Eye movements reveal the time-course of anticipating behaviour based on complex, conflicting desires

Heather J Ferguson[1]

Richard Breheny[2]

[1] University Of Kent, England, UK

[2] University College London, England, UK

Correspondence to:

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

email: h.ferguson@kent.ac.uk
Tel: +44 (0) 1227 827120 Fax: +44 (0) 1227 827030

Author Notes.

This work was carried out with the support of a grant to Richard Breheny from the Arts and Humanities Research Council (Ref: AH/E002358/1) and the Centre for the Study of Mind in Nature (Oslo), ‘Linguistic Agency’ project. Thanks are due to Ruth Filik, Gerry Altmann and two anonymous reviewers for helpful comments on this work.

Eye movements reveal the time-course of anticipating behaviour based on complex, conflicting desires

Abstract

The time-course of representing others' perspectives is inconclusive across the currently available models of ToM processing. We report two visual-world studies investigating how knowledge about a character's basic preferences (e.g. *Tom's favourite colour is pink*) and higher-order desires (*his wish to keep this preference secret*) compete to influence online expectations about subsequent behaviour. Participants' eye movements around a visual scene were tracked while they listened to auditory narratives. While clear differences in anticipatory visual biases emerged between conditions in Experiment 1, post-hoc analyses testing the strength of the relevant biases suggested a discrepancy in the time-course of predicting appropriate referents within the different contexts. Specifically, predictions to the target emerged very early when there was no conflict between the character's basic preferences and higher-order desires, but appeared to be relatively delayed when comprehenders were provided with conflicting information about that character's desire to keep a secret. However, a second experiment demonstrated that this apparent 'cognitive cost' in inferring behaviour based on higher-order desires was in fact driven by low-level features between the context sentence and visual scene. Taken together, these results suggest that healthy adults are able to make complex higher-order ToM inferences without the need to call on costly cognitive processes. Results are discussed relative to previous accounts of ToM and language processing.

Keywords: Theory of Mind, eye movements, visual world paradigm, discourse processing

Theory of mind (ToM) is a commonly used term to describe the ability to understand and predict events in terms of other peoples' mental states, such as their intentions, beliefs and desires. As such, it is not just an *everyday* task, but an *every-time-we-encounter-another-person* task. So it is something that as adults we perform frequently and, with the well-documented exceptions of disorders such as autism and schizophrenia (e.g. Baron-Cohen, 2000; Frith & Corcoran, 1996), seemingly also without a great deal of difficulty. However, a long tradition of research suggests that ToM reasoning incurs a certain degree of difficulty. Research on children's development of ToM abilities strongly suggests that children only develop fully functioning ToM abilities, sufficient to pass false-belief tasks, for instance, in their fourth or fifth year (Wellman, Cross & Watson, 2001). Research with adults also suggests that, at least on certain tasks or in certain domains, ToM inferences are not automatic (e.g. Apperly, Riggs, Simpson, Samson, & Chiavarino, 2006) or as well integrated with other inferences as one might otherwise expect (e.g. Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003). Indeed even among these healthy adults, ToM performance has been related to Executive Function abilities, with recent research demonstrating correlations between ToM abilities and inhibitory control (Brown-Schmidt, 2009b; German & Hehman, 2006) and working memory capacity (Lin, Keysar, & Epley, 2010; McKinnon & Moscovitch, 2007).

Taken together, this research supports the view that those aspects of ToM involving higher-order inferences about mental states are set apart from more basic inferences about goals based on situational constraints (Perner, 1991; Tomasello & Rakoczy, 2003). On this view, higher-order ToM reasoning draws on resource consuming controlled processes. Many challenges to this view exist, however. On the one hand, infants as young as 18 months seem to be able to pass non-verbal false-belief tasks (Onishi & Baillargeon; 2005; Southgate, Senju, & Csibra, 2007), while it has been argued that older toddlers' difficulty with verbal false-belief tasks engages aspects of the task not germane to ToM inference (Bloom & German, 2000; Friedman & Leslie, 2004, Baillargeon, Scott, & He, 2010). In the domain of adult online processing, the view that perspective taking in discourse does not occur in first-pass processing has been challenged (e.g. Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tananhaus, 2008), as has the view that ToM inferences are not automatic (Cohen & German, 2009). In this paper we will address these issues by examining the timecourse with which adults are able to make appropriate ToM inferences about a character's behaviour, and by considering whether conflicts between that character's basic preferences and higher-order desires *necessarily* demand increased cognitive effort.

