# Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task

Heather J Ferguson [1]

Ian Apperly [2]

Jumana Ahmad [1 2]

Markus Bindemann [1]

James Cane [1]

[1] University of Kent, England, UK

[2] University of Birmingham, England, UK

Correspondence to:

Heather Ferguson
School of Psychology
University of Kent
Keynes College
Canterbury, Kent
CT2 7NP, UK

Email: H.Ferguson@kent.ac.uk
Phone: +44 (0)1227 827120 Fax: +44 (0)1227 827030

Author Notes.

This work was carried out with the support of a grant to HF and IA from the Leverhulme Trust (Ref: RPG-063). Thanks are due to Philip Ulrich for assistance with data collection.

## Abstract

Interpreting other peoples' actions relies on an understanding of their current mental states (e.g. beliefs, desires and intentions). In this paper, we distinguish between listeners' ability to infer others' perspectives and their explicit use of this knowledge to predict subsequent actions. In a visual-world study, two groups of participants (passive observers *vs*. active participants) watched short videos, depicting transfer events, where one character ('Jane') either held a true or false belief about an object's location. We tracked participants' eye-movements around the final visual scene, time-locked to related auditory descriptions (e.g. "Jane will look for the chocolates in the container on the left."). Results showed that active participants had already inferred the character's belief in the 1 second preview period prior to auditory onset, before it was possible to use this information to predict an outcome. Moreover, they used this inference to correctly anticipate reference to the object's initial location on false belief trials at the earliest possible point (i.e. from "Jane" onwards). In contrast, passive observers only showed evidence of a belief inference from the onset of "Jane", and did not show reliable use of this inference to predict Jane's behaviour on false belief trials until much later, when the location ("left/right") was auditorily available. These results show that active engagement in a task activates earlier inferences about others' perspectives, and drives immediate use of this information to anticipate others' actions, compared to passive observers, who are susceptible to influences from egocentric or reality biases. Finally, we review evidence that using other peoples' perspectives to predict their behaviour is more cognitively effortful than simply using one's own.

Keywords: Theory of Mind; false belief; eye-tracking; perspective use; cognitive effort

## Introduction

Until relatively recently, most researchers assumed that adults are fully capable 'mindreaders', having developed the necessary skills to pass even complex Theory of Mind (ToM) tasks between the ages of 2 and 7 years old (Wellman, Cross, & Watson, 2001). However, a growing body of research has emerged over the last couple of decades, demonstrating that ToM continues to develop through adolescence (e.g. Blakemore, 2008; Dumontheil, Küster, Apperly, & Blakemore, 2010), and that even healthy adults can suffer interference when considering other peoples' perspectives, showing a 'reality bias' (Mitchell, Robinson, Isaacs, & Nye, 1996) or 'curse of knowledge' (Birch & Bloom, 2007). The ability to see things from someone else's point of view is commonly referred to as perspective-taking. As such, perspective-taking relies heavily on ToM abilities to understand other peoples' mental states (which might be different from one's own), and how this might affect their knowledge, beliefs and actions. Perspective-taking has been examined along two key dimensions: one that assesses the *spatial* perspective of another person, including how this influences what objects they see; and another that assesses *mental* perspectives in terms of another person's beliefs, desires or intentions. Only mental perspectives give rise to the distinctive forms of explanation, prediction and justification of behaviour that have been extensively studied in the literature on "theory of mind" (e.g. Apperly, 2010; Doherty, 2008). The current paper targets mental perspectives, using an eye-tracking false belief study to examine how involvement in a task influences one's ability to infer others' beliefs and use this knowledge to predict subsequent behaviour.

Perspective-taking has become increasingly prominent in the context of language, particularly in relation to the timecourse with which people are able to interpret referentially ambiguous expressions (e.g. "the cup" when two cups are visible) based on a speaker's visual perspective. Initially, this research was conducted using simple command-based tasks, where participants followed the instructions of a confederate ‘director’ to select (“click on the…”) or move (“Move/pick up the…”) target objects around a visual display. Crucially, to examine the role of perspective in these tasks, the speaker’s knowledge of available objects could be limited by the presence of a physical barrier to the speaker’s (but not the listener’s) view. In this way, participants need to use perspective to infer the speaker’s knowledge, and predict reference to the mutually visible object. Much of this early research reported a delay in selecting the perspective-appropriate object when the speaker and listener held conflicting knowledge about the available objects, as well as overt errors of selecting a perspective-inappropriate referent (e.g. Barr & Keysar, 2002; Epley, Morewedge & Keysar, 2004; Keysar, & Barr, 2005; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003). Keysar and colleagues interpret this delay as an initial bias to relate information to one’s own egocentric perspective, suggesting that successful perspective use is cognitively challenging and operates only as a secondary and controlled correction mechanism.

In contrast, an earlier effect of perspective has been found in similar reference assignment tasks when linguistic markers, such as colour (e.g. red), (in)definite expressions (e.g. the/one of the), or scalar adjectives (e.g. big/small) were available to narrow down the relevant contrast set (e.g. Chambers et al., 2002; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008). Nevertheless, even here some interference from the privileged competitor was evident when looking at eye movement patterns over time, showing that listeners cannot completely ignore their egocentric perspective. Together, this research demonstrates that shared knowledge and perspective can have immediate effects on reference resolution when two equally fitting referents are available and strong constraints are provided in the discourse to narrow down the intended referent (Heller et al., 2008). Importantly, it argues against an automatic egocentric-first bias in interpreting perspective-sensitive language when the appropriate contrasts are available in the reference set.

Indeed, even earlier effects of perspective have been found within rich discourse contexts, where participants are actively engaged in a task with another person/character. For example, Brown-Schmidt, Gunlogson and Tanenhaus (2008) employed a similar reference assignment task to those described above, but here participants communicated with their partner through a question-answer discourse. In this study, conversational context explicitly established what the speaker did and did not know through the use of questions (e.g. “What’s above the cow?”), which facilitated the listener’s use of perspective, and eliminated egocentric biases (see also Brown-Schmidt, 2012). In a very different task, Hanna and Tanenhaus (2004) showed that participants can rapidly use perspective to modify the possible domains of reference when following instructions from a confederate playing the role of a chef (e.g. “could you put *the cake mix*…”). Here, participants restricted their referential search to objects in their own domain (as opposed to the chef’s domain), only when the chef’s hands were empty. When the chef’s hands were full, participants inferred that instructions could relate to objects in either domain (i.e. theirs or the chef’s), and widened their visual search accordingly. 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). As such, results from these reference assignment tasks demonstrate that the degree to which ToM is spontaneously employed in comprehension depends on the specific contributions of discourse context and knowledge about the speaker’s perspective.

Interestingly, a different timecourse of anticipatory bias has been reported across eye-tracking tasks where the participant is a passive observer to a narrated scenario (i.e. not engaged in an explicit task). Ferguson and colleagues have demonstrated that listeners can rapidly and accurately predict other peoples’ actions based on an explicit description of their (false) beliefs (e.g. “Bill will look for his watch on the table/chair”, Ferguson, Scheepers, &

Sanford, 2010), or conflicting desires (e.g. “John doesn’t want anyone to know that his favourite colour is pink”, Ferguson & Breheny, 2011). However, when a real-life conversational partner’s true/false belief was visually encoded (Ferguson & Breheny, 2012), listeners suffered a delay in successfully anticipating the speaker’s intended referents, despite showing clear and immediate sensitivity to the other person’s differing perspective. That is, participants were sensitive to the fact that the speaker’s knowledge differed from their own, but did not use this perspective to anticipate the appropriate location in this interactive false belief task. Taken together, results from these passive tasks suggest that even without an explicit reason to track another person’s mental state, we may spontaneously update our understanding of events to include inferences about other peoples’ perspectives, but we do not routinely use this knowledge to set up predictions about others’ actions. These contrasting results highlight a limitation in previous work, which has not typically distinguished between processes involved in inferring people’s perspectives and processes involved in using them.

In much previous work it is assumed that participants are already in possession of information about the other person’s mental state during the period of interest. For example, Ferguson and Breheny (2012) assumed that an inference about the speaker’s knowledge had already been made prior to the critical language input during which participants had to use this inference. Similarly, most tasks involving ambiguous reference assignment rely on participants having already computed the listeners’ visual perspective (i.e. which objects the speaker can and can not see) before instructions to select an object are uttered e.g. Brown-Schmidt et al., 2008; Hanna et al., 2003; Heller et al., 2008; Keysar et al., 2000). Moreover, previous research has demonstrated a distinction between early effects of perspective on referent anticipation, and effects from low-level cues emerging during integration (Barr, 2008; c.f. Brennan & Hanna, 2009; Brown-Schmidt & Hanna, 2011), but has not investigated how and when these early effects originate in participants’ perspective-taking. Yet in most interaction situations, particularly when another person's mental state has not been explicitly mentioned, people have to infer others' mental states before they can use them, and this is likely to contribute to processing costs (but see Horton, 2007).

Assistance with this problem may come from research outside of language comprehension, which has recently begun addressing the cognitive basis of ToM. As depicted in Figure 1, in a summary of the literature at the time, Apperly (2010) distinguished between processes involved in making a ToM inference, storing this information, and using it for predicting behaviour or making further ToM inferences. Distinguishing between these steps opens the way for asking when each step occurs, and whether each step responds in the same way to manipulations of motivation and cognitive effort.

### *Current Study*

Here, we sought to further understand the mechanisms involved in making and using inferences about others' mental states. Dynamic video stimuli depicted true and false belief scenarios, while participants' eye movements around the visual scene were recorded. Experimental videos began with two actors (introduced as Sarah and Jane) standing behind a table with a target object (e.g. a chocolate) in the centre and three possible containers on the left, middle and right side of the table. In the first part of the video, Sarah moves the object into one of the three boxes while Jane looks on. In the second part of the video, Sarah moves the object into one of the other boxes- either while Jane is watching or after she has left the scene (thus, Jane was not aware of the events that followed) therefore setting up true and false beliefs for Jane respectively. In order to examine the influence of involvement in a task, participants were split into two groups, and each group was given a different set of instructions. Passive observers were simply told to 'look and listen', while active participants were instructed to press one of three keys to select the container that would complete the sentence. Thus, active participants were given an explicit reason to keep track of the characters' perspectives, but passive observers were not. In addition, the relationship between the object and the relevant container was manipulated, such that on half the trials the first container used in the transfer event (i.e. the 'belief' box) predictably matched properties of the target object (e.g. a chocolate box), thus providing an additional semantic cue to support the belief inference in FB-predictable conditions. On the other half of the trials (unpredictable conditions), the target object predictably matched properties of the unused distractor container, and therefore semantic cues did not support either the reality or belief inference. Thus, the experiment crossed belief (true *vs.* false), task (passive *vs.* active) and predictability of the initial container (predictable *vs*. unpredictable). Participants' eye movements around the final visual scene (i.e. the three closed containers) were tracked, time-locked to a concurrent auditory description, of the form, "Jane will look for the chocolate in the container on the [left/ middle/ right]". In this way, we were able to track how referential expectations (as revealed by looking preferences) towards each of the possible containers emerged over time. In addition to analysing fixation patterns, this design allowed us to record the accuracy and timing of behavioural responses in the active participant group, and to monitor participants' pupil size in each experimental condition during the task. Research in Cognitive Psychology has provided clear evidence of a link between pupil diameter and cognitive effort (e.g. Kahneman & Beatty, 1966; Just & Carpenter, 1993).

One advantage of embedding the perspective reference at the end of a longer narrative context is that it allows listeners time to set up expectations about forthcoming referents, prior to the onset of disambiguating information (e.g. Altmann & Kamide, 1999). Crucially, this design allows us to distinguish the process of making an *inference* about other peoples' mental states (i.e. simply acknowledging that they hold a different level of knowledge or visual perspective than oneself) from actively *using* that knowledge to adjust expectations about that person's actions in a given situation. Figure 1 shows Apperly's (2010) distinction between ToM inference, storage and use, applied to the visual world of the current task. Firstly, note the general principle that ToM inferences can of course be executed either while directly observing behaviour (e.g., of people in our videos), or at a later point, from a memory record of this observed behaviour or other information. Next, from the figure it is clear that prior to his/her belief being inferred, the character's false belief can have no influence on patterns of fixation. Once their false belief has been inferred, merely storing this information may influence fixation patterns by raising the salience of the container where the character falsely believes the object is located. However, only the use of this information- ie. during comprehension of the stimulus sentence- can produce predictive fixations that are biased towards the container where the character falsely believes the object is located. Below we make clear how this analysis yields predictions that vary in informative ways across the conditions of our study.

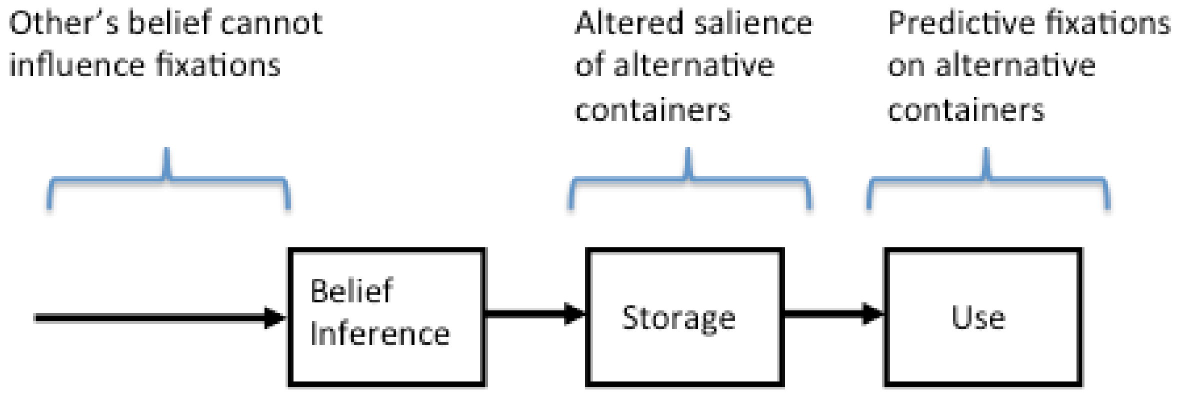
Other's belief cannot influence fixations
Altered salience of alternative containers
Predictive fixations on alternative containers
Belief Inference
Storage
Use

Figure 1:

An illustration of how perspective inference and use might be organized within a model of ToM, showing how each stage is reflected by eye movements within the current false belief paradigm.

