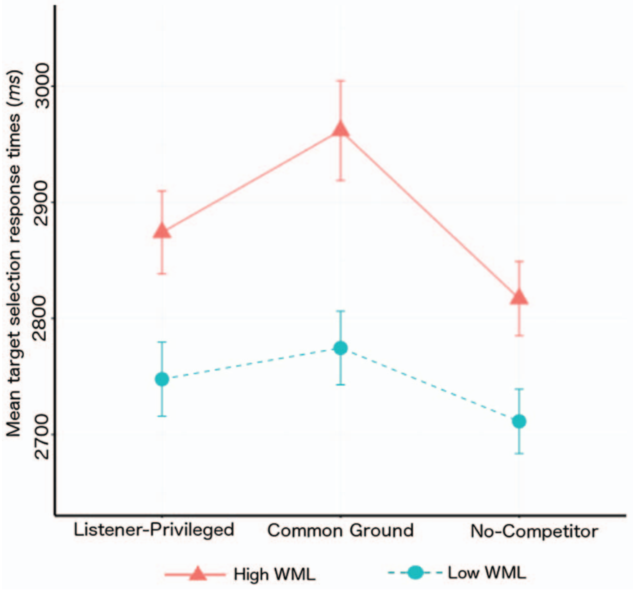

Figure 4. Target selection response times by competitor condition and working memory load condition for Experiment 1. See the online article for the color version of this figure.

Eye-tracking measures. Figure 5 shows the full time-course of target biases by competitor and WML conditions.

Pretarget-noun onset. Preonset mixed-model estimates, t values, and standard errors are shown in Table 2.

Analyses for this preonset period did not reveal any significant effects of competitor or WML or interactions between competitor and WML on fixations to the target object (all $ps > .05$), Similarly there were no significant interactions between competitor and individual difference variables of inhibitory control or WMC in this preonset period.

Posttarget-noun onset. Postonset mixed-model estimates, t values, and standard errors are shown in Table 3. Analyses in this postonset period revealed a significant effect of competitor for both intercept and quadratic fits (Intercept: all $ps < .05$; Quadratic: Contrast 1, $p < .001$; Contrast 2, *ns*). For the intercept, the target preference was significantly smaller in the listener-privileged condition compared with the no-competitor condition (Estimate $= -.16$, $SE = .015$, $t = -10.70$, $p < .001$) but significantly greater in the listener-privileged competitor condition compared with the common ground condition (Estimate $= .30$, $SE = .015$, $t = 2.16$, $p < .05$). For the quadratic fit, there was significantly less curvature in the no-competitor condition compared with the listener-privileged condition (Estimate $= -.46$, $SE = .048$, $t = -8.83$, $p < .001$) and the common ground condition (Estimate $= -.42$, $SE = .048$, $t = -8.65$, $p < .001$), but no significant difference between the listener-privileged and common ground condition ($p > .1$). Thus, though the results from the intercept

---

[4] Of participants, 31 out of 36 had 100% accuracy rates, only 5 participants made a limited amount of errors on trials.

Table 1
Means and Standard Deviations for Working Memory Load Accuracy, Selection Response Times, and Selection Accuracy for Experiment 1 and 2

			WML accuracy (probability)		Selection RT (ms)		Selection accuracy (probability)	
Experiment	Perspective	WML	*M*	*SD*	*M*	*SD*	*M*	*SD*
1: No motivation	Listener-privileged	High	.838	.193	2874.98	309.40	1.000	—
		Low	.995	.028	2752.36	293.10	.995	.028
	Common ground	High	.838	.185	2966.33	407.93	.986	.061
		Low	.977	.058	2775.32	256.62	.991	.039
	No-competitor	High	.824	.178	2816.34	368.69	1.000	—
		Low	.981	.053	2709.93	303.09	.995	.028
2: Motivation	Listener-privileged	High	.720	.190	2104.19	416.37	.995	.030
		Low	1.000	—	1977.27	504.71	.989	.042
	Common ground	High	.800	.210	2185.05	419.10	.957	.086
		Low	.990	.040	2154.24	407.77	.979	.057
	No-competitor	High	.820	.160	1862.17	546.12	.995	.030
		Low	.990	.040	1889.82	504.60	.995	.030

analysis indicate that participants were overall more likely to fixate on the target object in the listener-privileged compared with common ground condition, results from the quadratic fit provide no evidence that this bias developed earlier in the listener-privileged condition than in the common ground condition.

Although there was no effect of WML (all *p*s > .1), there were significant interactions between competitor × WML on the intercept and quadratic fits (Intercept: Contrast 2, $p < .001$; Quadratic: Contrast 1, $p < .05$). Analyses revealed a significant difference in WML effects for the common ground condition only, where a decision between the target and competitor objects was required. Here, WML impacted fixations for both the intercept and quadratic components—there were significantly more fixations to the target under low load compared to high load for the intercept (Estimate = −.038, $SE = .016$, $t = -2.34$, $p < .05$), and there was significantly more curvature in the high-load condition compared to the low-load condition for the quadratic component (Estimate = .14, $SE = .066$, $t = 2.09$, $p < .05$). The quadratic effect represented a later onset of fixation shift toward the target object in the high-load condition compared with the low-load condition. There were no differences between WML conditions in either the listener-privileged condition or the no-competitor condition (all *p*s > .1). Between competitor conditions, there was significantly greater target preference in the listener-privileged condition com-

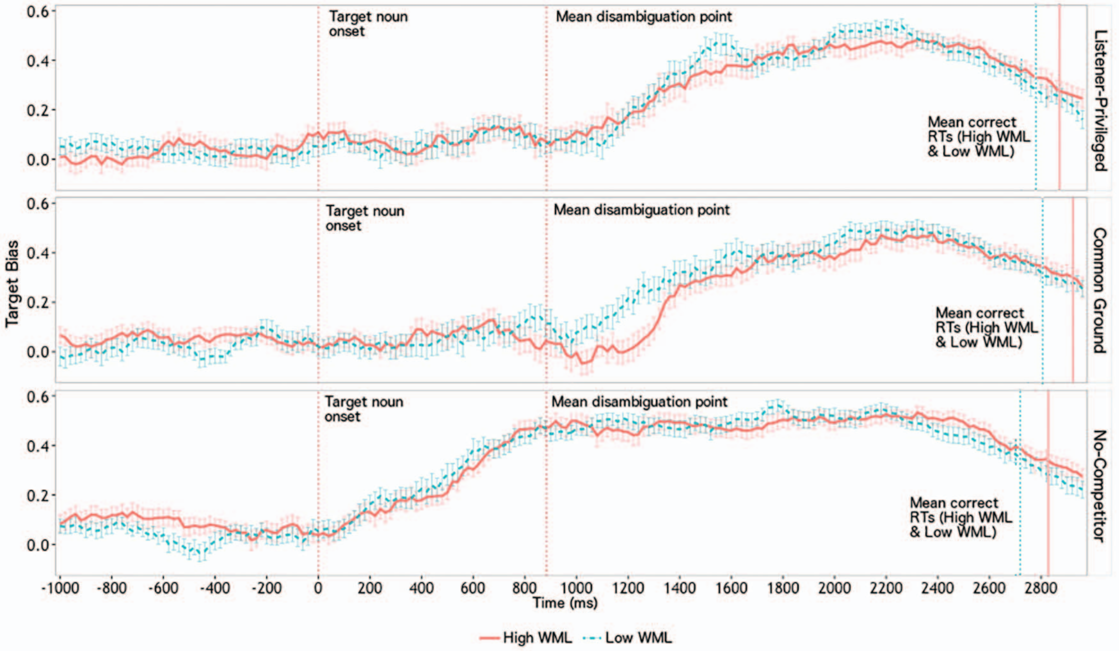

Figure 5. Full time-course of target preference by competitor condition and working memory load for Experiment 1. See the online article for the color version of this figure.