To date, various approaches have been used to investigate ToM reasoning in adults, with many of these tasks adapting a version of the false belief paradigm (Baron-Cohen, Leslie, & Frith, 1985). In this task, participants hear a short story that involves the transfer of a target object either with or without the knowledge of a specified character (thus establishing a basis for false belief in some conditions). Success on this task is typically measured through participants' responses to questions that require an inference about the character's false belief ("where will X look for the Y?"). Clearly then this task can provide valuable information on the developmental 'transitional phase' of ToM ability (e.g. de Villiers & Pyers, 2002; Flavell, Flavell, Green, & Moses, 1990; Wellman, Cross & Watson, 2001). However, since adults do not normally make errors on this task, its application to adult populations is more limited (Stone, Baron-Cohen & Knight, 1998; Perner & Wimmer, 1985). Over the last decade, attempts to make the task more complicated for adults by, for example, including multiple embeddings (as in, 'Lauren thinks that Tara knows that Alison ate the cheese'; Kinderman, Dunbar & Bentall, 1998; Rutherford, 2004), or increasing cognitive demands (e.g., with a concurrent memory task; Mckinnon & Moscovitch, 2007) have revealed a disruption in processing when making inferences about mental states. However, a more precise indicator of ability in such a task has been gained from recent studies that have measured both accuracy of responses and reaction times (e.g. Apperly, Back, Samson & France, 2008; German & Hehman, 2006). Generally, these studies suggest that inferring false beliefs engage more effortful cognitive processes relative to other inferences, as reflected in slower judgments relating to both reality and beliefs. For example, Apperly, Riggs, Simpson, Chiavarino and Samson (2006) found increased response times to belief compared to reality probe questions following a false belief video when participants were not explicitly instructed to track the person's beliefs.

While reaction time methods such as these have provided valuable insights into the relative availability of reality and belief concepts, they remain limited as to the detail they can provide relating to the exact time course of ToM inferences during language processing or in general. To address this issue, researchers have begun to investigate how inferences about others' knowledge evolve over time during social interaction using the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey, Eberhard & Sedivy, 1975). These studies have used eye tracking to monitor listeners' eye movements while they interpret a speaker's utterances and manipulate objects in a visual scene. Importantly, some of the visual objects are occluded from the speaker's but not the listener's view, thus creating a disparity between the set of information in the common ground (see Clark, 1996; Stalnaker, 1978) versus privileged ground. Generally, these studies agree that the addressees have rapid access to common ground information, allowing them to restrict referential candidates to those in common ground; however, inconsistencies arise regarding *timing* of such referential biases.

One view, typically referred to as the egocentric account, suggests a dissociation between peoples' ability to reflect on information from their own versus other peoples' knowledge and the routine ability to apply it in social situations (e.g. Barr & Keysar, 2002;

Epley, Morewedge & Keysar, 2004; Keysar, & Barr, 2005; Keyser et al., 2000; Keysar et al., 2003). In an example of one such study, Keysar et al. (2003) set up communication scenarios where listeners first hid a 'private competitor' object (e.g. a roll of sellotape) in an opaque brown bag and placed it alongside other items that were mutually visible to both of them, including a 'common-ground competitor' object (e.g. a cassette tape). Test trials then involved the speaker asking the listener to "move the tape" and compared trials where the target and competitor matched with those where they mismatched. Results showed that on critical trials, listeners continued to consider the private competitor as a possible referent, despite their knowledge of the speakers' ignorance about the object that is in the bag. Indeed, many listeners even attempted to move the bag itself at some point during the experiment. Experiments such as this suggest that comprehension is initially biased to the comprehenders' own knowledge rather than to the speaker's (often partially ignorant) perspective (a phenomenon clearly related to the "curse of knowledge", e.g. Birch & Bloom, 2007). This would predict that integration of other peoples' perspectives operates only as a subsequent and controlled correction mechanism that requires increased cognitive effort.

An alternative to the egocentric view suggests that addressees show a strong and early preference for potential referents in common ground. Hanna et al. (2003) argue that certain aspects of previous perspective-taking studies were detrimental to the process of integrating perspective into interpretation, such as the use of a recording in Keysar et al. (2000) or the absence of grounding referents in common ground. Hanna et al. (2003) attempted to remedy these perceived shortcomings and subsequently showed that participants were much more sensitive to the speaker's perspective, virtually without delay. However, even in these studies there is some small interference from the privileged ground competitor. Similar results are reported in Heller et al. (2008), which eliminates a global ambiguity present in the Hanna et al. studies that may have encouraged participants to overcome any 'strategic egocentrism' bias. Still better evidence that perspective can be integrated immediately comes from Brown-Schmidt et al. (2008) who use a different type of discourse to previous studies. In Keysar et al. (2000) and most other perspective research, the speaker gives orders to a participant. In Brown-Schmidt et al's study, the speaker and hearer are engaged in a joint action and the critical sentences are questions.