In the current study, the additional processing time available during the discourse (as well as the 1000 msec preview period prior to discourse onset) will allow us to track the trajectory of perspective biases at four key points during the trial (corresponding to the image onset, and auditory onsets of the character “Jane”, the object “chocolate”, and the location “left”). The design gives us the potential to distinguish between effects due to participants having inferred a belief versus effects of using this information to predict the behaviour described in the stimulus sentence.

The literature overview above highlights one important mechanism that might influence the degree to which other peoples’ perspectives can be inferred and/or used online: involvement in the task. Different effects have been reported across different tasks that either require participants to follow a speaker’s instructions, engage in an interactive question-answer discourse or simply attend to a passive narrative/visual scenario, therefore it is vital now to understand how these effects might be modulated by the specific task constraints (see Salverda, Brown, & Tananhaus, 2010). Indeed, many of the studies described previously explicitly set out to enhance engagement with a communication partner (real or on-screen) by increasing task involvement (e.g. Brown-Schmidt et al., 2008; Hanna & Tanenhaus, 2004). However, since these studies did not manipulate involvement (and lack-of) within a single paradigm, it is unclear whether involvement, or other differences between studies (e.g. design, task, or analysis), is responsible for the enhanced perspective use that they report. Specifically, the individual experiments run in previous work have differed according to numerous factors including: (i) whether participants need to take the perspective of a real (live) person or an on-screen avatar, (ii) whether perspective-taking is needed to fulfill an explicit instruction or to support comprehension of a narrative, (iii) whether there is ambiguity regarding what the intended target object is, or its location, (iv) whether objects are hidden from the other person's view/reach, or simply not desired by the other person, and (v) the statistical procedures/comparisons adopted to analyse eye movements as a measure of perspective inference and use. Moreover, most previous eye-tracking studies that have examined perspective-taking in language comprehension have employed a version of the referential communication task (described above). While this task provides valuable information on how listeners can use a speaker's visual access to potential referents to resolve referential ambiguities, it does not guarantee that participants tap into another person's mental perspective- their beliefs or desires- to examine how these mental states might influence perspective inferences and use (but see Mozuraitis, Chambers, & Daneman, 2014).

In sum, this paper aims to gain a fuller understanding of perspective-taking during communication by examining the role of the self in that task, and explicitly manipulating task involvement in our experimental design. Psycholinguistic research suggests that people suffer impairments in perspective use when they are passive overhearers (e.g. Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992), or when interpreting non-interactive dialogue (Brown-Schmidt, 2009). Similarly, in a language production study, participants were more likely to consider their partner's perspective when their goal was to make a request compared to when their goal was to inform (Yoon, Koh, & Brown-Schmidt, 2012, Experiment 2). The current study will extend this work by systematically manipulating participants' involvement in a single true/false belief paradigm, which eliminates the influence of other differences between studies (as described above), and allows us to isolate the different cognitive processes that are recruited when inferring and using beliefs. Note that using this paradigm, biases based on the visual presence of the target object (as typically seen in the referential communication task) were eliminated by hiding the target object inside one of three opaque containers. The semantic association manipulation between objects and containers (similar to Samson, Apperly, & Humphreys, 2007) allows us to examine whether the cognitive effort involved in making the ToM inference is reduced by the availability of consistent appearance-based cues. This proposal relates to previous studies that have demonstrated improved use of perspective when the perceptual properties of referents are a good match with the linguistic descriptor (e.g. Hanna et al., 2003). The question is, does this semantic cue result in earlier or stronger anticipation of the perspective-appropriate container?

### *Predictions*

To appreciate how we might detect effects of belief inference[1] versus use, let us first imagine the looking pattern if participants are entirely ignorant of the target character's belief and interpret the stimulus sentence only according to information from their own perspective (i.e. they have neither inferred the character's perspective or used it to predict their described behaviour). Before the sentence begins, the one clear expectation must be that patterns of fixation between the containers will be the same in the true belief and false belief conditions. This is because participants are ignorant of the character's belief and so there is no reason to suppose that the information they are holding in memory differs systematically in a way that might bias their attention to the containers. As the sentence progresses one possibility is that participants attend passively, and only preferentially fixate one container after that container is described. Alternatively, they may predict the container on which the character will act, but since they are ignorant of the character's belief, their fixations will always be biased in favour of the container that they know contains the object, in both the true and false belief conditions.

Now let us imagine the pattern if participants have inferred the character's belief in advance of the stimulus sentence but do not use it to predict the character's behaviour. In the

[1] Note that for ease of presentation we use the term 'inference' throughout this paper for comparison with perspective use, however, given the long lead-in time used here it is unlikely that participants are making the inference for the first time during the period of analysis, but rather they are accessing the stored representation of this inference for prediction.

true belief condition such a participant should have a memory representation of the object’s true location (possibly alongside the object’s initial location), which is consistent with the character’s belief that the object is in that true location. In contrast, in the false belief condition the participant should have a memory representation of the object’s true location (again, alongside the object’s initial location), which conflicts with the character’s belief that it is in the initial location. Since it is known that memory representations of items in a visual world can drive fixations to their locations (e.g. Altmann, 2011; Altmann & Kamide, 2009; Richardson & Spivey, 2000), we might expect this participant to show different patterns of fixation to the containers in the true belief and false belief conditions. These different patterns reflect the fact that in the true belief condition listeners have reason to fixate the container that actually contains the object, since it features in both their own memory record of the object’s location and their record of where the character thinks the object is located. In contrast, in the false belief condition listeners have reason to fixate both the container that actually contains the object, since it features in their memory record of the object’s location, and the container where the character thinks the object is located, since they also have a record of the character’s false belief. However, since (by hypothesis) they are only storing the character’s false belief, and not using it to predict the character’s behaviour, they will have no reason to attend preferentially to the location where the character falsely believes the object is located.

Finally, let us imagine the pattern if participants have inferred the character’s belief and are using this information to predict the action described in the stimulus sentence. In the true belief condition prediction would be evidenced in a fixation pattern that was biased in favour of the container holding the object. Note, however, that this pattern is indistinguishable from the predictions made if participants are entirely ignorant of the character’s perspective and predict egocentrically, because the character’s true belief coincides with the participant’s own perspective. In the false belief condition use of the character’s perspective to direct predictions about their behaviour would be evidenced in a fixation pattern that was biased in favour of the empty container that the character incorrectly thought contained the object (the initial location). Note that the false belief condition is the critical condition that demonstrates a non-egocentric prediction from the perspective of the character.

In relation to the timing of these effects, the most informative differential pattern for belief inference should be in the period before the sentence begins (i.e. the 1000 msec preview), because later effects of having inferred a belief could be combined with effects due to predictions about the unfolding stimulus sentence. However, the earliest point that perspective use could be activated is the auditory onset of the character’s name, though evidence that the participant has made the inference about differing perspectives should be apparent prior to this point.

In sum, evidence of participants having made a belief inference would come from *any* differential pattern of fixations to the containers between the true belief and false belief conditions. This is likely to be driven by a reduced bias to fixate the object’s real (final) location on false belief trials, irrespective of whether there is an overall bias in favour of the container holding the object, another container or no container. In contrast, evidence that a belief inference was being used to predict the outcome of the action described in the sentence would come only from a significant bias in favour of whichever container the character believed to hold the object, with this evidence being decisive in the false belief condition. That is, in the false belief condition successful use of the perspective inference would lead to a clear preference to fixate the object’s initial location rather than its final location (since the story character is not aware of this). Notably, inferences about the character’s perspective can be made at any point from the video sequence onwards, while participants can only use this knowledge to predict future events once they hear whose perspective to adopt in the auditory narrative (i.e. Jane or Sarah). We will explore the time-course of these effects as events unfold. Finally, by manipulating whether participants were instructed to be passive listeners or active predictors of the stimulus sentence we aim to assess the degree to which either belief inference or belief use was spontaneous or elicited by specific task requirements. Predictions on the exact timing of these effects will be detailed below.

## Method

### *Participants*

A total of eighty native English speakers from the University of Kent were paid to participate in the study (50% female). Of these, half (N=40) took part in the experiment as ‘passive observers’ ($M_{\text{age}}$ = 21.89, $SD_{\text{age}}$ = 4.84), and the other half were ‘active participants’ ($M_{\text{age}}$ = 22.84, $SD_{\text{age}}$ = 5.81).

### *Stimuli and Design*

Twenty four sets of experimental videos and pictures were paired with an auditory description in one of four conditions. Video clips were recorded in a single session involving two female ‘actors’ (distinguished by white/blacks tops) and edited using Adobe Premier. All visual images were presented on a 17 inch colour monitor in 1024 x 768 pixels resolution. Auditory sentences were recorded in a single session from a female native British English speaker who used a neutral intonation[2]. The auditory files were presented as 44.1 KHz stereo sound clips via headphones connected to the eye-tracker PC. The temporal onsets and offsets of critical words were hand-coded with millisecond resolution using the WavePad sound-editing package.

Four different video scenarios depicted a series of transfer events that set up the relevant contexts (see Figure 2). All experimental videos began with the two actors

[2] Contact the authors for example audio recordings.

(introduced as Sarah and Jane) standing behind a table with a target object in the centre and three possible containers on the left, middle and right side of the table. In the first part of the video, Sarah moved the target object into one of the three containers while Jane looked on. To set up the two perspective states, a second part of the video depicted Sarah moving this target object into one of the other containers. Importantly, Jane was either present for this second transfer event (meaning that she held a true belief about the object's location), or had left the scene after the first transfer events, and was absent when the object was moved (meaning that Jane held a false belief about the object's location). All videos ended with Sarah standing alone behind the table with the three closed containers. To set up the two predictability states, the target object was manipulated so that it was predictably related to either the first container used in the transfer sequence (i.e. the 'belief' box) or to the unused distracter box[3], thus providing an additional semantic cue to facilitate the belief inference in some conditions (Yee & Sedivy, 2006). Predictable pairings were established through typical object-container relations (e.g. chocolates and a chocolate box, video camera and camera case), as in Samson et al. (2007). Twelve everyday object-container pairings were used twice across experimental trials; once in a predictable pairing (i.e. object matched initial container) and once in an unpredictable pairing (object matched distractor container), with equal numbers of occurrences in true- and false-belief trials. Subsequent pictures depicted the final state from each of these scenarios (i.e. Sarah with the three closed boxes), and were created by extracting the final frame from each video clip. Systematic viewing strategies were prevented by counterbalancing the spatial arrangement of the objects across items.

Sound files consisted of a single pre-recorded sentence, of the form, "Jane will look for the [object] in the container on the [left/ middle/ right]"[4]. In order to examine the influence

[3] Note that target containers remained the same across all four versions of a trial.

[4] Note that this verbal description of the object's location prompts listeners to assign left/right according to their own view, however this was the same across all items and conditions.

of involvement in a task, participants were split into two groups, and each group was given a different set of instructions. Passive observers were simply told to ‘look and listen’, while active participants were instructed to press one of three keys to select the container that would complete the sentence. Thus, the experiment employed a 2 x 2 x 2 mixed design, with task (passive observer *vs*. active participant) as the between-subjects factor, and belief (true belief *vs*. false belief) and predictability of the initial container (predictable *vs*. unpredictable) as the repeated-measures factors.

One version of each item was assigned to one of four presentation lists, with each list containing twenty-four unique experimental items, six in each of the four conditions. Participants were randomly assigned to one of these four lists, which ensured that across these lists (and therefore across participants) each video was seen in all four conditions. By using this fully counterbalanced design, we can be confident that any differences between conditions cannot be due to natural differences in the video stimuli themselves. In addition, thirty filler items were interspersed randomly among the twenty-four experimental trials to create a single random order. All fillers depicted similar transfer events to those used in the experimental trials and were included to disguise the purpose of the study and to prompt participants to consider the various characters’ perspectives over the course of the experiment. Of these fillers, fifteen depicted events where the target object was simply replaced into the same container in the second part of the video. This was to ensure that the object did not always move location in the second part of the video, which may have caused participants to make predictions based on learned patterns. As in the experimental trials, Jane watched the entire transfer sequence in half the filler trials, and left the scene part-way through the transfer sequence in the other half. Additionally, the auditory descriptions were manipulated so that participants had to either infer events according to Jane (as in experimental trials, N=8), Sarah (the fully informed character, N=5), a stranger (e.g. “A stranger will look for the…”, N=5), or in reality (e.g. “It’s true that the X is in…”, N=12). This manipulation in the filler items ensured that on experimental trials, which all referred to Jane’s perspective, participants did not know whose name was going to be mentioned until they heard “Jane”. Binary comprehension questions that tested participants’ memory of events followed half of the experimental and half of the filler trials (see Appendix A), and all participants responded to these questions. All participants scored at or above 90% accuracy on these comprehension questions.