Table 2

Estimates and t Values for the Pre-Onset Period for Experiments 1 and 2

	Experiment 1: No motivation			Experiment 2: Motivation		
Fixed effect	Estimate	*SE*	*t*	Estimate	*SE*	*t*
Intercept	.039***	.011	3.67	.061***	.011	5.41
Competitor contrast						
C1: LP and CG vs. NC	−.030	.017	−1.81	−.044**	.017	−2.59
C2: LP vs. CG	−.001	.018	−.05	.034	.021	1.66
WML	.018	.015	1.19	−.010	.016	−.60
WMC	−.002	.008	−.22	.003	.009	.33
Inh. Cont.	.016*	.008	2.08	−.003	.009	−.38
C1 × WML	−.038	.031	−1.25	−.020	.033	−.61
C2 × WML	−.017	.035	−.49	−.076*	.039	−1.96
C1 × WMC	.009	.017	.54	.0002	.017	.02
C2 × WMC	.0004	.018	.03	.016	.021	.74
C1 × Inh. Cont.	.003	.017	.16	−.001	.017	−.06
C2 × Inh. Cont.	.030	.018	1.72	−.011	.021	−.52
WML × WMC	.003	.015	.19	−.022	.016	−1.35
WML × Inh. Cont.	.024	.015	1.62	−.024	.016	−1.45
C1 × WML × WMC	−.003	.032	−.10	.032	.034	.93
C2 × WML × WMC	−.026	.036	−.70	.052	.040	1.31
C1 × WML × Inh.Cont.	.025	.031	.82	.062	.034	1.80
C2 × WML × Inh. Cont.	−.030	.035	−.85	.043	.040	1.09
N		36			31	

Note. C1 = Contrast 1: Listener-Privileged + Common Ground vs. No-Competitor; C2 = Contrast 2: Listener-Privileged vs. Common Ground; WML = working memory load; WMC = working memory capacity; Inh. Cont. = inhibitory control.
* $p < .05$. ** $p < .01$. *** $p < .001$.

pared with the common ground condition under high WML (Estimate = .049, *SE* = .016, $t = 3.02$, $p < .01$), but not under low WML (*p*s > .1). In contrast, the no-competitor condition showed significantly greater target bias compared with the common ground and listener-privileged conditions under both high and low WML (high WML: listener-privileged, Estimate = .15, *SE* = .016, $t = 9.39$, $p < .001$; common ground, Estimate. = .20, *SE* = .016, $t = 12.41$, $p < .001$; low WML: listener-privileged, Estimate = .16, *SE* = .016, $t = 10.05$, $p < .001$; common ground, Estimate = .18, *SE* = .016, $t = 10.93$, $p < .001$).

There were a number of significant interactions between competitor and individual inhibitory control scores for both the linear and quadratic fits (see Table 3), however post hoc analyses revealed that effects of inhibitory control were only present in the common ground condition. Specifically, greater inhibitory control was significantly associated with a steeper linear fit (Estimate = .14, *SE* = .040, $t = 3.54$, $p < .001$) and greater quadratic curvature (Estimate = .089, *SE* = .038, $t = 2.36$, $p < .05$) in this common ground condition.

There was also a significant three-way interaction between Competitor × WML × WMC for both the linear and quadratic fits (linear: *p*s < .05; quadratic: Contrast 1, $p < .001$; Contrast 2, *ns*). Follow-up analyses revealed only nonsignificant relationships for relating to the linear fits (all *p*s > .1), however both the no-competitor condition and the listener-privileged condition showed significant interaction effects of WMC × WML for the quadratic fit (listener-privileged, $p < .05$; no-competitor, $p < .01$). For the no-competitor condition under high load, greater WMC was related to a greater curvature in target preference (Estimate = .092, *SE* = .07, $t = 1.99$, $p < .05$) meaning that the onset in shift to target was steeper with greater WMC. Under low load, there was no significant relationship between WMC and shifts in target preference ($p > .1$). Further analyses did not reveal any significant effects or relationships for the listener-privileged condition (all *p*s > .1).

In sum, behavioral data from target selection RTs in Experiment 1 failed to show a RT advantage for the listener-privileged condition compared to the common ground condition. In addition, the eye-movement results showed no evidence that participants were sensitive to the perspective of the speaker during the ambiguous period just prior to target-noun onset (cf. Barr, 2008b). Biases to the target object in this listener-privileged condition did not differ from biases in the common ground condition during the pretarget-noun period. In contrast, there was some evidence of sensitivity to perspective in the listener-privileged condition after the onset of the target object noun: there was a stronger target bias in the listener-privileged condition than the common ground condition for the intercept in this postonset period. However, this intercept term aggregates data across the 2-s postonset period (including disambiguating information). This effect would therefore be expected if the onset of the noun prompted participants to process the speaker's perspective, irrespective of whether the speaker's perspective was then used to disambiguate reference. The critical evidence to demonstrate that the speaker's perspective was being used to disambiguate reference would come from a faster-changing bias toward the correct referent in the listener-privileged condition compared with the common ground condition, but there was no such evidence from effects on the linear or quadratic components.

Effects of WML were apparent on RTs in Experiment 1, where a concurrent high WML delayed target selection responses in the