### *Procedure*

Participants sat in front of a colour monitor while eye movements were recorded from the right eye using an EyeLink 1000 eye-tracker (viewing was binocular), running at 1000 Hz sampling rate. Distance from the screen was kept at a constant 60 cm for all participants using a fixed chin rest. Participants in the ‘passive observer’ group were given the following instruction: “In this experiment you will watch short videos, each of which will be followed by a still frame from that video and a spoken description of events. Your task is simply to watch and listen and respond to the comprehension questions when prompted”. In contrast, participants in the ‘active participant’ group were told: “In this experiment you will watch short videos, each of which will be followed by a still frame from that video and a spoken description of events. Your task is to predict, as quickly and accurately as possible, which container will complete the spoken sentence. You should indicate your answer using the keyboard”. Thus, while the active participants were required to make an explicit key-press response based on their perspective inference, the passive observers were not.

The experiment was controlled using Experiment Builder software and the experimental procedure is illustrated in Figure 2. At the start of the experiment, participants were introduced by name (Sarah and Jane) to the two characters that would be featured in the videos, and were instructed to remember the name-person pairings. Each trial began with a centrally-located drift correction point. Following successful fixation on this point, a video depicting a transfer event was initiated, as described above. Video clips lasted on average 28 seconds (range = 16 to 53 seconds) and were followed by a blank screen for 500 msec. Next, the corresponding picture was presented to the participant, along with the relevant auditory target sentence. Picture onset preceded audio onset by 1000 msec. This picture stayed onscreen for a total of 7000 msec, with the corresponding sentence typically ending 1-2 seconds before the end of the trial. A 500 msec blank screen separated trials. Active participants were able to select the relevant container until the end of the trial.

At the beginning of the experiment, and once every ten trials thereafter, the eye-tracker was calibrated and validated against nine fixation points, using the standard EyeLink calibration procedure. This procedure took about half a minute and an entire session lasted for about 45 minutes.

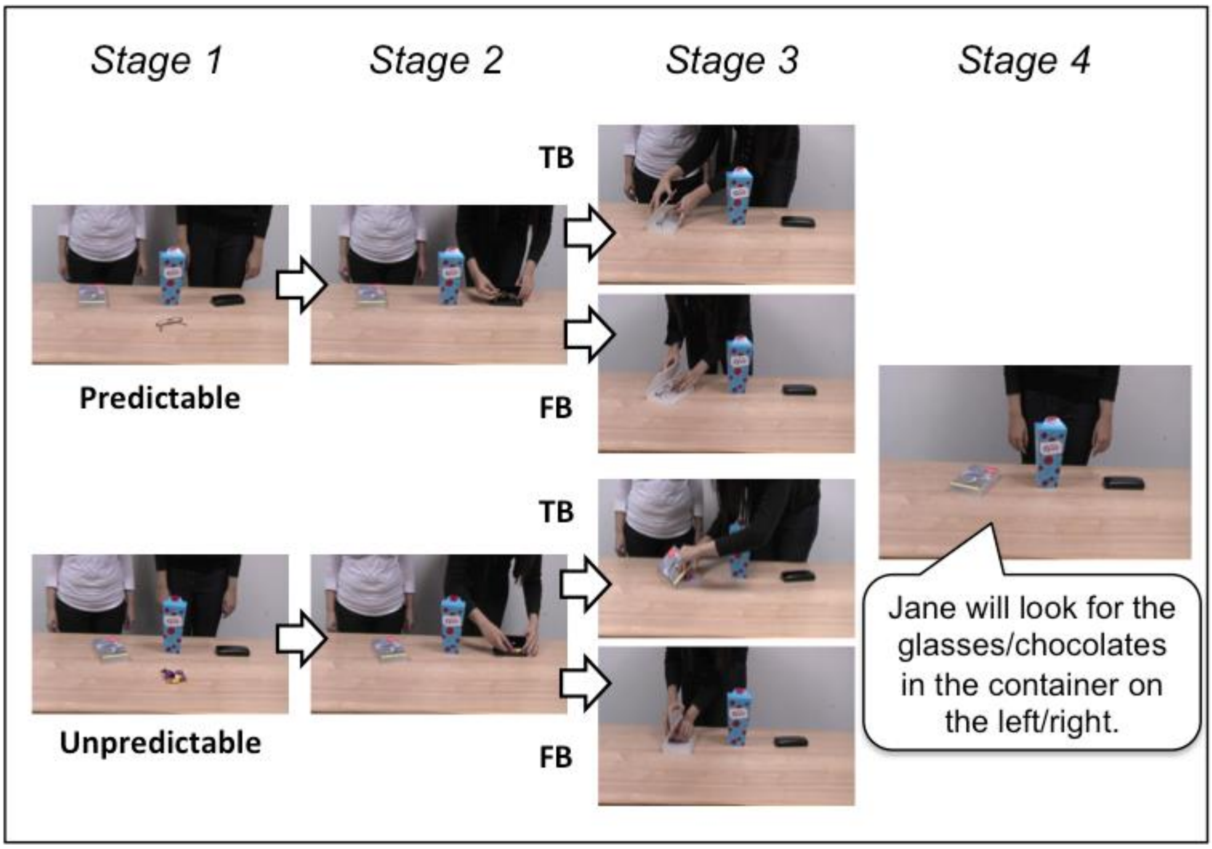
Stage 1
Stage 2
Stage 3
Stage 4
TB
FB
TB
FB
Predictable
Unpredictable
Jane will look for the glasses/chocolates in the container on the left/right.

Figure 2:

Schematic trial sequence of visual displays presented to participants. Stage 1 depicts the 'start state'. In stage 2, a video showed Sarah moving the target object into one of three containers. Stage 3 showed Sarah move the object into one of the other containers, either while Jane was still present (true belief, TB) or after Jane had left the scene (false belief, FB). Finally, Stage 4 shows the 'final state' picture that participants saw while they listened to the audio sentence.

## Results & Discussion

### *Eye-tracking data processing and analysis*

Eye movements that were initiated while the target image was onscreen were processed according to the relevant picture and sound onsets on a trial-by-trial basis. The spatial coordinates of fixations (in pixels) were mapped onto the appropriate regions of analysis, corresponding to the container locations (left, middle and right) for each image. If a fixation was located within 20 pixels around a container’s perimeter, it was coded as belonging to that object; otherwise, it was coded as background.

To visualise the data, visual preferences to the final location (i.e. the box that actually contains the target object) and the initial location (i.e. the first box involved in the transfer sequence - Jane’s belief on FB trials) were plotted by calculating a location advantage score as a function of time (i.e. the probability of fixating the final location *minus* probability of fixating the initial location). This measure is symmetrical around zero such that higher proportions of fixations on the final location result in a positive score, whereas higher proportions of fixations on the initial location result in a negative score. The resulting plots are shown separately for passive observers (Figure 3) and active participants (Figure 4) for ease of exposition, and illustrate when visual interpretations became biased to either container as the auditory sentence progressed. Note that eye movements and auditory input have been resynschronized according to individual word onsets (see Altmann & Kamide, 2009), and as such represent more accurate plots of evolving visual biases around the scene.

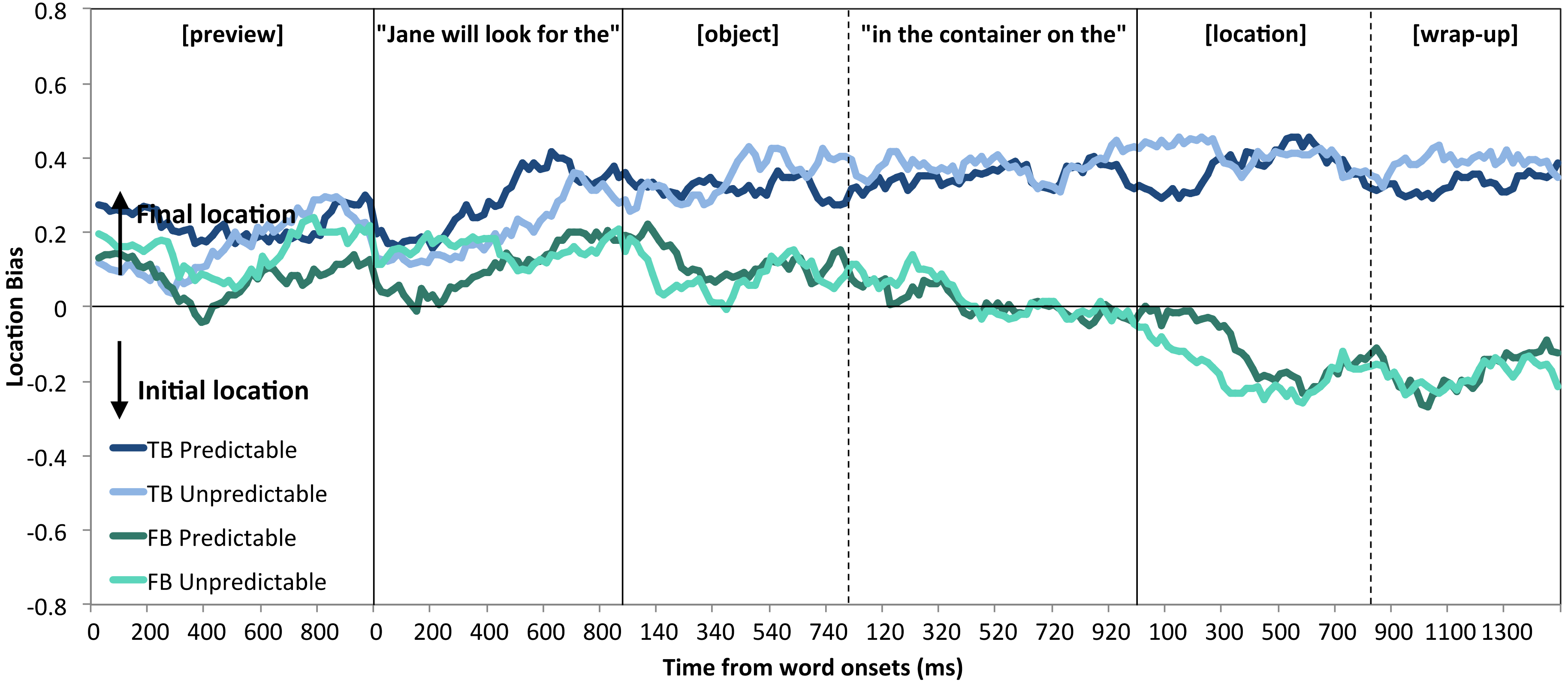

Figure 3:

The average location advantage scores for each condition in the ‘passive observers’ task group. Note that the dashed vertical lines indicate the absolute onsets and average offsets of words in the target sentence, as labelled.

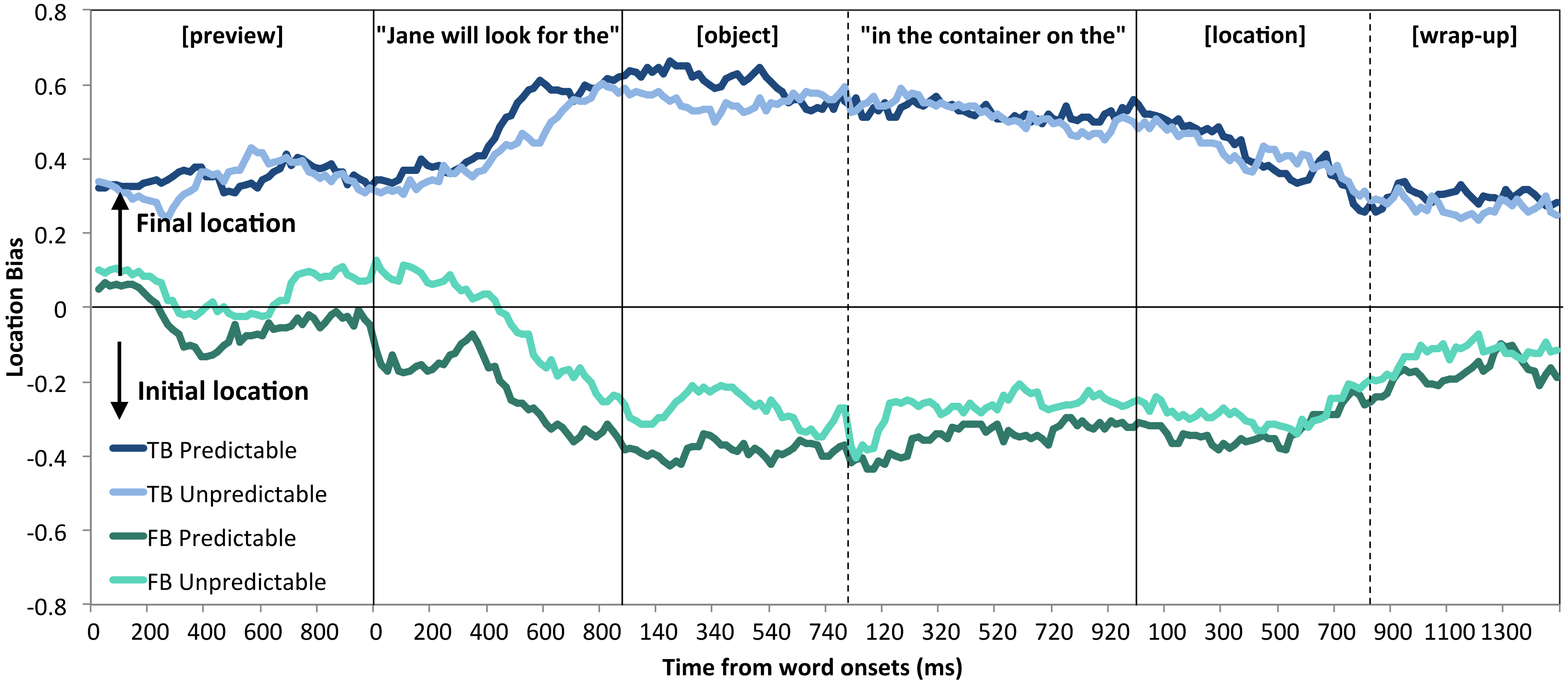

Figure 4:

The average location advantage scores for each condition in the ‘active participants’ task group. Note that the dashed vertical lines indicate the absolute onsets and average offsets of words in the target sentence, as labelled.

For statistical analysis, we report the location of individual fixations[5] at four key points during the trial, synchronized to the absolute onsets and offsets of relevant events in the target sentence on a by-trial basis. The dependent measure for each time point compared the probability of making at least one fixation on the final location with the probability of making at least one fixation on the initial location. Fixations were coded as a binary response for each time window, and probabilities were calculated relative to the total number of fixations within that time window (including fixations on the distractor location and background). A 'final location advantage score' was calculated separately for participants and items as the probability of fixating the final location minus the probability of fixating the initial location ((P(final location))-(P(initial location))).

The first measure examines the location of the first fixation that was launched during the preview (i.e. immediately after image onset, but before audio onset). This measure allows us to determine whether participants have already computed the belief inference while observing the actors in the video. If listeners have already inferred Jane's belief, then we should see a difference in the visual biases between TB and FB conditions immediately upon image onset. The reason for this is that although participants are likely to consider their memory of the object's final (real) location and the character's belief about their location on all trials, these representations are only likely to compete with each other on FB trials, when their own memory is of the object in its final location whereas the character's false belief is that the object is in the initial location (see Altmann & Kamide, 2009). Examining the presence of this difference between passive and active participant groups will establish whether participants only engage in this spontaneous perspective-taking under specific task demands. None of the participants are expected to use an inference about Jane's belief to direct expectations to the initial location on this first fixation measure, since they have not

[5] Fixations were defined according to Eyelink's Cognitive configuration.

received auditory information on whose perspective to adopt (recall that filler trials tapped into different perspectives, i.e. Sarah, a stranger, or in reality). The second measure examines the probability of making at least one eye movement towards the final/initial locations from the auditory onset of “Jane” but before the onset of the [Object] (e.g. “Jane will look for the”). This period is the first time that participants could use their knowledge of Jane’s belief to explicitly predict events according to her perspective. Thus, we predicted that if listeners have inferred Jane’s FB, and if they are using this knowledge, then we should observe anticipatory attention to the belief-appropriate container. In other words, we would expect participants to show a clear preference for the initial location over the final location on FB trials. Alternatively, if listeners have not inferred Jane’s FB or if they experience strong conflict between the reality and belief representations of the object, this might be manifest in either a bias to fixate the final location (suggesting a pull of reality or egocentric bias), or simply no bias to either location. The third measure examines the probability of fixating the final/initial location between the [Object] onset and [Location] onset (e.g. “chocolates in the container on the”). This period is the last point at which participants can launch anticipatory eye movements towards relevant containers. These predicted patterns are the same as for the second measurement period but with a higher likelihood that they would be observed: since the target object is auditorily mentioned here, it is considered an important point for participants to use knowledge about Jane’s beliefs to explicitly anticipate her actions, as the mention of the object should direct the eyes to the relevant location (Cooper, 1974). Finally, the fourth measure examines the probability of directing at least one eye movement towards the final/initial location during the [Location] (e.g. “left”). This measure examines effects during integration of the target location, with the prediction that all participants will rapidly accommodate the described location. Differences between conditions on this measure would indicate integration difficulties, perhaps due to mismatching expectations about events. The average durations for these time-regions are shown in Table 1, and the average probabilities of fixating the final location, initial location and distractor at each time point are displayed in Table 2. Note that for ease of interpretation, the values in Table 2 collapse across predictability since this factor did not produce any significant effects in the statistical analyses of eye movement data (see details below).

| | "Jane will look for the" | [Object] | [location] |
|---|---|---|---|
| TB Predictable | 834 | 1802 | 686 |
| TB Unpredictable | 857 | 1802 | 674 |
| FB Predictable | 848 | 1846 | 703 |
| FB Unpredictable | 839 | 1799 | 691 |

Table 1:

Average time-region durations for each condition (timings in ms).

| | Passive task | | Active task | |
|---|---|---|---|---|
| | True Belief | False Belief | True Belief | False Belief |
| [Preview] *p(first fixation)* | | | | |
| Final location | 0.34 (.19) | 0.32 (.19) | 0.47 (.23) | 0.34 (.24) |
| Initial location | 0.16 (.15) | 0.17 (.18) | 0.15 (.17) | 0.28 (.22) |
| Distractor box | 0.21 (.16) | 0.21 (.16) | 0.15 (.16) | 0.15 (.14) |
| "Jane" *p(fixation)* | | | | |
| Final location | 0.42 (.21) | 0.31 (.18) | 0.55 (.2) | 0.26 (.19) |
| Initial location | 0.20 (.17) | 0.24 (.17) | 0.13 (.14) | 0.47 (.21) |
| Distractor box | 0.18 (.15) | 0.20 (.16) | 0.14 (.12) | 0.10 (.12) |
| [Object] *p(fixation)* | | | | |
| Final location | 0.46 (.19) | 0.35 (.21) | 0.59 (.22) | 0.18 (.16) |
| Initial location | 0.20 (13) | 0.29 (.18) | 0.14 (.15) | 0.5 (.22) |
| Distractor box | 0.14 (13) | 0.16 (.14) | 0.09 (.1) | 0.12 (.14) |
| [Location] *p(fixation)* | | | | |
| Final location | 0.47 (.23) | 0.24 (.2) | 0.44 (.24) | 0.22 (.16) |
| Initial location | 0.16 (.14) | 0.42 (.21) | 0.17 (.15) | 0.40 (.25) |
| Distractor box | 0.17 (.15) | 0.15 (.14) | 0.21 (.15) | 0.18 (.16) |

Table 2:

Mean proportions of fixations (by participants) in each condition to the final location, initial location and distractor at each time point. Standard deviations are shown in parentheses.

Statistical analyses were carried out separately for each fixation measure, using the lmer function in the lme4 package (Bates & Maechler, 2010) using R (version 3.0.1, R Development Core Team, 2013). Participants and items were entered into separate models as random effects, and task (Passive *vs.* Active), belief (FB *vs*. TB) and predictability (Unpredictable *vs*. Predictable) as fixed effects. The final location advantage score was used as the dependent variable, which is described above as participants' bias to fixate the initial/final object location at four key points (i.e. preview period, "Jane", "Object", and "left/middle/right"). The two levels of each fixed factor were coded using contrasts (-.5 *vs.* +.5, respectively). Models using the fixed effects as "maximal" random slopes were run (Barr, Levy, Scheepers, & Tily, 2013), but Chi-squared tests showed that including these did not improve model fit on any fixation measure. Where post-hoc analyses were required to follow up on significant interactions, models were re-leveled to each one of the IV levels of interest and the model updated, using the mcposthoc function in R. Statistical effects on the intercept test whether the fixation bias is significantly different from zero (i.e. biased to the initial or final location) at a given time-point. For all tests a significance level of 5% was used.

Table 3 shows the fixed and random effects for the model adopted within each time window. Note that the fixed effect of predictability did not emerge as a significant effect in any of the full models, and model fitting using REML (Newman, Tremblay, Nichols, Neville, & Ullman, 2012) showed that including this variable did not significantly improve the model fit. Additionally, in the final [Location] measure, neither predictability nor task significantly improved the model fit, and both these variables were removed during back-fitting. Therefore, for ease of understanding Table 3 reports effects from the optimal back-fitted model, including only those factors that significantly contributed to the model. Results from the full model, including maximal random slopes, can be seen in Appendix B.

| | [Preview] *p(first fixation)* | | "Jane" *p(fixation)* | | [Object] *p(fixation)* | | [Location] *p(fixation)* | |
|---|---|---|---|---|---|---|---|---|
| | Estimate (SE) | t-value | Estimate (SE) | t-value | Estimate (SE) | t-value | Estimate (SE) | t-value |
| By participants model fit (LogLik) | -105.6 | | -84.65 | | -86.41 | | -115.7 | |
| Fixed effects (by participants): | | | | | | | | |
| Intercept | 0.17 (.02) | 9.45*** | 0.15 (.02) | 8.43*** | 0.11 (.02) | 6.09*** | 0.05 (.02) | 2.58** |
| Task | 0.02 (.04) | 0.54 | -0.03 (.03) | -0.9 | -0.11 (.04) | -3.30*** | | |
| Belief | 0.15 (.04) | 4.01*** | 0.39 (.03) | 11.29*** | 0.48 (.35) | 13.81*** | 0.5 (.04) | 13.07*** |
| Task:Belief | 0.24 (.07) | 3.20** | 0.48 (.07) | 6.95*** | 0.56 (.07) | 8.07*** | | |
| Random effects: | | | | | | | | |
| Participant (Var) | <0.001 | | <0.001 | | <0.001 | | <0.001 | |
| Residual (Var) | 0.11 | | 0.1 | | 0.1 | | 0.12 | |
| By items model fit (LogLik) | -39.6 | | -22.16 | | -32.55 | | -36.41 | |
| Fixed effects (by items): | | | | | | | | |
| Intercept | 0.17 (.04) | 4.38*** | 0.13 (.03) | 3.94*** | 0.12 (.03) | 4.87*** | 0.06 (.03) | 1.78* |
| Task | 0.02 (.04) | 0.62 | -0.05 (.03) | -1.4 | -0.13 (.04) | -3.38*** | | |
| Belief | 0.14 (.04) | 3.83*** | 0.39 (.03) | 11.15*** | 0.5 (.04) | 12.85*** | 0.48 (.04) | 12.30*** |
| Task:Belief | 0.23 (.08) | 3.00** | 0.47 (.07) | 6.79*** | 0.59 (.08) | 7.60*** | | |
| Random effects: | | | | | | | | |
| Item (Var) | 0.03 | | 0.02 | | 0.01 | | 0.02 | |
| Residual (Var) | 0.07 | | 0.06 | | 0.07 | | 0.07 | |

Table 3:

Parameter estimates by participants and items at each time point. Significance tests for fixed effects were estimated using MCMC sampling, where * = *p*<.05; ** = *p*<.01; *** = *p*<.001.

Analyses of the first fixation during the preview period revealed a significant effect of belief. As can be seen in Figures 3 and 4, this effect was due to a significantly stronger bias to fixate the final location when Jane witnessed the second transfer event (TB trials; Intercept *Est.*$_1$ = .25, SE$_1$ = .03, $t_1$ = 9.55; *Est.*$_2$ = .25, SE$_2$ = .04, $t_2$ = 5.6, *p*s < .001), compared to when Jane was ignorant to the second transfer event (FB trials; Intercept *Est.*$_1$ = .1, SE$_1$ = .03, $t_1$ = 3.86; *Est.*$_2$ = .1, SE$_2$ = .04, $t_2$ = 2.32, *p*s < .05). Interestingly, belief also interacted with task on this measure. Analysis of the effects of belief in each group revealed that active participants employed significantly different viewing strategies in TB *vs.* FB trials (*Est.*$_1$ = .27, SE$_1$ = .05, $t_1$ = 5.11; *Est.*$_2$ = .26, SE$_2$ = .05, $t_2$ = 4.84, *p*s < .001), but passive observers did not (*Est.* = .03, SE = .05, *ts* < 1). Specifically, analyses of the intercept term in the active group for each level of belief (i.e. comparing the grand mean bias in each condition to zero) showed that participants were significantly biased to launch a first fixation to the final location when they shared Jane's TB about the object's location (*Est.*$_1$ = .32, SE$_1$ = .04, $t_1$ = 7.79; *Est.*$_2$ = .31, SE$_2$ = .06, $t_2$ = 5.2, *p*s < .001), but showed no significant bias to either container when their knowledge of the object's real location conflicted with Jane's FB (*Est.*$_1$ = .05, SE$_1$ = .04, $t_1$ = 1.26; *Est.*$_2$ = .06, SE$_2$ = .06, $t_2$ = .97, *p*s > .1).

Fixations from the onset of "Jane" also showed a significant effect of belief, reflecting a significant bias to fixate the final location on TB trials (Intercept *Est.*$_1$ = .34, SE$_1$ = .02, $t_1$ = 13.9; *Est.*$_2$ = .33, SE$_2$ = .04, $t_2$ = 8.65, *p*s < .001), and a marginal bias to fixate the initial location on FB trials (Intercept *Est.*$_1$ = -.05, SE$_1$ = .02, $t_1$ = -2.02; *Est.*$_2$ = -.06, SE$_2$ = .04, $t_2$ = -1.68, *p*s < .1). Once again, this effect of belief was qualified by a significant interaction with task, reflecting a significant effect of belief among both active participants (*Est.*$_1$ = .63, SE$_1$ = .05, $t_1$ = 12.86; *Est.*$_2$ = .63, SE$_2$ = .05, $t_2$ = -12.58, *p*s < .001), and passive observers (*Est.*$_1$ = .15, SE$_1$ = .05, $t_1$ = 3.07; *Est.*$_2$ = .15, SE$_2$ = .05, $t_2$ = 3.06, *p*s < .005). Analyses of the intercept term in the active group for each level of belief showed that participants were more likely to fixate the final location when Jane held a TB about the object's location (*Est.*$_1$ = .44, SE$_1$ = .04, $t_1$ = 12.27; *Est.*$_2$ = .42, SE$_2$ = .05, $t_2$ = 9.17, *p*s < .001), but were more likely to fixate the initial location when Jane held a FB about the object's location (*Est.*$_1$ = -.18, SE$_1$ = .04, $t_1$ = -5.13; *Est.*$_2$ = -.21, SE$_2$ = .05, $t_2$ = -4.56, *p*s < .001). In contrast, passive observers showed a general preference to fixate the final location in both belief conditions, but this bias was stronger in the TB condition (*Est.*$_1$ = .24, SE$_1$ = .03, $t_1$ = 7.16; *Est.*$_2$ = .23, SE$_2$ = .05, $t_2$ = 4.76, *p*s < .001), compared to the FB condition (where the bias to the final location was statistically marginal) (*Est.*$_1$ = .09, SE$_1$ = .03, $t_1$ = 2.61; *Est.*$_2$ = .08, SE$_2$ = .05, $t_2$ = 1.64, *p*s < .1).