Table 3
Estimates and t Values for the Post-Onset Period for Experiments 1 and 2

	Experiment 1: No motivation			Experiment 2: Motivation		
Fixed effect	Estimate	*SE*	*t*	Estimate	*SE*	*t*
Intercept	.218***	.01	17.66	.257***	.012	21.86
Competitor Contrast						
C1: LP and CG vs. NC	−.172***	.013	−13.69	−.147***	.011	−13.48
C2: LP vs. CG	.031*	.015	2.16	.084***	.013	6.72
WML	−.018	.014	−1.31	−.028*	.013	−2.11
WMC	−.003	.010	−.28	.008	.010	.80
Inh. Cont.	.023*	.010	2.41	.006	.010	.58
C1: WML	−.007	.012	−.55	−.015	.013	−1.12
C2: WML	.035*	.014	2.48	−.007	.015	−.45
C1: WMC	.008	.013	.63	.001	.011	.05
C2: WMC	.002	.015	.14	.022	.013	1.70
C1: Inh. Cont.	−.021	.013	−1.68	−.007	.011	−.64
C2: Inh. Cont.	.021	.015	1.44	−.005	.013	−.41
WML: WMC	.002	.012	.13	−.003	.010	−.31
WML: Inh. Cont.	.018	.012	1.56	−.012	.010	−1.26
C1: WML: WMC	−.003	.012	−.27	−.004	.014	−.26
C2: WML: WMC	.026	.014	1.80	−.011	.016	−.69
C1: WML: Inh. Cont.	.004	.012	.35	−.021	.014	−1.56
C2: WML: Inh. Cont.	−.006	.014	−.43	.014	.016	.89
Linear fit						
Linear	.535***	.040	13.48	.482***	.039	12.40
C1: LP and CG vs. NC	.022	.037	.60	.278***	.055	5.02
C2: LP vs. CG	.049	.042	1.16	−.080	.064	−1.25
WML	−.038	.076	−.50	.018	.064	.29
WMC	.012	.032	.37	−.059	.032	−1.87
Inh. Cont.	.056	.032	1.78	.041	.032	1.31
C1: WML	−.071	.054	−1.31	.203***	.058	3.48
C2: WML	.046	.062	.74	.207**	.067	3.07
C1: WMC	−.011	.037	−.29	−.056	.057	−.97
C2: WMC	−.064	.043	−1.49	−.031	.066	−.46
C1: Inh. Cont.	.084*	.037	2.26	−.019	.057	−.34
C2: Inh. Cont.	−.116**	.043	−2.69	.065	.066	.98
WML: WMC	.002	.059	.04	−.041	.044	−.94
WML: Inh. Cont.	.032	.059	.55	−.006	.044	−.13
C1: WML: WMC	−.112*	.056	−2.02	−.004	.061	−.06
C2: WML: WMC	.153*	.064	2.39	−.110	.070	−1.56
C1: WML: Inh. Cont.	−.146**	.055	−2.63	−.012	.060	−.20
C2: WML: Inh. Cont.	.145*	.064	2.27	.099	.070	1.42
Quadratic fit						
Quadratic	.045	.031	1.43	−.058	.030	−1.93
C1: LP and CG vs. NC	.424***	.042	10.09	.551***	.042	13.19
C2: LP vs. CG	.009	.048	.18	−.140**	.048	−2.90
WML	.039	.055	.71	.046	.053	.88
WMC	.037	.025	1.49	.002	.024	.07
Inh. Cont.	−.003	.025	−.14	−.003	.024	−.11
C1: WML	.115*	.054	2.12	.028	.058	.47
C2: WML	−.120	.062	−1.93	.073	.067	1.08
C1: WMC	.031	.043	.73	−.113*	.043	−2.61
C2: WMC	−.060	.049	−1.21	−.049	.050	−.98
C1: Inh. Cont.	.071	.043	1.67	.020	.043	.47
C2: Inh. Cont.	−.138**	.049	−2.80	.056	.050	1.13
WML: WMC	−.019	.040	−.48	−.033	.038	−.86
WML: Inh. Cont.	−.025	.040	−.64	−.017	.038	−.44
C1: WML: WMC	−.257***	.055	−4.64	−.057	.061	−.93
C2: WML: WMC	−.017	.064	−.27	−.039	.070	−.55
C1: WML: Inh. Cont.	.049	.055	.89	.017	.060	.28
C2: WML: Inh. Cont.	.031	.064	.48	−.054	.070	−.78
N		36			31	

Note. C1 = Contrast 1: Listener-Privileged + Common Ground vs. No-Competitor; C2 = Contrast 2: Listener-Privileged vs. Common Ground; WML = working memory load; WMC = working memory capacity; Inh. Cont. = inhibitory control.
* $p < .05$. ** $p < .01$. *** $p < .001$.

referential communication task. However, because this effect did not interact with competitor type we can infer that this reflects a general processing delay rather than a specific WML effect on perspective taking ability. Nevertheless, eye-movement analyses did reveal different effects of WML between the competitor conditions; there was no effect of WML within the listener-privileged condition or no-competitor condition, however delayed fixations to the target were found under high load compared with low load in the common ground condition where the ambiguity could not be resolved by perspective taking. Furthermore, inspection of the time-courses suggests that after disambiguation, participants were quicker to initiate fixations to the target object under high load in the listener-privileged condition compared with under high load in the common ground condition. Effects of perspective after the point of disambiguation are potentially surprising, because by this point the language alone provides sufficient information to identify a unique referent. However, it is the case that if perspective is used in combination with the disambiguating language, then in the privileged ground condition participants need only use the disambiguated language to confirm the right referent from two possible objects, whereas in the common ground condition they must use the information to identify the correct referent given that perspective could not be used to narrow down the intended referent prior to this point. We suggest that it is this identification process that is impaired by high WML in the common ground condition. Finally, it is interesting to note that individual differences in inhibitory control and WMC did not modulate perspective use in this experiment.

Experiment 2

Although the competitor effects in the postnoun period of Experiment 1 suggest that participants were sensitive to the speaker's limited perspective, the finding that participants did not show a target bias prior to noun onset, or a steeper bias to fixate the target object following the noun onset, in the listener-privileged condition suggests that they experienced initial interference from the privileged competitor object, which delayed the target preference from emerging until later in the verbal instruction. Alternatively, it could be that participants simply learnt that the ambiguity would be resolved at the sentence end (e.g., "with the spots on"), and thus chose not to deploy perspective taking at all until this disambiguating information had become available. This account is consistent with previous research, which has shown that while listeners are spontaneously sensitive to others' perspectives, they can be delayed in the explicit *use* of this information to predict others' actions (Ferguson & Breheny, 2012; Ferguson et al., 2015), however it contrasts with other studies in which participants do not delay perspective use until a temporary ambiguity is resolved (e.g., Brown-Schmidt et al., 2008; Heller et al., 2008). The key difference between these studies is interactivity; those tasks in which participants are actively engaged in an interactive (e.g., question–answer) discourse report faster use of perspective to identify an ambiguous referent (see Ferguson et al. (2015) and Salverda, Brown, & Tanenhaus (2011) for further discussion of these effects). Either of these accounts could therefore be taken as evidence that perspective taking is subject to an initial egocentric bias (at least in low interactivity situations), or that it is not routinely deployed during social interactions, but is only activated as a later mechanism to resolve ambiguities according to need and when there is sufficient motivation to do so (Epley, Keysar, et al., 2004). The importance of motivating factors in perspective taking has been demonstrated by Savitsky et al. (2011) who examined the differences in perspective use between friends versus strangers. Results showed that when participants followed the directions of a friend they made more egocentric errors compared to when following the directions of a stranger. The authors interpret these effects as evidence that people are less motivated to track others' perspectives when they know them well; they assume that friends share their perspectives. In addition, it has been proposed that perspective is more salient when individuals are engaged in an interactive task that collaboratively establishes each person's perspective (Brown-Schmidt & Hanna, 2011). Thus, the presence of a real versus virtual director may act as a motivator to actively employ perspective taking abilities in the earliest moments of language processing.

Experiment 2 sought to examine these possibilities further by encouraging participants to use all available cues (including perspective) to interpret the avatar's instructions as quickly as possible, and to anticipate target objects prior to disambiguating information. This was achieved by replicating the experimental design from Experiment 1, with the addition of a monetary reward that motivated participants to make rapid and accurate selection responses. Participants were given feedback on their performance on a trial-by-trial basis, and correct responses that were made within a limited response period were rewarded. It was expected that this time-pressure would motivate participants to make use of all available cues to facilitate early reference disambiguation, including the speaker's perspective. Thus, if the delay in perspective use seen in Experiment 1 was due to a lack of motivation, then competitor effects should emerge during the anticipatory period and on behavioral measures in this new design. However, if the delay in Experiment 1 was due to a pervasive egocentric first competitor effect then the difference between privileged and common ground conditions will again be delayed here. Once perspective taking was fully engaged, we set out to examine how WML might influence its effects.

Method

Participants. Native English speaking University of Kent students ($N = 31$) were recruited through the university-wide research participation scheme. Twenty-three participants were female, and 8 were male. The mean age of participants was 21.26 years ($SD_{\text{age}} = 2.90$).