Analysis of the third measure, examining the probability of directing an eye movement to the final/initial location between the [Object] onset and [Location] onset, showed significant effects of belief and task, and an interaction between these fixed effects. The effect of belief reflects a significant bias to fixate the final location when Jane held a TB about the object's location (Intercept *Est.*$_1$ = .34, SE$_1$ = .02, $t_1$ = 14.07; *Est.*$_2$ = .37, SE$_2$ = .03, $t_2$ = 11.63, *p*s < .001), but a significant bias to fixate the initial location when Jane held a FB about the object's location (Intercept *Est.*$_1$ = -.13, SE$_1$ = .02, $t_1$ = -5.46; *Est.*$_2$ = -.12, SE$_2$ = .03, $t_2$ = -3.86, *p*s < .001). The significant effect of task revealed that participants experienced a stronger overall bias to the final location when they were passive observers (Intercept *Est.*$_1$ = .16, SE$_1$ = .02, $t_1$ = 6.64; *Est.*$_2$ = .19, SE$_2$ = .03, $t_2$ = 5.92, *p*s < .001), compared to when they were active participants in the task (Intercept *Est.*$_1$ = .05, SE$_1$ = .02, $t_1$ = 1.97; *Est.*$_2$ = .06, SE$_2$ = .03, $t_2$ = 1.85, *p*s < .06). Moreover, the significant interaction between these two fixed effects showed that from the [Object] onset, participants experienced effects of belief in both the active participants (*Est.*$_1$ = .76, SE$_1$ = .05, $t_1$ = 15.47; *Est.*$_2$ = .79, SE$_2$ = .05, $t_2$ = 14.47, *p*s < .001) and passive observers (*Est.*$_1$ = .2, SE$_1$ = .05, $t_1$ = 4.06; *Est.*$_2$ = .2, SE$_2$ = .05, $t_2$ = 3.72, *p*s < .001) groups. Analyses of the intercept term for each level of belief in each group revealed that active participants were significantly more likely to make a fixation to the final location when they shared Jane's TB about the object's location (*Est.*$_1$ = .43, SE$_1$ = .04, $t_1$ = 11.77; *Est.*$_2$ = .45, SE$_2$ = .04, $t_2$ = 11.22, *p*s < .001), but were more likely to make a fixation to the initial location when Jane held a FB about the object's location (*Est.*$_1$ = -.33, SE$_1$ = .04, $t_1$ = -9.12; *Est.*$_2$ = -.34, SE$_2$ = .04, $t_2$ = -8.3, *p*s < .001). In contrast, passive observers showed a significant bias to fixate the final location on TB trials (*Est.*$_1$ = .26, SE$_1$ = .03, $t_1$ = 7.96; *Est.*$_2$ = .29, SE$_2$ = .05, $t_2$ = 6.36, *p*s < .001), and a weaker bias to fixate this final location on FB trials, which did not reach statistical significance (*Est.*$_1$ = .06, SE$_1$ = .03, $t_1$ = 1.92; *Est.*$_2$ = .09, SE$_2$ = .05, $t_2$ = 1.93, *p*s < .06).

Finally, fixations that were made during the auditory [Location] showed a significant effect of belief, and no effects involving task. Here, all participants, regardless of task group, were more likely to fixate the final location when Jane held a TB about the object's location (Intercept *Est.*$_1$ = .3, SE$_1$ = .03, $t_1$ = 11.11; *Est.*$_2$ = .3, SE$_2$ = .04, $t_2$ = 7.94, *p*s < .001), and were more likely to fixate the initial location when Jane held a FB about the object's location (Intercept *Est.*$_1$ = -.2, SE$_1$ = .03, $t_1$ = -7.49; *Est.*$_2$ = -.18, SE$_2$ = .04, $t_2$ = -4.91, *p*s < .001).

From this fixation data we can infer that task manipulations elicited different timings of perspective inference and use between the two groups. Specifically, active participants were spontaneously sensitive to the characters' perspectives, as they employed significantly different viewing strategies in TB *vs.* FB trials immediately from the image onset; they were not waiting until prompted to do so by the audio description. The active group also showed early use of perspective information, as they correctly anticipated reference to the initial location on FB trials as soon as they heard whose perspective to take (i.e. from "Jane" onwards). In contrast, passive observers only showed significantly different visual biases between TB and FB conditions from the onset of "Jane", and in fact, did not show a reliable preference to fixate the initial location on FB trials until the location was auditorily available. This suggests that while the auditory information on whose perspective to take activated the inference about Jane’s belief, passive participants did not resolve the conflict between reality and FB representations to explicitly predict her behaviour. Both groups showed the appropriate biases to the final location throughout the time-series on TB trials, showing that participants used the visual scenes to facilitate language understanding, regardless of task. Therefore, active engagement in a task appears to activate earlier inferences about others’ perspectives, and drives immediate use of perspective information to anticipate others’ actions, compared to passive observers who are susceptible to influences from egocentric or reality biases. We will consider the source of such interference in the general discussion.

*Pupil diameter analysis*

Additional analyses monitored how participants’ pupil size was influenced by the constraints of each experimental condition during the task. Historically, pupil dilation has been thought to reflect simple low-level changes in the environment (e.g. luminance and colour; Beatty & Lucero-Wagoner, 2000; Reeves, 1920), or the valence of an individual’s emotional state (Hess & Polt, 1960; Partala & Surakka, 2003). However, more recent psychological research has established a link between pupil diameter and cognitive load, whereby increasing the cognitive demands of a task can increase pupil dilation. For example, the size of digit-span in a number memory task has been shown to be positively correlated with the size of the pupillary response (Kahneman & Beatty, 1966). Similarly, researchers have used pupillary reflexes to demonstrate the cognitive effort involved in the ‘other-race effect’, where other-race faces are more difficult to recognize than same-race faces (Goldinger, He, & Papesh, 2009). In language research, changes in pupil diameter have been found in response to manipulations of word frequency in a lexical decision task (Kuchinke, Võ, Hofmann, & Jacobs, 2007), and increasing complexity of syntactic structure during reading (Just & Carpenter, 1993). To date, pupil diameter has never been used to examine the cognitive effort that is required when participants reason about others' actions based on their true or false beliefs. The general prediction in the current study was that FB trials would incur a greater degree of cognitive effort compared to TB trials, as reflected in larger pupil size.

Analyses examined participants' average pupil diameter in each condition during the auditory sentence, while the target image was on-screen. During this relatively long exposure time (6000 msec), visual scenes were static, and were near-identical in content and lighting across the different conditions (i.e. they depicted the same three closed containers and the actress, Sarah), meaning that low-level differences in the environments were minimal (see Porter, Troscianko, & Gilchrist, 2007 for a discussion). Further, the initial 1000 msec preview period was excluded from analysis since previous research has demonstrated that the pupil needs time to adjust its size to a change in screen luminance (Pomplun, Sunkara, Fairley, & Xiao, submitted). Statistical analyses were carried out using mixed-effects models in R using the lme4 package, with participant and item templates as random effects, and task (Passive *vs.* Active), belief (FB *vs*. TB), and predictability (Unpredictable *vs*. Predictable) as fixed effects. The two levels of each fixed factor were coded using contrasts (-.5 *vs.* +.5, respectively). In addition, models were run that included random intercepts for each image's mean luminance and root mean squared contrast values to assess the variance that can be explained by low-level differences between individual images. Luminance and contrast were calculated post-hoc for each target image. Chi-squared statistical comparisons showed that including random intercepts for mean luminance (Var. = 466.5, SD = 21.6) and root mean squared contrast (Var. <.01, SD = .02) did not significantly improve the model fit, thus these low-level factors had very little impact on our experimental effects. Maximal models were also run with random slopes for the fixed effects, but statistical comparisons showed that a better model fit was achieved without including these slopes.

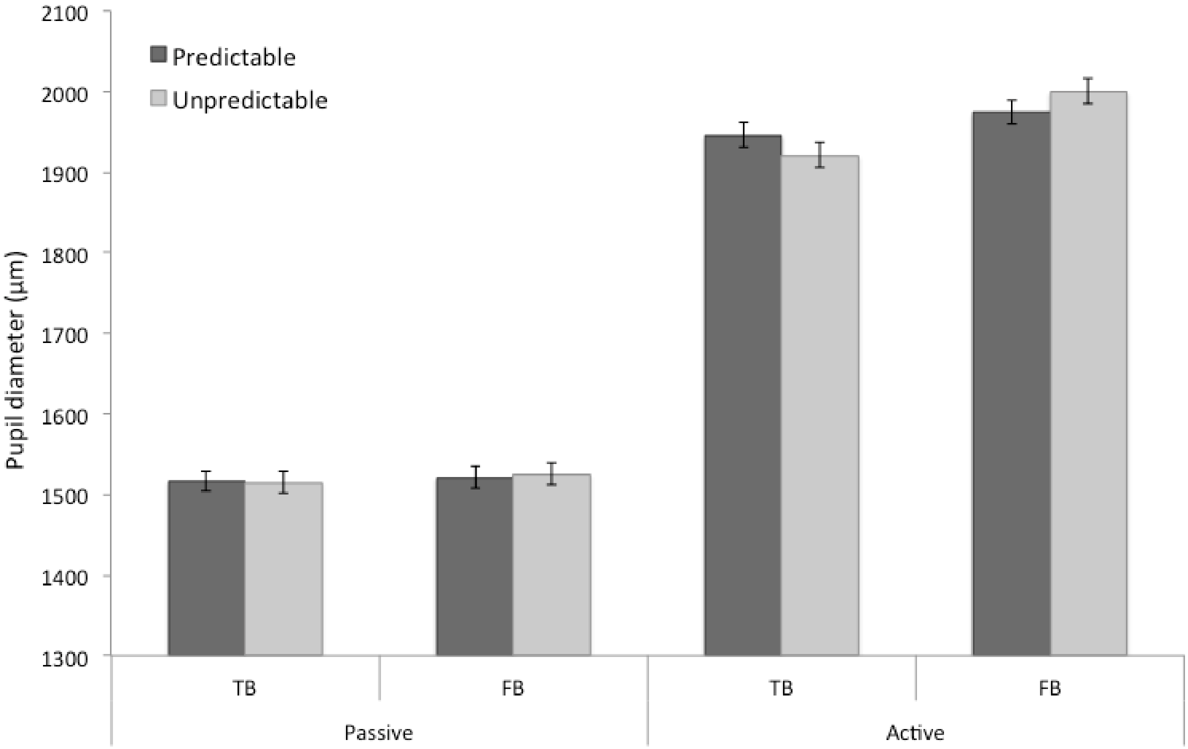

Figure 5:

Average pupil size per condition and participant group. Error bars show standard errors.

As can be seen in Table 4, analyses revealed a significant effect of task, with participants in the ‘active’ group showing increased pupil dilation (Intercept *Est.* = 1959) compared to the ‘passive’ group (Intercept *Est.* = 1519). Additionally, belief emerged as a significant effect, and as part of an interaction with task. Overall, pupil diameter was larger for FB (Intercept *Est.* = 1754) compared to TB trials (Intercept *Est.* = 1725), but when looking at the two task groups separately this effect was only present in the active participant group (*Est.* = -52.6, SE = 11.6, *t* = -4.55, *p* < .001) and not in the passive observer group (*Est.* = -5.5, SE = 11.6, *t* = -.47). Finally, we found a significant interaction between belief and predictability. Analysis of the underlying effects revealed that pupil diameter was significantly larger on unpredictable FB trials compared to TB trials (*Est.* = -45.1, SE = 11.6, *t*

= 3.9, *p* < .001), but did not differ when predictable object-container relations supported the FB interpretation (*Est.* = -13.0, SE = 11.6, *t* = -1.11).

| | Pupil Diameter | |
|---|---|---|
| | Estimate (SE) | t-value |
| Model fit (LogLik) | -12822 | |
| Fixed effects: | | |
| Intercept | 1739.2 (58.2) | 29.9*** |
| Task | 440.6 (111.0) | 3.97*** |
| Belief | -29.1 (8.2) | -3.55*** |
| Predictability | -0.2 (8.2) | -0.02 |
| Task:Belief | -47.1 (16.4) | -2.88** |
| Task:Predictability | 2.3 (16.4) | 0.14 |
| Belief:Predictability | 32.2 (16.4) | 1.97* |
| Task:Belief:Predictability | 47.8 (32.8) | 1.46 |
| Random effects: # | | |
| Participant (Var) | 244997 (495) | |
| Item (Var) | 7342 (85.7) | |
| Residual (Var) | 31959 (178.8) | |

Table 4:

Parameter estimates for pupil diameter (# random effects show variance and standard deviations). Significance tests for fixed effects use * = *p*<.05; ** = *p*<.01; *** = *p*<.001.

*Behavioural responses analysis*

Recall that participants in the ‘active task’ group were instructed to press one of three keys on a keyboard to select the container that would complete the sentence (i.e. keys corresponded to containers on the left, middle and right of the table). Behavioural analyses examined the speed and accuracy with which the ‘active participants’ group were able to select the appropriate sentence continuation during the target sentence. Reaction times were calculated relative to the onset of the auditory sentence. Responses were coded as being correct when they matched Jane’s belief about the object’s location (TB or FB), and errors were further broken down into those that were based on an appearance error (i.e. responding based on predictable object-container relations), an egocentric error (i.e. responding based on one's own knowledge of the object's location), or 'other' error. Mean reaction times and error rates for each condition are shown in Figures 6 and 7 respectively.