Stimuli, design, and procedure. The stimuli and design were the same as those described in Experiment 1, however in this task additional cues prompted participants to respond quickly and accurately; participants were motivated to respond quickly and accurately through financial rewards. It was expected that under time pressure, participants would make use of all available cues, including the speaker's perspective, to facilitate fast and accurate responses. To implement the time-pressure a beep was introduced on practice trials if participants had not selected an object to move within 2.75 s of the target-noun onset. This timing was chosen because it was marginally shorter than the mean RTs in Experiment 1 but was long enough to allow participants to hear the full instructions from the avatar. In addition, participants were informed that they would receive 2 pence (~0.03 USD) for each correct object selected, 2 pence for each object selected within the time limit, and a further 2 pence if the WML task digit was recalled correctly. Feedback on performance, as well as the total money won on a given trial and cumulate across the task, was presented on-screen to participants at the end of each trial after the memory recall prompt.

In total, participants could receive a maximum of 14 pence (~0.20 USD) per trial, and they were told the maximum they could win over the whole task was £6 (~7.50 USD). After completing the task, all participants were informed that anyone who had made over £5 (~6.40 USD) would receive the £6 maximum (based on a pilot test of the task we knew that all participants completing the task correctly would achieve over £5). Thus, all participants received the same payments, which ensured that they did not tell others that you receive the maximum irrespective of performance.

Results

Data preparation and analysis procedures were identical to those reported for Experiment 1. Descriptive statistics for accuracy in the WML manipulation, selection RTs and accuracy in the referential communication task for Experiment 2 are shown in Table 1. Reliability analyses for the Stroop task and OSPAN task revealed high reliability in all conditions for the Stroop task (incongruent: α = .91; congruent: α = .89; neutral: α = .89) and high reliability for the OSPAN task (α = .95).

WML accuracy. The mean response accuracy in the WML task was high (88.60%). Mixed-model analysis revealed a significant effect of WML (Estimate = −.21, *SE* = .027, *t* = −7.62, *p* < .001), reflecting higher accuracy on low WML trials compared to high WML trials (99% vs. 78% respectively, see Figure 6). There was also a marginally significant interaction effect between WML and Competitor (Estimate = −.081, *SE* = .042, *t* = −1.91, *p* = .058). WML effects were shown in all competitor conditions, with lower accuracy under high load than under low load (listener-privileged: Estimate = −.27, *SE* = .036, *t* = −7.46, *p* < .001; common ground: Estimate = −.19, *SE* = .036, *t* = −5.22, *p* < .001; no-competitor: Estimate = −.17, *SE* = .036, *t* = −4.92, *p* < .001; see Table 1). With regard to differences between competitor conditions, under high load there was significantly decreased accuracy in the listener-privileged condition (72%) compared with both the no-competitor (82%: Estimate = .081, *SE* = .030, *t* = −2.71, *p* < .01) and the common ground condition (80%: Estimate = .069, *SE* = .030, *t* = 2.35, *p* < .05). The type of competitor did not influence accuracy under low load (all *p*s > .5), where all scores were at or near ceiling. For individual differences of WMC and inhibitory control, only WMC influenced accuracy in the WML task. Here, greater WMC was significantly related to greater overall accuracy (Estimate = .027, *SE* = .013, *t* = 1.981, *p* = .05), however this relationship was not affected by WML or competitor condition.

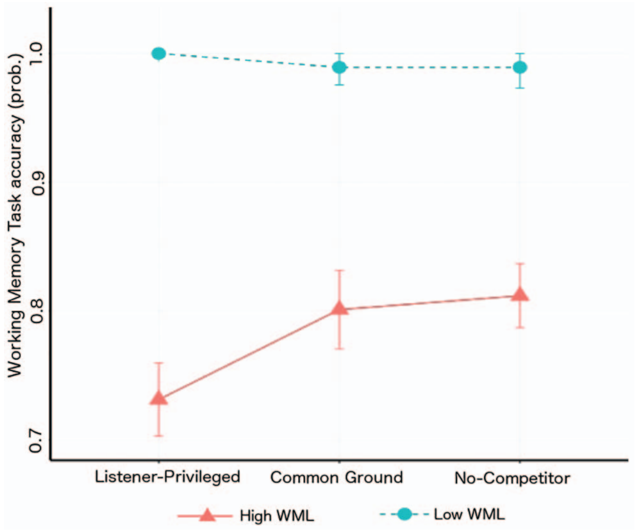

Figure 6. Working memory load task accuracy by competitor condition and working memory load condition for Experiment 2. See the online article for the color version of this figure.

Accuracy and selection response times. As in Experiment 1, the mean overall accuracy for correct target selection was high (98.48%) and given that the accuracy rates were near ceiling no further analyses were carried out.[5]

For RTs, examination of outliers indicated that only 0.72% of RTs could be considered outliers—these were removed from subsequent analyses. Mixed-model analysis of the remaining RTs revealed that while there was no significant effect of WML (*p* > .1), there was a significant effect of competitor (Contrast 1: Estimate = 228.33, *SE* = 77.18, *t* = 26.26, *p* < .001; Contrast 2: Estimate = −119.35, *SE* = 42.56, *t* = −2.81, *p* < .01), and a significant Competitor × WML interaction for Contrast 1 (Contrast 1: Estimate = −119.35, *SE* = 42.56, *t* = −2.81, *p* < .01; Contrast 2: *p* = .11; see Figure 7). Post hoc analyses revealed a significant effect of WML in the listener-privileged condition only (Estimate = 121.96, *SE* = 45.56, *t* = 2.68, *p* < .01), with significantly longer target selection RTs under high WML compared with low WML (see Table 1). RTs were significantly different between the competitor conditions under low load, with responses in the common ground condition significantly slower than in the listener-privileged condition (Estimate = −166.74, *SE* = 51.69, *t* = −3.23, *p* < .01), but no significant difference in RTs in the no-competitor condition compared to the listener-privileged condition (*p* > .1). Under high load, there were no differences in RTs between the common ground and listener-privileged conditions (*p* > .1), but responses in the no-competitor condition were significantly faster than in the listener-privileged condition (Estimate = 243.17, *SE* = 60.00, *t* = 4.05, *p* < .001). There were no significant effects or interactions relating to the individual difference variables of inhibitory control and WMC for RTs in Experiment 2 (all *p*s > .1).

Eye-tracking measures.

Fixation time-course. Examination of the grand mean of fixations for Experiment 2 revealed the effects to be within the same time limits shown in Experiment 1; so to provide a relevant contrast the same analysis, time-windows were kept from −1,000

---

[5] Of participants, 20 out of 31 had 100% accuracy on object selection, and a further 8 participants made a selection error on only one trial.

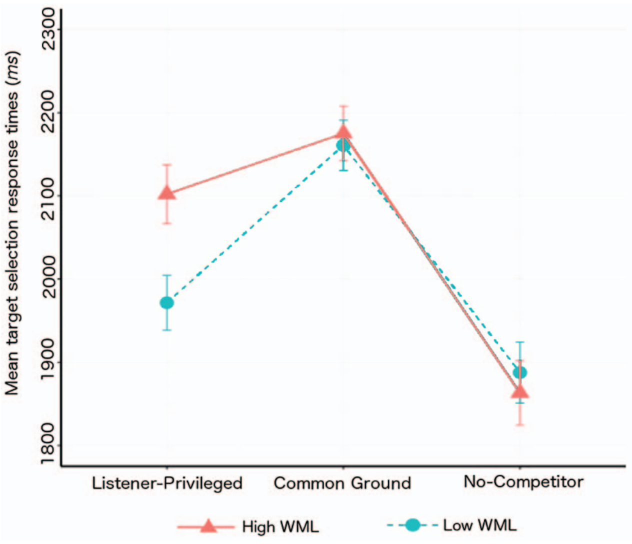

Figure 7. Target selection response times by competitor condition and working memory load condition for Experiment 2. See the online article for the color version of this figure.

ms to target-noun onset (preonset analysis time-window) and from target-noun onset (0 s) to 2,000 ms (postonset analysis time-window). The time-course of fixations to the target object versus the competitor object in each competitor and WML condition is shown in Figure 8.

Pretarget-noun onset. As for Experiment 1, we only analyzed effects on the intercept for the pretarget-noun onset period. Preonset mixed-model estimates, standard errors, and *t* values are shown in Table 2.

Examination of Figure 8 indicated an anticipatory fixation bias toward the target object in the low-load condition of the listener-privileged condition. Analyses revealed that, for this preonset period, there was a significant effect of competitor (Contrast 1: $p < .01$; Contrast 2: *ns*), with significantly greater target preference in the no-competitor condition compared with the common ground condition only (Estimate = .061, *SE* = .020, $t = 3.03$, $p < .001$). Target preference in the listener-privileged condition did not significantly differ from the common ground condition (all $ps > .1$). There was, however, a significant Competitor × WML interaction (Contrast 1: *ns*; Contrast 2: $p = .05$). Analysis of this interaction effect revealed a significant effect of WML on target preference for the listener-privileged condition only, with a reduced target preference under high load compared with low load (Estimate = .054, *SE* = .027, $t = 1.97$, $p < .05$). Between competitor conditions, there was a significantly greater target bias in the listener-privileged condition compared with the common ground condition under low WML (Estimate = −.072, *SE* = .028, $t = -2.55$, $p < .05$) but no difference under high load ($p > .5$). In contrast, there was a significantly lower target bias in the listener-privileged condition compared to the no-competitor condition under high WML (Estimate = −.055, *SE* = .027, $t = -2.02$, $p < .05$) but no difference under low load ($p > .5$). These findings indicate that prior to the target object being referred to participants were able to ignore objects not in common ground under low load as if they were not present (as in the no-competitor condition). However, under high-load participants were distracted by privileged objects as if these objects were in common ground. There were no significant competitor effects relating to individual dif-

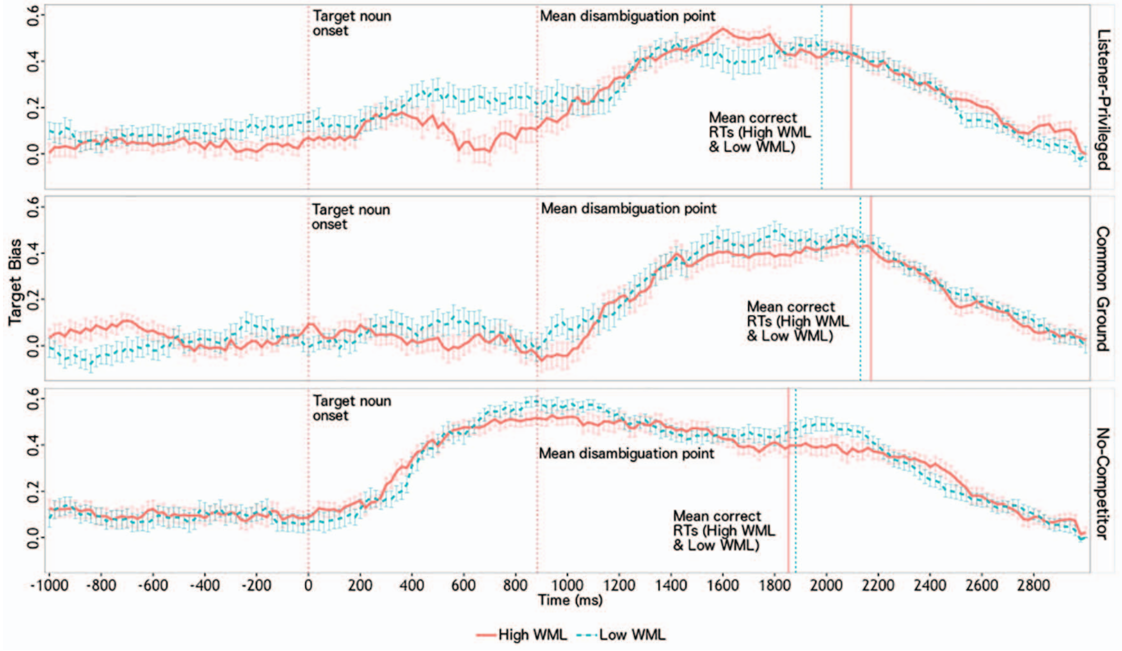

Figure 8. Full time-course of target preference by competitor condition and working memory load for Experiment 2. See the online article for the color version of this figure.

ferences of WMC or inhibitory control for this preonset period (all *p*s $> .05$).

Posttarget-noun onset. Postonset by-participant mixed-model estimates, standard errors, and *t* values are shown in Table 3. Examination of Figure 8 indicates differences between the high and low WML conditions in the listener-privileged condition. In particular, under high load the fixation pattern shifts toward the competitor object between 400 ms and 750 ms, after which listeners' target preference increases. In contrast, the fixation pattern under low load shows a shift toward the target object within 200 ms of the target-noun onset. Comparable differences in fixation patterns between the high and low WML were not apparent in the common ground or no-competitor conditions.

Analysis of these fixation patterns revealed that, in the postonset period, competitor type significantly influenced target preference across the intercept, linear, and quadratic fits (all *p*s $< .001$). There was significantly greater target preference but significantly smaller quadratic curvature in the listener-privileged condition compared with the common ground condition (Intercept: Estimate $= .084$, $SE = .013$, $t = 6.72$, $p < .001$; Quadratic: Estimate $= .14$, $SE = .048$, $t = 2.90$, $p < .01$; Linear: $p > .1$). In contrast, the listener-privileged condition showed a decreased target preference compared with the no-competitor condition, a steeper linear increase, and greater quadratic curvature over time (Intercept: Estimate $= -.105$, $SE = .013$, $t = -8.32$, $p < .001$; Linear: Estimate $= .238$, $SE = .064$, $t = 3.18$, $p < .001$; Quadratic: Estimate $= .481$, $SE = .048$, $t = 9.97$, $p < .001$). These findings indicate that participants were able to use perspective cues to aid earlier selection of the target when a viable competitor was present (i.e., listener-privileged vs. common ground), however target selection remained slower than when the absence of a competitor meant that perspective taking was unnecessary (i.e., listener-privileged vs. no-competitor).