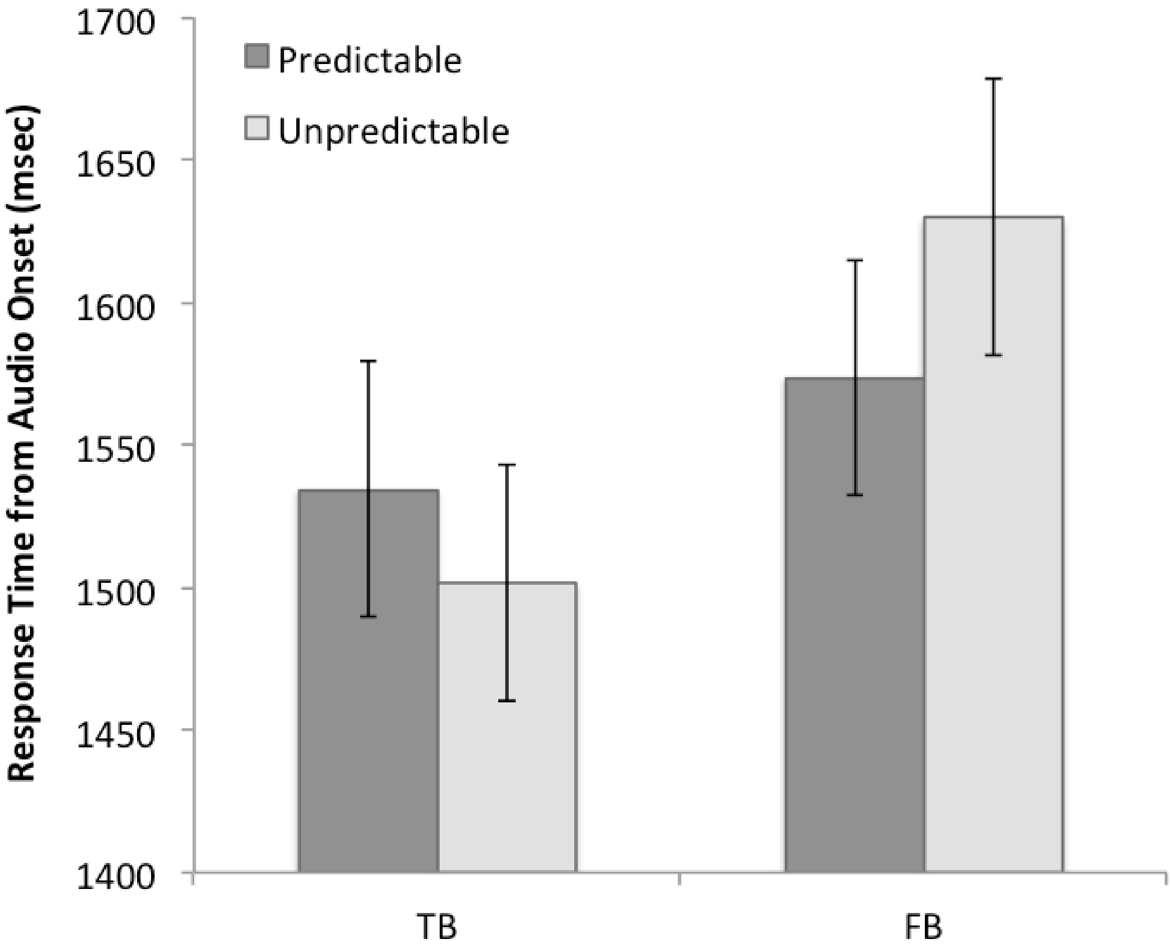

Figure 6:

Average response time from the audio onset per condition, in the active participant group. Error bars show standard errors.

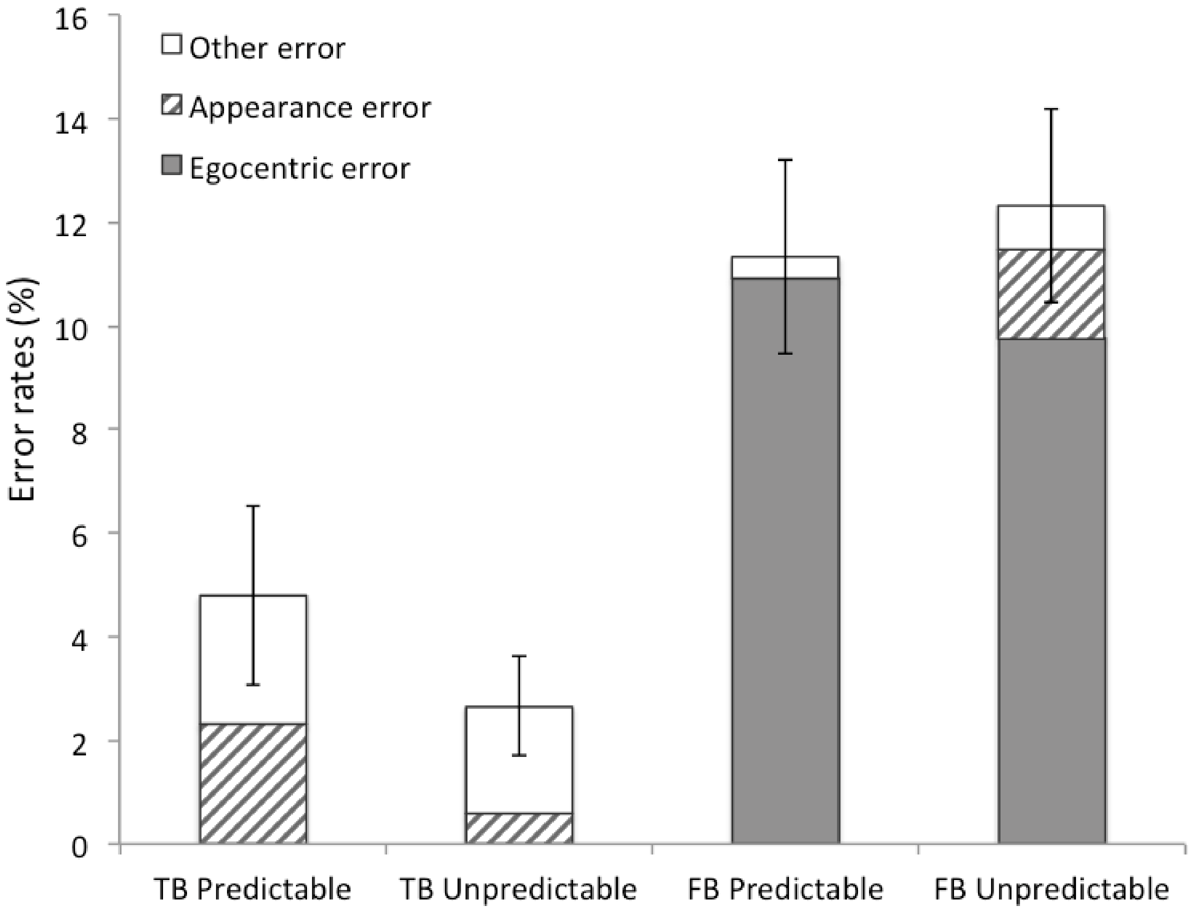

Figure 7:

Average error rates per condition, in the active participant group, broken down by type of error made. Error bars show standard errors.

Statistical analyses were carried out separately for reaction time and accuracy data. Reaction times were analysed using the lmer program, and accuracy were analysed using the glmer program (appropriate for binary data). Both participant and item templates were entered as random effects, and belief (FB *vs*. TB) and predictability (Unpredictable *vs*. Predictable) as fixed effects. The two levels of each fixed factor were coded using contrasts (-.5 *vs.* +.5, respectively). Models with random slopes were run, but Chi-squared tests showed that including these did not improve model fit. Statistical results are reported in Table 5.

| | Response Time | | Accuracy | |
|---|---|---|---|---|
| | Estimate (SE) | t-value | Estimate (SE) | z-value |
| Model fit (LogLik) | -7079 | | -196.6 | |
| Fixed effects: | | | | |
| Intercept | 1561.9 (83.2) | 18.76*** | 3.56 (.32) | 11.05*** |
| Belief | -81.9 (38.7) | -2.12* | 1.18 (.33) | 3.57*** |
| Predictability | -19.4 (38.7) | -0.5 | -0.24 (.33) | -0.74 |
| Belief:Predictability | 115.3 (77.5) | 1.49 | -0.66 (.66) | -1.01 |
| Random effects: # | | | | |
| Participant (Var) | 247997 (498) | | 0.33 (.57) | |
| Item (Var) | 8491 (92.1) | | 1.38 (1.17) | |
| Residual (Var) | 336989 (580.5) | | | |

Table 5:

Parameter estimates for response times and accuracy (# random effects show variance and standard deviations). Significance tests for fixed effects use * = *p*<.05; ** = *p*<.01; *** = *p*<.001.

Analysis of reaction time data revealed a significant difference between the two belief conditions, showing that correct responses were slower on FB trials compared to TB trials (1602 msec *vs.* 1518 msec). Neither the predictability effect or the interaction reached significance[6]. Note that the average response time across all conditions was 1559 msec after auditory onset, which corresponds to the average offset of the target object (e.g. ‘chocolates’) in the auditory input. Thus it seems that on average, participants clicked to select the correct object location prior to hearing this location information in the auditory input.

Analysis of accuracy also revealed a significant effect of belief, reflecting lower accuracy on FB trials compared to TB trials (88% *vs*. 96%). Looking at Figure 7 reveals the

[6] The same pattern of effects was found when reaction times were log transformed, which corrects for the positively skewed distribution of reaction time data, with a significant effect of belief (*Est.* = 0.06, SE = .02, *t* = 2.41, *p* < .02), but not predictability (*Est.* = 0.01, SE = .02, *t* = .54) or an interaction (*Est.* = 0.06, SE = .05, *t* = 1.34).

pattern of errors elicited in each condition; it is clear that this increased error rate on FB trials was due to participants responding incorrectly due to egocentric, rather than appearance, biases. No other effects reached significance.

## General Discussion

The current experiment adapted the visual world paradigm, combining dynamic visual scenes, language comprehension and eye-tracking to investigate how adults use knowledge about others' beliefs online to understand and predict their actions. Specifically, we examined the strength and timecourse of predictive eye movements during a false belief task, in order to understand the processes involved in making and using such inferences. There were three key manipulations in this study. Firstly, we compared comprehension in TB trials, where the participant held the same belief as the character, with FB trials, where the participant held two conflicting mental representations of the target object (i.e. according to self and character perspectives). Secondly, we manipulated participants' degree of involvement in the task by comparing performance in a group of passive observers with a group of active participants. Third, we manipulated the availability of appearance-based cues between the target object and container. The paradigm allowed investigation of both explicit (behavioural responses in active group) and implicit (eye-tracking) measures underlying these tasks. Moreover, by examining effects in each condition at four key points during the trial, we aimed to distinguish participants' inference and storage of the character's beliefs (evidenced by a difference in fixation patterns between TB and FB conditions), and their use of this information to predict the character's subsequent behaviour (evidenced by a preference to fixate the initial location on FB trials).

Across all measures, we found evidence that taking another person's perspective is more cognitively demanding when that person's knowledge is at odds with our own. Firstly, looking at participants' behavioural responses to the task (active participants), we see that FB trials elicited higher error rates and increased response times, compared to TB trials. This fits with previous research that has compared response times to judgements of true and false beliefs (e.g., Apperly, Carroll, Samson, Qureshi, Humphreys, & Moffatt, 2010; Back & Apperly, 2010; German & Hehman, 2006). Similarly, increased pupil size on FB trials compared to TB trials implies that increased cognitive effort is required to inhibit the object's real location. However, these measures cannot typically inform us whether participants are *inferring* the protagonist's mental state at the point of responding, or whether this response is indicative of participants *using* an existing ToM inference (but see Back & Apperly, 2010; Apperly, Riggs, Simpson, Samson, & Chiavarino, 2006). Previous work on ToM use during language comprehension has implicitly assumed that the key inferences have taken place, without necessarily giving participants either the time or the motivation for this to be certain. Our listeners were given a sufficiently long lead-in time, within a context that made it highly likely that participants would have inferred the protagonist's beliefs by the disambiguating point. Combined with the distinction between processed involved in ToM inferences, storage of the resulting information, and use of this information (to predict behaviour in the present study), we were able to use eye-tracking data to provide additional insights into the timecourse of ToM processing.

Participants showed significantly different visual biases between TB and FB conditions, with participants only considering the object's real location on TB trials, but also considering the object's prior location on FB trials. This looking pattern indicated that participants had inferred the character's perspective and were storing this information. However, analysis of the fixations data revealed that the active participant group had already made the perspective inference during the video sequence, while in the passive observer group this perspective inference was delayed until the auditory description revealed whose perspective to take (i.e. “Jane will look for the”). These findings support our second prediction; both groups inferred the character’s perspective prior to disambiguation, but this inference was delayed in the passive group.

Despite this sensitivity to others’ perspectives prior to the point of disambiguation, the data suggest that clear predictions (i.e. a visual preference to the appropriate location) based on false beliefs may emerge later than those based on true beliefs in both participant groups (as in Ferguson & Breheny, 2012). However, in the case of the active participant group we do not see this as delayed perspective use, since participants launched fixations to the FB-appropriate location as soon as the auditory input revealed whose perspective to adopt. Thus in line with our third prediction, active participants used perspective to direct their expectations in the earliest possible moments of language comprehension; the passive observers did not.

The current study is the first to examine how one’s involvement in a task influences perspective-taking, by directly comparing processing between active participants (i.e. actively engaged in a task that required perspective-taking) and passive observers (i.e. no explicit reason to track others’ perspectives) within a single true/false belief paradigm. Here, listeners’ performance, measured through predictive eye movements, was significantly enhanced by being actively involved in the task (i.e. explicitly predicting the characters’ actions).This pattern resembles effects observed in previous studies, where more interactive communication tasks resulted in evidence of earlier predictive eye movements. The experimental context in tasks that have shown early use of perspective, such as Brown-Schmidt et al.’s (2008) question-answer discourse, and Hanna and Tanenhaus’s (2004) chef scenario, ensured that participants were actively engaged in a task with the speaker. A full understanding of the speaker’s utterance in each case was essential to facilitate conversational interactions- and therefore to maintain the social relationship between interlocutors. In contrast, most tasks that have shown later effects of perspective asked participants to simply ‘follow the speaker’s instructions’ with no tangible benefit to the listener if they were successful or not (e.g. Keysar et al., 2000). Thus, providing a reason for participants to adopt the speaker’s perspective emerges as an important element in successful ToM use, despite the fact that inferences about others’ perspectives were activated in both conditions, whether or not predictions of behavior were explicitly required.

Interestingly, improved performance in the current study was specifically related to the speed with which participants were able to infer and use Jane’s false beliefs to predict her actions; performance on TB trials was consistently good in active and passive groups. Recall that the active participants had already made the perspective inference (i.e. showed significantly different viewing strategies in TB *vs.* FB trials) during the language-free preview period, and showed clear prediction of the appropriate initial location on FB trials in the earliest possible moments of processing (i.e. from “Jane will look for the” onwards). In contrast, passive observers did not show evidence of having inferred the character’s perspective until the verbal perspective cue was uttered (“Jane”), and did not show a reliable preference to fixate the initial location on FB trials until the location had become auditorily available. This suggests that interference from one’s own perspective needs to be actively suppressed to enable listeners to fully adopt another person’s perspective. This account draws further support from pupil diameter analyses, which showed increased pupil size in the active compared to the passive group, suggesting that the active participants were more cognitively involved in the task than the passive observers.

We argue that both active and passive tasks provide equally valuable insights into online processing of ToM. The fact that healthy adults consider other peoples’ perspectives spontaneously when observing third party interactions, even without an explicit reason to do so, is consistent with claims that at least some of these inferences are automatic in nature (see also Kovács, Téglás, & Endress, 2010; Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Lam, Bayliss, & Dux, 2012; Schneider, Nott, & Dux, 2014). However, the fact that these inferences appeared delayed in passive participants suggests that they are not automatically stimulus-driven, but instead occur spontaneously at a rate determined by participants' motivations in the task. While this implicit sensitivity to others' perspectives does not lead to clear predictions based on those beliefs in passive observers (as shown by the lack of a significant anticipatory bias to fixate the initial location), it does clearly alter the normal processing strategies that one would use without the presence of an alternative perspective. The cognitive efficiency of this process is demonstrated here by smaller pupil size in the passive group, compared to the active group. In contrast, when participants were given an explicit reason to infer another person's perspective, their performance improved, with active participants directing their anticipatory visual attention towards the perspective-appropriate box. This suggests that different cognitive processes are activated under different contexts.

It is also important to consider which cognitive processes may have been delaying the FB predictions in this study, particularly in the passive observers group. Some previous research has described a default egocentric tendency to process incoming information according to our own knowledge of reality, prior to accommodating other peoples' perspectives (e.g. Keysar et al., 2003). Indeed, in the current study these egocentric tendencies may have been boosted by the visible presence of the character Sarah, who shared knowledge of the object's real location with the participant in all conditions. However, it is also possible that egocentric biases exist as 'background tendencies' that can come into play when a representation of the alternative perspective has not been actively maintained. This would account for the fact that our active participants were able to make clear predictions about Jane's actions from the earliest possible moments of processing- the explicit task ensured that participants actively maintained the belief interpretation. Clearly the intrusion of egocentric biases on listeners' expectations cannot be ruled out by the current data, especially given that passive observers continued to show a marginal preference to fixate the final location until the onset of the target location. Moreover, the majority of response errors on FB trials in the active group were based on participants responding according to their own knowledge of events (10%), while appearance-related cues accounted for only 2% of errors.

However, looking at the eye movement data we find some evidence against the view that egocentric biases are the default perspective during comprehension. Active participants' looking behaviour on FB trials did not show an initial egocentric preference prior to predicting the initial location; participants simply divided their attention between the two potential locations. This pattern of behaviour is consistent with them holding multiple representations of the same object, one in each of the perspective-appropriate locations, which compete with each other during the ambiguous period (Altmann & Kamide, 2009). Once cues are provided from the language input (regarding whose perspective to adopt), this directs listeners to favour one of these representations over the other, resulting in a clear bias to the initial location on FB trials. Thus, the lack of egocentric bias in active participants, coupled with the finding that visual biases differed significantly between TB and FB conditions in both groups on all fixation measures (except the preview period in passive observers), supports the proposal that listeners were simultaneously aware of the different perspectives in each case; they held an egocentric representation of events alongside the perspective-appropriate representation. Indeed, the delayed bias to the initial location in the passive compared to the active group can then be explained in terms of reduced active maintenance of the alternative perspective, leading to increased influence from the egocentric perspective. In relation to Altmann and Kamide's multiple representations account, this suggests that passive participants have applied different weights to each instantiation of the target object given their task-related experience that fully adopting the character's perspective was not required to complete the task. As such, we suggest that any delay in setting up FB predictions reflect ongoing interpretation of depicted events and beliefs, which is likely to incur processing costs due to working memory demands and information suppression.

Finally we consider how the predictability of semantic relations between objects and containers influenced perspective-taking. Incorrect responses on FB trials in the active group were dominated by egocentric errors, with only a minority of errors being driven by visual constraints. Moreover, experimental manipulations of predictability did not influence fixation patterns between the final and initial locations at any time, nor did it alter response accuracy or speed in active participants. This suggests that the availability of semantic cues did not enhance the efficiency of perspective-taking. However, some support for the facilitation role of semantic information is provided here by the pupil diameter analysis, which suggests that semantically related object- initial container pairings reduced the cognitive effort required on those FB trials. This pupillometry data should be considered with caution, however, given recent reports that have questioned the reliability of eye-trackers for estimating pupil diameter when the eyes are moving (Brisson et al., in press).

Taken together, results from the current study provide online evidence that comprehenders spontaneously infer others' perspectives, and do so even without an explicit reason. However, they also demonstrate that being actively engaged in a task activates these belief inferences earlier, and leads to faster and stronger use of that knowledge to predict their subsequent actions. In contrast, when people are simply passive observers they remain susceptible to egocentric influences, which impairs their use of perspective to predict specific actions as they consider unfolding events according to both the reality and alternative perspectives.

## 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 Psychologia*, *137,* 190-200.

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.

Altmann, G.T.M. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

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

Apperly, I.A., Riggs, K.J., Simpson, A., Samson, D., & Chiavarino, C. (2006). Is belief reasoning automatic? *Psychological Science, 17*, 841-844.

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

Back, E. & Apperly, I. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition, 115*, 54-70.

Barr, D.J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109*, 18-40.

Barr, D.J. & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*, 391-418.

Barr D.J., Levy R., Scheepers C., & Tily, H. (2013) Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255-278.

Beatty, J. & Lucero-Wagoner, B. (2000). The pupillary system. In J. Caccioppo, L.G. Tassinary, & G. Berntson (Eds.). *The Handbook of Psychophysiology,* Hillsdale, NJ: Cambridge University Press.

Birch, S.A.J. & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*, 382-386.

Blakemore, S-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience, 9,* 267-277.

Brennan, S.E. & Hanna, J.E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science (Special Issue on Joint Action), 1,* 274-291.

Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (In Press). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavioural Research Methods*.

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.

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language, 61*, 171-190.

Brown-Schmidt, S. & Hanna, J.E. (2011). Talking in another person’s shoes: Incremental perspective-taking in language processing. *Dialog and Discourse, 2*, 11-33.

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.

Chambers, C.G., Tanenhaus, M.K., Eberhard, K.M., Filip, H., & Carlson, G.N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47,* 30-49.

Cooper, R.M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84-107.

Doherty, M.J. (2008). Theory of Mind: How Children Understand Others' Thoughts and Feelings. Hove, UK: Psychology Press.

Dumontheil, I., Küster, O., Apperly, I.A., & Blakemore, S-J. (2010). Taking perspective into account in a communicative task. *Neuroimage, 52*, 1574-1583.

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.

Ferguson, H.J. & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology, 48*, 257-263.

Ferguson, H.J. & Breheny, R. (2011). Eye movements reveal the time-course of anticipating behaviour based on complex, conflicting desires. *Cognition, 119*, 179-196.

Ferguson, H.J., Scheepers, C., & Sanford, A.J. (2010). Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes, 25*, 297-346.

German, T.P. & Hehman, J.A. (2006). Representational and executive selection resources in ‘theory of mind’: Evidence from compromised belief-desire reasoning in old age. *Cognition, 101*, 129-152.

Goldinger, S.D., He, Y., & Papesh, M. (2009). Deficits in cross-race face learning: Insights from eye-movements and pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1105-1122.

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.

Hanna, J.E., Tanenhaus, M.K., & Trueswell, J.C. (2003). The effects of common ground andperspective on domains of referential interpretation. *Journal of Memory and Language,49*, 43-61.

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

Hess, E.H. & Polt, J.M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132,* 349-350.

Horton, W.S.  (2007).  The influence of partner-specific memory associations on language production: Evidence from picture naming.  *Language and Cognitive Processes*, *22,* 1114-1139.

Just, M.A. & Carpenter, P.A. (1993). The intensity dimension of Thought: Pupillometricindices of sentence processing. *Canadian Journal of Experimental Psychology, 47,* 310-339.

Kahneman, D. & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154,* 1583-1585.

Keysar, B. & Barr, D.J. (2005). Coordination of action and belief in communication. In J.C. Trueswell & M.K. Tanenhaus (Eds.), *Approaches to Studying World Situated Language Use: Bridging the Language-as-Product and Language-as-Action*. Cambridge, MA: MIT Press.

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.

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

Kovács, Á.M.. Téglás, E., & Endress, A.D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330, 1830-1834.

Kuchinke, L., Võ, M.L-H., Hofmann, M., & Jacobs, A. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology, 65*, 132-140.

Mitchell, P., Robinson, E.J., Isaacs, J.E., & Nye, R.M. (1996). Contamination in reasoning about false belief: An instance of realist bias in adults but not children. *Cognition, 59*, 1-21.

Mozuraitis, M., Chambers, C.G., & Daneman, M. (2014). Privileged vs. shared knowledge about object identity in real-time referential processing. Paper presented at *27th Annual CUNY conference* on human sentence processing, Ohio State University, USA.

Newman, A.J., Tremblay, A., Nichols, E.S., Neville, H.J., & Ullman, M.T. (2012). The influence of language proficiency on lexical-semantic processing in native and late learners of English: ERP evidence. *Journal of Cognitive Neuroscience, 24*, 1205-1223.

Partala, T. & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59,* 185-198.

Pomplun, M., Sunkara, S., Fairley, A.V., & Xiao, M. (submitted). Using pupil size as a measure of cognitive workload in video-based eye-tracking studies.

Porter, G., Troscianko, T., & Gilchrist, I.D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly Journal of Experimental Psychology, 60,* 211–229.

Reeves, P. (1920). The response of the average pupil to various intensities of light. *Journal of the Optical Society of America, 4,* 35-43.

Richardson, D.C. & Spivey, M.J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition, 76*, 269-295.

Salverda, A.P., Brown, M., & Tananhaus, M.K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica, 137*, 172-180.

Samson, D., Apperly, I.A., & Humphreys, G.W. (2007). Error analyses reveal contrasting deficits in "theory of mind": neuropsychological evidence from a 3-option false belief task. *Neuropsychologia, 45*, 2561-2569.

Schneider, D., Bayliss, A.P., Becker, S., & Dux, P.E. (2012). Sustained implicit belief processing revealed by eye movements. *Journal of Experimental Psychology: General, 141*, 433-438.

Schneider, D., Lam, R. Bayliss, A.P., & Dux, P.E. (2012). Cognitive load disrupts theory of mind processing. *Psychological Science*, *23*, 842-847.

Schneider, D., Nott, Z.E., & Dux, P.E. (2014). Task instructions and implicit theory of mind. *Cognition, 133,* 43-47.

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

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *62*, 655–684.

Wilkes-Gibbs, D. & Clark, H.H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Cognition*, *31*, 183-194.

Yee, E. & Sedivy, J.C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*, 1-14.

Yoon, S.O., Koh, S., & Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin and Review, 19*, 699-707.

## Appendix A

Experimental items

*Note that for each of the items below, sentence conditions are listed in the order: TB-predictable, FB-predictable, TB-unpredictable, FB-unpredictable. Target locations for each items are shown in square brackets, from left to right, as they were laid out on the table.*

1.[video case; washing tablet box; jewellery box]
Jane will look for the video camera in the container on the right.
Jane will look for the video camera in the container on the left.
Jane will look for the washing tablets in the container on the right.
Jane will look for the washing tablets in the container on the left.

2. [jaffa cakes box; egg carton; cheese spread box]
Jane will look for the egg in the container on the left.
Jane will look for the egg in the container in the middle.
Jane will look for the cheese triangle in the container on the left.
Jane will look for the cheese triangle in the container in the middle.

3. [chocolate box; tea box; video case]
Jane will look for the teabags in the container on the right.
Jane will look for the teabags in the container in the middle.
Jane will look for the chocolates in the container on the right.
Jane will look for the chocolates in the container in the middle.