There was also a significant effect of WML ($p < .05$, see Table 3) on target preference and a significant Competitor × WML interaction for the linear fit (Contrast 1: $p < .001$; Contrast 2: $p < .01$). Although all competitor conditions showed greater target preference in the low-load condition compared to the high-load condition (Differences: listener-privileged, Estimate $= -.036$; common ground, Estimate $= -.030$; no-competitor, Estimate $= -.018$), the difference between the WML conditions only reached significance for the listener-privileged condition ($t = -2.29$, $p < .05$; common ground, $p = .066$; no-competitor, $p > .1$). The listener-privileged condition also showed significantly steeper linear fit in the high WML condition compared with the low WML condition (Estimate $= .190$, $SE = .074$, $t = 2.53$, $p < .05$) due to later onset of fixations to target compared with under high WML compared with low WML. These linear effects were not apparent in either the common ground condition or the no-competitor condition (all *p*s $> .1$). Together these findings indicate that WML impedes use of information about the speaker's perspective to resolve referential ambiguity in favor of the target.

For individual difference variables, analyses revealed a significant interaction of Competitor × WMC for the quadratic fit only. Post hoc analyses revealed that the relationship between target preference and WMC was only present in the no-competitor condition, where greater WMC was related to increased quadratic curvature (Estimate $= .077$, $SE = .038$, $t = 2.05$, $p < .05$), as in Experiment 1. There were no effects of WMC or inhibitory control relating to perspective use in the listener-privileged condition or the common ground condition (*all p*s $> .05$).

In summary, Experiment 2 finds evidence of sensitivity to the speaker's perspective even before such information can be integrated with his instructions and of use of this information to resolve reference. Specifically, target selection responses were faster in the listener-privileged condition compared to the common ground condition, and eye movements (i.e., target biases) revealed faster anticipation and integration of the mutually available target object in the listener-privileged condition compared with the common ground condition. Crucially however, this enhanced visual bias to the target object in the listener-privileged condition was only apparent under low load; under high-load participants suffered persistent interference from the privileged competitor which delayed target preference to the same extent observed in the common ground condition. Eye movement analyses during integration revealed working memory effects in the listener-privileged condition for the linear component, reflecting a lower gradient in shift to the target object under low WML in the listener-privileged condition, indicating that eye movements toward the target objects began at an earlier time-point under low load. Finally, we did not find any evidence that individual differences in inhibitory control and WMC modulate perspective use, even when perspective taking was fully engaged.

General Discussion

The two experiments reported here examined three potential influences on perspective taking ability: time, cognitive resources, and motivation. Eye movements and behavioral responses were recorded, whereas participants engaged in a referential communication task involving temporary referential ambiguity (e.g., *glass* when more than one glass was visually present). We compared effects in three conditions: a listener-privileged condition where a competitor object was only available to the participant, a common ground condition where target and competitor objects were available to both participant and speaker, and a no-competitor condition where a competitor was not visually available. In addition, WML was manipulated within each condition using a dual-task design that required participants to hold a sequence of digits in memory during the referential communication task (low load vs. high load). In this way we extended previous work examining the degree of cognitive effort required to consider others' perspectives, and how perspective taking might be impacted when these cognitive resources are allocated elsewhere. Crucially, we examined how cognitive load affects anticipation and integration of ambiguous referential objects, and employed sensitive growth curve analyses to examine the time course with which cognitive load shows its effects during integration. In Experiment 2, we tested whether motivating participants to use perspective to resolve reference assignment modulated perspective use and WM effects.

In Experiment 1, where there was no reward or time pressure, listeners showed very limited evidence of having used perspective cues to disambiguate the target object from the competitor in the listener-privileged condition. During the anticipation pretarget-noun period the overall target bias was not different between the listener-privileged condition and the common ground conditions, suggesting that participants had not ruled out the competitor object as a potential target, even though it was hidden from the speaker's view. After the onset of the target-noun participants did show a greater bias toward the target in the listener-privileged compared with the common ground condition, but, as discussed earlier, this only warrants the inference that they were now attending to the speaker's perspective; not that they were using this information to constrain reference. The only sign that participants were actually using perspective information came after the noun onset (which includes the point of disambiguation in the verbal instruction), by which time information about perspective was no longer necessary. Although it was no longer necessary, our data are compatible with participants using perspective information, resulting in more efficient target selection. Specifically, under high WML, earlier shifts in target bias fixations occurred when perspective cues were present (i.e., in the listener-privileged condition) compared with when they were not (i.e., in the common ground condition). Surprisingly, there was no advantage of the presence of perspective cues under low-load conditions wherein similarly rapid shifts to the target object were shown in both the listener-privileged condition and the common ground condition following the noun onset.

In contrast in Experiment 2, where a reward for speed and accuracy of responses was applied, listeners showed clear evidence both of early sensitivity to perspective and of having used perspective cues to disambiguate the target object from the competitor object. Here, target selection responses were faster in the listener-privileged condition compared with the common ground condition (though only under low WML). Analysis of eye movements also revealed facilitation effects when perspective could be used to narrow down the intended target object (i.e., listener-privileged condition); participants anticipated a mutually available object prior to the onset of the noun, and were faster to integrate that target object following onset of the noun. Target biases were also stronger under low load than high load, in both competitor conditions. Crucially, under low-load participants showed a stronger visual bias to the target object in the listener compared with the common ground condition, affecting both the pretarget-noun period and the posttarget-noun period. In other words, participants directed their expectations about forthcoming referents to those in common ground (see also Barr, 2008b), and showed greater ease of integration for those objects in common ground (i.e., they suffered less interference from objects in privileged ground). Furthermore, differences were shown in target preference for the linear component, indicating, in line with our predictions, that not only was there a greater overall target bias in the listener-privileged condition but that the shift to target over time began at an earlier time-point. Under high WML, there was a delayed target preference shift in the listener-privileged condition, as fixation patterns during the pretarget-noun period were comparable with when both objects were in common ground.

Taken together, these results show that perspective cues are used particularly when there is some explicit motivation for fast and accurate performance, and when there are sufficient cognitive resources to do so (we return to this second point shortly). However, in Experiment 1, participants did not show any evidence of perspective taking during the pretarget-noun period; they did not limit their search to objects in the common ground. This pattern suggests either that this inference was only activated once the target noun had been uttered (perhaps because of initial egocentric interference), or that participants had inferred the speaker's perspective prior to the noun but simply did not use this inference immediately to constrain their looking behavior. Further research is necessary to distinguish between these two possibilities, however the fact that participants were able to use perspective early in Experiment 2 (also influencing the speed of their behavioral responses), suggests that lack of motivation for rapid perspective use, and not a default egocentric bias, is responsible for the weaker perspective effect in Experiment 1. This is particularly relevant in our design where participants could simply delay their response until disambiguating information was available in the language input (in contrast to other referential communication tasks in which the target is never fully disambiguated; e.g., tape in Keysar et al., 2003). Thus, the findings fit with previous studies that have suggested that perspective taking abilities are preferentially activated to resolve ambiguities in social interactions when there is sufficient motivation to do so (Epley, Keysar, et al., 2004). Indeed, recent research has demonstrated that perspective inferences are facilitated when doing so is relevant to the task at hand, such as being explicitly instructed to track another person's beliefs (Back & Apperly, 2010; Ferguson et al., 2015) or when prompted by the context in a collaborative task (Brown-Schmidt et al., 2008; Hanna & Tanenhaus, 2004).