*Question:* The objects were actually in the video case? True/ False

4. [stock cube box; cigarette packet; paracetamol box]
Jane will look for the tablets in the container on the left.
Jane will look for the tablets in the container on the right.
Jane will look for the cigarettes in the container on the left.
Jane will look for the cigarettes in the container on the right.

*Question:* Which container was NOT involved in the object transfer? Cigarette packet/ Paracetamol box

5. [makeup bag; jaffa cake box; sunglasses case]
Jane will look for the sunglasses in the container in the middle.
Jane will look for the sunglasses in the container on the right.
Jane will look for the brush in the container in the middle.
Jane will look for the brush in the container on the right.

6. [fruit pastilles tube; glasses case; pencil case]
Jane will look for the glasses in the container on the left.
Jane will look for the glasses in the container in the middle.
Jane will look for the pens in the container on the left.
Jane will look for the pens in the container in the middle.

7. [egg carton; video case; tea box]
Jane will look for the video camera in the container on the right.

Jane will look for the video camera in the container in the middle.
Jane will look for the egg in the container on the right.
Jane will look for the egg in the container in the middle.

*Question:* Which of these objects was NOT pictured? Jewellery box/ Egg carton

8. [washing tablets box; egg carton; jewellery box]
Jane will look for the egg in the container on the right.
Jane will look for the egg in the container in the middle.
Jane will look for the washing tablets in the container on the right.
Jane will look for the washing tablets in the container in the middle.

*Question:* Which container was NOT involved in the object transfer? Egg carton/ washing box

9. [sunglasses case; teabag box; mince pie box]
Jane will look for the teabags in the container on the right.
Jane will look for the teabags in the container in the middle.
Jane will look for the sunglasses in the container on the right.
Jane will look for the sunglasses in the container in the middle.

10. [cheese spread box; paracetamol box; Dove soap box]
Jane will look for the tablets in the container on the right.
Jane will look for the tablets in the container in the middle.
Jane will look for the cheese triangle in the container on the right.
Jane will look for the cheese triangle in the container in the middle.

11. [glasses case; mince pie box; Roses chocolate box]
Jane will look for the sunglasses in the container in the middle.
Jane will look for the sunglasses in the container on the left.
Jane will look for the chocolates in the container in the middle.
Jane will look for the chocolates in the container on the left.

12. [video case; Roses chocolate box; glasses case]
Jane will look for the glasses in the container on the left.
Jane will look for the glasses in the container on the right.
Jane will look for the chocolates in the container on the left.
Jane will look for the chocolates in the container on the right.

13. [makeup bag; shoe box; video recorder case]
Jane will look for the brush in the container in the middle.
Jane will look for the brush in the container on the left.
Jane will look for the video camera in the container in the middle.
Jane will look for the video camera in the container on the left.

*Question:* Which container was NOT involved in the object transfer? Shoe box/ Camera case

14. [mince pie box; egg carton; teabag box]
Jane will look for the teabags in the container on the left.
Jane will look for the teabags in the container on the right.

Jane will look for the egg in the container on the left.
Jane will look for the egg in the container on the right.

15. [sunglasses case; cigarette packet; jaffa cakes box]
Jane will look for the cigarettes in the container on the right.
Jane will look for the cigarettes in the container in the middle.
Jane will look for the sunglasses in the container on the right.
Jane will look for the sunglasses in the container in the middle.

*Question:* The objects are actually in the cigarette packet? True/ False

16. [teabag box; video case; pencil case]
Jane will look for the pens in the container in the middle.
Jane will look for the pens in the container on the right.
Jane will look for the teabags in the container in the middle.
Jane will look for the teabags in the container on the right.

17. [washing tablets box; video recorder case; shoe box]
Jane will look for the washing tablets in the container on the right.
Jane will look for the washing tablets in the container on the left.
Jane will look for the video camera in the container on the right.
Jane will look for the video camera in the container on the left.

18. [jewellery box; Roses chocolate box; washing tablets box]
Jane will look for the washing tablets in the container on the left.
Jane will look for the washing tablets in the container on the right.
Jane will look for the chocolates in the container on the left.
Jane will look for the chocolates in the container on the right.

*Question:* Which container was NOT involved in the object transfer? Chocolate box/ Jewellery box

19. [cheese spread box; paracetamol box; plasters box]
Jane will look for the cheese triangle in the container on the right.
Jane will look for the cheese triangle in the container on the left.
Jane will look for the tablets in the container on the right.
Jane will look for the tablets in the container on the left.

20. [jaffa cakes box; cigarette packet; cheese spread box]
Jane will look for the cheese triangle in the container on the left.
Jane will look for the cheese triangle in the container on the right.
Jane will look for the cigarettes in the container on the left.
Jane will look for the cigarettes in the container on the right.

21. [Roses chocolate box; pencil case box; fruit pastilles tube]
Jane will look for the pens in the container on the right.
Jane will look for the pens in the container in the middle.
Jane will look for the chocolates in the container on the right.
Jane will look for the chocolates in the container in the middle.

22. [cigarette packet; paracetamol box; Dove soap box]
Jane will look for the cigarettes in the container on the right.
Jane will look for the cigarettes in the container on the left.
Jane will look for the tablets in the container on the right.
Jane will look for the tablets in the container on the left.

*Question:* Which of these containers was NOT pictured? Chocolate box/ Paracetamol box

23. [mince pie box; makeup case; Roses chocolate box]
Jane will look for the brush in the container on the left.
Jane will look for the brush in the container in the middle.
Jane will look for the chocolates in the container on the left.
Jane will look for the chocolates in the container in the middle.

*Question:* The object is actually in the chocolate box? True/ False

24. [pencil case; glasses case; video case]
Jane will look for the pens in the container on the right.
Jane will look for the pens in the container on the left.
Jane will look for the glasses in the container on the right.
Jane will look for the glasses in the container on the left.

<u>Filler items</u>

1.
Jane will look for the stock cubes in the container on the left.

2.
It’s true that the photo is in the container in the middle.

3.
Sarah will look for the shoe in the container on the right.

4.
It’s true that the soap is in the container on the right.

5.
Jane will look for the sweets in the container on the left.

6.
Sarah will look for the watch in the container in the middle.

7.
Jane will look for the sugar cubes in the container on the right.

8.
Jane will look for the cake in the container on the right.

9.
It’s true that the keys are in the container on the right.

10.
It’s true that the cake is in the container on the right.

11.
Sarah will look for the soap in the container on the right.

12.
It’s true that the money is in the container in the middle.

13.
Sarah will look for the cigarettes in the container in the middle.

14.
It’s true that the chocolates are in the container in the middle.

15.
A stranger will look for the biscuits in the container on the right.

16.
It’s true that the sweets are in the container in the middle.

17.
A stranger will look for the photo in the container on the left.

18.
Sarah will look for the watch in the container on the right.

19.
It’s true that the phone is in the container on the right.

20.
Jane will look for the nail varnish in the container on the left.

21.
Jane will look for the money in the container on the right.

22.
Jane will look for the keys in the container in the middle.

23.
A stranger will look for the cake in the container in the middle.

24.
Jane will look for the stock cubes in the container on the left.

25.
It’s true that the soap is in the container on the left.

26.

A stranger will look for the sugar cubes in the container in the middle.

27.
It’s true that the cake is in the container on the right.

28.
A stranger will look for the shoe in the container on the left.

29.
It’s true that the cake is in the container on the left.

30.
It’s true that the video is in the container on the right.

## Appendix B

| | [Preview] *p(first fixation)* | | "Jane" *p(fixation)* | | [Object] *p(fixation)* | | [Location] *p(fixation)* | |
|---|---|---|---|---|---|---|---|---|
| | Estimate (SE) | t-value | Estimate (SE) | t-value | Estimate (SE) | t-value | Estimate (SE) | t-value |
| By participants model fit (LogLik) | -99.69 | | -79.19 | | -77.64 | | -110.8 | |
| Fixed effects (by participants): | | | | | | | | |
| Intercept | 0.17 (.02) | 9.06*** | 0.15 (.02) | 7.95*** | 0.11 (.02) | 5.95*** | 0.05 (.02) | 2.75** |
| Task | 0.02 (.04) | 0.52 | -0.03 (.04) | -0.84 | -0.11 (.04) | -3.23*** | -0.03 (.04) | -0.84 |
| Belief | 0.15 (.03) | 4.38*** | 0.39 (.03) | 11.74*** | 0.48 (.04) | 12.14*** | 0.5 (.04) | 11.61*** |
| Predictability | 0.03 (.04) | 0.89 | -0.04 (.03) | -1.29 | -0.03 (.03) | -0.76 | -0.04 (.04) | -1 |
| Task:Belief | 0.24 (.07) | 3.49*** | 0.48 (.07) | 7.23*** | 0.56 (.08) | 7.09*** | 0.01 (.09) | 0.06 |
| Task:Predictability | -0.12 (.07) | -1.65 | -0.08 (.06) | -1.26 | 0.02 (.07) | 0.29 | -0.15 (.07) | -2.14 |
| Belief:Predictability | 0.1 (.08) | 1.34 | 0.04 (.07) | 0.5 | -0.05 (.06) | -0.79 | -0.12 (.08) | -1.51 |
| Task:Belief:Predictability | -0.09 (.15) | -0.6 | 0.04 (.15) | 0.29 | 0.24 (.12) | 2.07 | 0.24 (.15) | 1.59 |
| Random effects: # | | | | | | | | |
| Participant | 0.01 (.12) | | 0.01 (.1) | | 0.01 (.12) | | 0.01 (.09) | |
| Task | 0.05 (.22) | | 0.05 (.22) | | 0.04 (.19) | | 0.05 (.23) | |
| Belief | 0.03 (.19) | | 0.04 (.2) | | 0.07 (.27) | | 0.07 (.26) | |
| Predictability | 0.05 (.21) | | 0.03 (.16) | | 0.04 (.21) | | 0.03 (.18) | |
| Task:Belief | 0.16 (.4) | | 0.14 (.37) | | 0.16 (.4) | | 0.25 (.5) | |
| Task:Predictability | 0.17 (.41) | | 0.13 (.37) | | 0.15 (.39) | | 0.19 (.44) | |
| Belief:Predictability | 0.23 (.48) | | 0.22 (.47) | | 0.12 (.34) | | 0.19 (.44) | |
| Task:Belief:Predictability | 0.72 (.85) | | 0.67 (.82) | | 0.43 (.66) | | 0.81 (.9) | |
| Residual | 0.02 (.12) | | 0.01 (.12) | | 0.01 (.1) | | 0.02 (.13) | |

| | [Preview] *p(first fixation)* | | "Jane" *p(fixation)* | | [Object] *p(fixation)* | | [Location] *p(fixation)* | |
|---|---|---|---|---|---|---|---|---|
| | Estimate (SE) | t-value | Estimate (SE) | t-value | Estimate (SE) | t-value | Estimate (SE) | t-value |
| By items model fit (LogLik) | -18.81 | | -0.07 | | -15.8 | | -18.7 | |
| Fixed effects (by items): | | | | | | | | |
| Intercept | 0.17 (.04) | 4.37*** | 0.13 (.03) | 3.95*** | 0.12 (.03) | 4.84*** | 0.06 (.03) | 1.76* |
| Task | 0.02 (.06) | 0.4 | -0.05 (.05) | -1.01 | -0.13 (.05) | -2.69*** | -0.04 (.05) | -0.76 |
| Belief | 0.14 (.04) | 3.31*** | 0.39 (.05) | 8.74*** | 0.5 (.05) | 10.3*** | 0.48 (.04) | 11.1*** |
| Predictability | 0.04 (.03) | 1.29 | -0.02 (.02) | -0.86 | -0.03 (.04) | -0.86 | 0.001 (.04) | 0.04 |
| Task:Belief | 0.23 (.06) | 3.75*** | 0.47 (.08) | 5.65*** | 0.59 (.09) | 6.52*** | -0.02 (.09) | -0.18 |
| Task:Predictability | -0.11 (.07) | -1.56 | -0.03 (.05) | -0.57 | 0.02 (.06) | 0.25 | -0.12 (.06) | -1.97* |
| Belief:Predictability | 0.09 (.06) | 1.41 | 0.06 (.05) | 1.4 | -0.04 (.05) | -0.78 | -0.1 (.06) | -1.54 |
| Task:Belief:Predictability | -0.13 (.11) | -1.17 | 0.04 (.11) | 0.36 | 0.28 (.11) | 2.46* | 0.2 (.12) | 1.62 |
| Random effects: # | | | | | | | | |
| Item | 0.04 (.19) | | 0.02 (.16) | | 0.01 (.11) | | 0.02 (.15) | |
| Task | 0.07 (.26) | | 0.05 (.21) | | 0.04 (.21) | | 0.06 (.24) | |
| Belief | 0.03 (.19) | | 0.04 (.2) | | 0.04 (.21) | | 0.04 (.19) | |
| Predictability | 0.01 (.1) | | 0.01 (.09) | | 0.02 (.13) | | 0.02 (.14) | |
| Task:Belief | 0.04 (.2) | | 0.14 (.37) | | 0.14 (.38) | | 0.15 (.39) | |
| Task:Predictability | 0.07 (.27) | | 0.04 (.19) | | 0.05 (.21) | | 0.05 (.21) | |
| Belief:Predictability | 0.04 (.21) | | 0.02 (.15) | | 0.02 (.14) | | 0.06 (.24) | |
| Task:Belief:Predictability | 0.13 (.37) | | 0.2 (.45) | | 0.1 (.32) | | 0.21 (.46) | |
| Residual | 0.02 (.15) | | 0.01 (.12) | | 0.03 (.16) | | 0.02 (.14) | |

Parameter estimates for the full LMM model, by participants and items at each time point (# random effects show variance and standard deviations). Significance tests for fixed effects were estimated using MCMC sampling, where * = *p*<.05; ** = *p*<.01; *** = *p*<.001.