It is interesting to note that here, participants in both experiments were explicitly instructed to keep track of the speaker's perspective in order to accurately follow their instructions: The key difference between the two experiments was in their motivation to use this knowledge to facilitate faster responses. As such, the results are also compatible with constraint-based accounts of perspective taking, which suggest that perspective only has immediate effects on reference resolution when strong constraints are provided in the discourse to narrow down the intended referent (Hanna et al., 2003). In the current experiment, participants could resolve the ambiguity and respond correctly to the instructions by simply waiting for the disambiguating information to follow the noun (e.g., "with the spots on"). Thus, this verbal disambiguating information provides a stronger constraint on language comprehension than inferences about perspective, which might be either delayed or simply not computed. However, in Experiment 2 where participants were motivated to respond quickly through monetary rewards, perspective provides a strong constraint on referential selection and thus shows its effects early on in processing. Thus, it is possible that while the same fixed set of constraints may be available to interlocutors in a discourse, varying cues from the situational context, including motivation, can influence whether and when listeners integrate perspective information to constrain referential interpretation. However, further research is needed to identify whether these findings generalize beyond financial and time-limiting motivations to other types of motivation (e.g., intrinsic/explicit communication concerns, personal goals, social goals).

The results here also showed that when participants are motivated to use perspective (i.e., Experiment 2), WML modulated peoples' ability both to show sensitivity to the speaker's perspective and to use perspective to narrow down the intended target object. Specifically, target selection RTs were significantly faster on listener-privileged trials when WML from the secondary task was low compared to when it was high. In the eye-movement data, participants showed a significantly increased bias to the target object under low load than high load during both the anticipatory pretarget-noun period, and the posttarget-noun period. Indeed, visual biases in the listener-privileged condition under high WML were comparable to when both objects were in common ground (i.e., the common ground condition). Notably, during the posttarget-noun period, growth curve analyses of the time course of biases showed a steeper linear slope in the high WML condition compared with the low WML condition. Further analysis of the quadratic effects showed that under high load, participants considered the competitor object between 400 ms and 750 ms after the ambiguous noun onset, then switched to show a preference to fixate the target object just prior to the point of disambiguation. In contrast, under low-load fixation patterns indicated a shift toward the target object within 200 ms of the ambiguous target-noun onset. This steeper and delayed target bias under high WML reflects the need to compensate for not having restricted expectations toward shared objects prior to the target-noun onset. From these data, we can infer that increased WML disrupts the earliest stages of perspective taking, either by preventing listeners from inhibiting competitor objects in privileged ground or by preventing them from inferring perspective at all. The fact that these effects of WML were evident during the pretarget-noun period suggests that increased WML might operate by preventing encoding of common/privileged ground in the first place, suggesting that working memory is necessary for perspective inferences. Alternatively, inferences about perspective might still take place, but the high WML prevents their use to bias attention prior to the informative part of the message. Further research is required to explore these alternatives.

An interesting finding was that the effects of WML and perspective were bidirectional. That is, though perspective taking ability was impaired when a secondary task placed high demands on WML, the need to use perspective also impacted on recall accuracy in the WML task itself. Recall accuracy was significantly worse under high load in the listener-privileged condition compared with the common ground condition or the no-competitor condition. Because the listener-privileged condition was the one condition that required suppression of a privileged competitor object, we can assert that using another person's perspective draws from limited cognitive resources and detrimentally impacts on subsequent cognitive events. This suggests that both processes rely on an overlapping network of cognitive functions. These effects provide further evidence that considering others' perspectives can be cognitively effortful. This finding is consistent with previous studies that have reported impairments in one's ability to infer others' mental states when cognitive resources are occupied by a secondary task (Bull et al., 2008; McKinnon & Moscovitch, 2007; Schneider et al., 2012; cf. Qureshi et al., 2010). Indeed, the fact that cognitive load began exerting its effects on performance when participants were merely using perspective to anticipate available target objects (i.e., biasing visual attention to objects in common ground during the pretarget-noun period), suggests that either perspective taking, or the influence of this information on eye-movements, may place substantial demands on memory and related executive functions. Future research may build on these findings by exploring whether different kinds of memory loads (e.g., verbal vs. spatial) affect perspective taking differently as has been shown in other cognitive domains (see Winawer et al., 2007).

In contrast to these effects of WML, individual differences in WMC or inhibitory control were not found to be reliable predictors of perspective taking performance in either of the two experiments reported here (so regardless of motivational pressures). While this contrasts with previous studies that have found correlations between individual differences in executive function and performance on referential perspective taking tasks (Brown-Schmidt, 2009b; Lin et al., 2010), it is consistent with more recent research that has raised questions about the robustness of this relationship (see Brown-Schmidt, 2012; Ryskin et al., 2014). Considering methodological differences between the present study and previous studies that have found an executive function–perspective-taking relationship provides further context to the contrasting results. For instance, Brown-Schmidt (2009b) reported a relationship between inhibitory control and perspective use only when the inhibitory control task required inhibition of a verbal response and not when the task required inhibition of a nonverbal (i.e., manual) response. The present study used a version of the Stroop task where participants made nonverbal responses to stimuli. Thus, it seems possible that the specific ability to inhibit a verbal response, rather than a general behavioral response, is important in successful perspective taking in referential communication tasks. Note however that direct comparisons of the verbal and manual response versions of the Stroop task have indicated only minor differences between the modalities (see Redding & Gerjets, 1977), suggesting that verbal information is being inhibited with both types of modality. With regard to the relationship between WMC and perspective taking ability, Lin et al. (2010) used dichotomous grouping to examine the impact of WMC, with participation restricted to those with OSPAN scores within the top 20th or bottom 20th percentiles of their sample (labeled as *high-* and *low-WMC* groups, respectively). In contrast, WMC (OSPAN) scores were operationalized on a continuous scale in the present study (i.e., were ungrouped), which allowed us to take into account individual variation in WMC scores and examine the potential relationship with individual variation in perspective use. It is likely that when such individual variation is taken into account the impact of WMC over other individual difference factors is less important. Further research is required to explore this fully.

Finally, we note that the potential hidden object (behind the occluded grid space) and overinformative verbal description in the no-competitor condition did not delay the reference resolution process. Participants showed a clear and early preference to fixate the target object (i.e., before the onset of the object noun), suggesting that they did not interpret the speaker's description as contrastive and did not search for a speaker-privileged competitor object. Nevertheless, this condition was a useful comparison for the common ground and listener-privileged conditions because it shows how reference assignment progresses in the absence of a visual competitor.

In sum, the present research demonstrates that a speaker's perspective can, in the right circumstances, be integrated rapidly with one's own knowledge to disambiguate between potential referents. Importantly, we show that perspective is only used early when there is sufficient motivation and cognitive resources to do so. When cognitive capacities were allocated elsewhere (i.e., to a demanding secondary task) participants were slower to narrow down the search for a target referent. Analysis of the time-course of these effects reveal that WML disrupts the earliest stages observable in the present study, either by preventing listeners from inhibiting competitor objects in privileged ground or by preventing them from inferring perspective at all. It also disrupts the use of this information to constrain reference once the speaker's message reveals specific alternative possibilities. In addition, we show that using another person's perspective itself is cognitively costly, which has detrimental effects on responses to a secondary task. Together these findings go some way to explain the limitations on, and features of, perspective taking ability—providing evidence emphasizing the importance of domain-general processes in perspective taking and referential communication.

References

Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica, 137,* 190–200. http://dx.doi.org/10.1016/j.actpsy.2010.09.009

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111,* 55–71. http://dx.doi.org/10.1016/j.cognition.2008.12.005

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition, 76,* B13–B26. http://dx.doi.org/10.1016/S0010-0277(00)00073-1

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition, 115,* 54–70. http://dx.doi.org/10.1016/j.cognition.2009.11.008

Barr, D. J. (2008a). Analyzing "visual world" eye-tracking data using multilevel logistic regression. *Journal of Memory and Language, 59,* 457–474. http://dx.doi.org/10.1016/j.jml.2007.09.002

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

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language, 46,* 391–418. http://dx.doi.org/10.1006/jmla.2001.2815

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, S., Vasishth, R., & Baayen, H. (2015). *Parsimonious mixed models*. Retrieved from http://www.ling.uni-potsdam.de/~vasishth/pdfs/BatesEtAlParsimonious2015.pdf

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using "S4" classes*. Retrieved from http://CRAN.R-project.org/package=lme4

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes, 28,* 443–467. http://dx.doi.org/10.1080/01690965.2011.649040

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science, 1,* 274–291. http://dx.doi.org/10.1111/j.1756-8765.2009.01019.x

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

Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16,* 893–900. http://dx.doi.org/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 and Cognitive Processes, 27,* 62–89. http://dx.doi.org/10.1080/01690965.2010.543363

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

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another persons shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse, 2,* 11–33. http://dx.doi.org/10.5087/dad.2011.102

Bull, R., Phillips, L. H., & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of theory of mind and executive function. *Cognition, 107,* 663–672. http://dx.doi.org/10.1016/j.cognition.2007.07.015

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. http://dx.doi.org/10.1037/a0013283

Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua, 120,* 1956–1972. http://dx.doi.org/10.1016/j.lingua.2010.02.005

de Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science, 291,* 1803–1806. http://dx.doi.org/10.1126/science.1056496

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.neuroimage.2010.05.056

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language, 54,* 554–573. http://dx.doi.org/10.1016/j.jml.2005.12.009

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. http://dx.doi.org/10.1037/0022-3514.87.3.327

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

Falissard, B. (2012). *psy: Various procedures used in psychometry*. *R package* (Version 1.1). Retrieved from http://CRAN.R-project.org/package=psy

Ferguson, H. J., Apperly, I., Ahmad, J., Bindemann, M., & Cane, J. (2015). Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task. *Cognition, 139,* 50–70. http://dx.doi.org/10.1016/j.cognition.2015.02.010

Ferguson, H. J., & Breheny, R. (2011). Eye movements reveal the timecourse of anticipating behaviour based on complex, conflicting desires. *Cognition, 119,* 179–196. http://dx.doi.org/10.1016/j.cognition.2011.01.005

Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology, 48,* 257–263. http://dx.doi.org/10.1016/j.jesp.2011.08.007

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

Grodner, D., & Adler, R. (2013). The influence of perspective and communicative goals on how speakers choose to refer. In L. Goldstein (Ed.), *Brevity* (Vol. 15, pp. 264–279). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199664986.003.0016

Grodner, D., Dalini, M., Pearlstein-Levy, S., & Ward, A. (2012, March 26). *Factors that contribute to the use of perspective in referent identifica-*

tion. Paper presented at the 25th CUNY Conference on Human Sentence Processing, New York, NY.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 78–101). New York, NY: Wiley.

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. http://dx.doi.org/10.1207/s15516709cog2801_5

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

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

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

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

Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics, 4*. Advance online publication. http://dx.doi.org/10.1515/IP.2007.004

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. http://dx.doi.org/10.1111/1467-9280.00211

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

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2015). *lmerTest: Tests in linear mixed effects models. R package* (Version 2.0–29). Retrieved from http://CRAN.R-project.org/package=lmerTest

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1118–1133. http://dx.doi.org/10.1037/0278-7393.16.6.1118

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. http://dx.doi.org/10.1016/j.jesp.2009.12.019

Lindsay, S., Scheepers, C., & Kamide, Y. (2013). To dash or to dawdle: Verb-associated speed of motion influences eye movements during spoken sentence comprehension. *PLoS One, 8*(6), e67187. http://dx.doi.org/10.1371/journal.pone.0067187

Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology, 7,* 181–191. http://dx.doi.org/10.1177/0261927X8800700403

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition, 102,* 179–218. http://dx.doi.org/10.1016/j.cognition.2005.12.011

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59,* 475–494. http://dx.doi.org/10.1016/j.jml.2007.11.006

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58,* 220–249. http://dx.doi.org/10.1016/j.cogpsych.2008.07.002

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level 1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition, 117,* 230–236. http://dx.doi.org/10.1016/j.cognition.2010.08.003

Psychology Software Tools, Inc. (2012). *Solution for research, assessment, and research* [E-Prime 2.0]. Pittsburgh, PA. Retrieved from http://www.pstnet.com

R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Redding, G. M., & Gerjets, D. A. (1977). Stroop effect: Interference and facilitation with verbal and manual responses. *Perceptual and Motor Skills, 45,* 11–17. http://dx.doi.org/10.2466/pms.1977.45.1.11

Rubio-Fernández, P. (2013). Perspective tracking in progress: Do not disturb. *Cognition, 129,* 264–272. http://dx.doi.org/10.1016/j.cognition.2013.07.005

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science, 24,* 27–33. http://dx.doi.org/10.1177/0956797612447819

Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T. (2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. *Journal of Memory and Language, 74,* 46–76. http://dx.doi.org/10.1016/j.jml.2014.04.003

Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica, 137,* 172–180. http://dx.doi.org/10.1016/j.actpsy.2010.09.010

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1255–1266. http://dx.doi.org/10.1037/a0018729

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

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science, 23,* 842–847. http://dx.doi.org/10.1177/0956797612439070

Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition, 133,* 43–47. http://dx.doi.org/10.1016/j.cognition.2014.05.016

Schurz, M., Kronbichler, M., Weissengruber, S., Surtees, A., Samson, D., & Perner, J. (2015). Clarifying the role of theory of mind areas during visual perspective taking: Issues of spontaneity and domain-specificity. *NeuroImage, 117,* 386–396. http://dx.doi.org/10.1016/j.neuroimage.2015.04.031

Sedivy, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language, 46,* 341–370. http://dx.doi.org/10.1006/jmla.2001.2812

Sedivy, J. C. K., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109–147. http://dx.doi.org/10.1016/S0010-0277(99)00025-6

SR Research Ltd. (2013). EyeLink Experiment Builder Softeware (Version 1.10.165). Osgoode, Ontario, Canada.

Stroop, R. J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662. http://dx.doi.org/10.1037/h0054651

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634. http://dx.doi.org/10.1126/science.7777863

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154. http://dx.doi.org/10.1016/0749-596X(89)90040-5

Wardlow Lane, L., & Liersch, M. J. (2012). Can you keep a secret? Increasing speakers' motivation to keep information confidential yields poorer outcomes. *Language and Cognitive Processes, 27,* 462–473. http://dx.doi.org/10.1080/01690965.2011.556348

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684. http://dx.doi.org/10.1111/1467-8624.00304

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 7780–7785. http://dx.doi.org/10.1073/pnas.0701644104

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science, 18,* 600–606. http://dx.doi.org/10.1111/j.1467-9280.2007.01946.x

Yoon, S. O., Koh, S., & Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review, 19,* 699–707. http://dx.doi.org/10.3758/s13423-012-0262-6

Received February 12, 2015
Revision received September 2, 2016
Accepted September 2, 2016 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Members of Underrepresented Groups:
Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit http://www.apa.org/pubs/authors/review-manuscript-ce-video.aspx.