The authors of these studies propose that perspective-taking in discourse can be explained within the Constraint-Based framework, according to which cues to mismatched perspective are treated with greater or lesser degrees of certainty and so perspective information is incorporated only where these cues overcome the very strong constraints provided by the form of the speaker's referring expression. Constraint-Based models of comprehension are automatic, interactive and frequency-based (see MacDonald & Seidenberg, 2006) and so the account offered in Hanna et al. (2003), Heller et al. (2008) and Brown-Schmidt et al. (2008; see also Brown-Schmidt, 2009b; Nadig & Sedivy, 2002) is compatible with the idea that some, common higher-order ToM tasks, involving multiple perspectives, do not make great processing demands, in terms of EF or other mechanisms of controlled processing.

Taken together, the eye-tracking literature on speaker's perspective provides very informative evidence relevant to our question and highlights the benefits of using online techniques, such as the visual world paradigm, to examine the incremental processes involved in language comprehension. However, thus far this work has focused on one particular case of higher-order ToM inference that involves an ambiguous noun or expression and appears to suffer from a strong 'bottom up' effect from the language input (Barr, 2008). Additionally, studies of this kind always involve the participant holding a different, more informed, perspective than the speaker (see also Gerrig, Brennan & Ohaeri, 2000; Gerrig, Ohaeri & Brennan, 2000; Keysar, 1994; 2000). As we are interested in the general question of whether higher-order ToM inferences necessarily engage effortful EF processes we can look beyond these tasks which require participants to overcome a 'curse of knowledge', something which is well known to provide much interference in ToM processing (see Birch & Bloom, 2007).

Recently, Ferguson, Scheepers and Sanford (2010) have examined the processing of stories that involve a false belief using the visual world paradigm. Here, participants heard a 'reality' (e.g. "John watched Mary move the...") or 'belief' (e.g. "While John was distracted, Mary moved the...") context, followed by a target sentence describing the character's looking behaviour (e.g. Later John wanted to find his watch so he looked on the…') that was paired with visually presented location-referents. Results showed that comprehenders were able to modify expectations based on their own knowledge about reality, to direct their visual attention to appropriate referents according that character's perspective. This was evident from contextually-driven visual biases that emerged from *before* the target location onset. Critically, these results point towards a language processor that is immediately able to use relevant information about the beliefs of others to predict their subsequent actions, without interference from narrative reality. In Ferguson et al.'s study of online belief reasoning, there was no evidence of disrupted processing when comparing false belief and true belief conditions. However, it is possible that this was due to the inclusion of anomalous continuations (i.e. on half the trials the character's behaviour conflicted with their beliefs), which may have prompted comprehenders to process *all* incoming information more deeply. To follow on from this work we have designed a new paradigm using scenarios that tap into reasoning about others' behaviour without conflicting knowledge from the self-perspective.

There are many discourse contexts that call on the use of higher-order ToM abilities and involve conflicting perspectives but where the curse of knowledge is not an issue. For example, a character in a story may have conflicting desires or mixed emotions. One familiar source of conflicting desires involves conflict over private basic desires, which can result in embarrassing or socially less acceptable behaviour and a desire to conform or be accepted in the eyes of others. For instance, a character may have a love of junk food but is embarrassed about this and does not want other people to know about it. Making a prediction about such a character's public behaviour requires taking into account different possible intentions (or courses of action) which conflict with one another and it requires judging whether other people's attitude to the character's truly desired action will affect that character's behaviour. Although any prediction about such a character's actions clearly involves a higher-order theory of mind inference and also a comparison of conflicting mental states, there is no interference from information in the discourse situation (such as 'pull of reality' or effects of linguistic input) that is found in the online studies discussed above. Additionally, such a 'secret desire' scenario should be familiar to most of us and should not pose too much difficulty due to its novelty.

A similarly complex scenario, in terms of ToM inference, is one where a character has a great love of junk food but is very happy for people to know about that. Here a character's public actions will be based both on basic preferences and higher-order intentions with regards to the attitude of others, but unlike the 'secret' scenario, the basic-level preference and the higher-order intentions do not result in conflicting possible courses of action. Like the secret scenario, characters that hold such 'open' attitudes should also be familiar to most of us.

In the current experiments, participants listened to short stories made up of two sentences. The first sentence introduced a property of a character (e.g. a personal preference) and set up a context in which that character was either happy about other people knowing about the property (basic preferences and intentions were consistent, as in (1a)) or where the character did not want people to know about the property (basic preferences and intentions were in direct conflict, as in (1b)).

(1) a. Elaine is very open about her love of junk food.

b. Elaine is very self-conscious about her love of junk food.

The second sentence (2) elaborated on this by describing the story character performing some action that was consistent with the preceding context (so it's appropriate for Elaine to eat a burger within (1a), but something other than junk food, i.e. a salad, within (1b)), while participants viewed a visual scene containing images that matched the character's action. Note that within both conditions, the relevant action is predictable given the background information provided in (1). However, both conditions involve fairly sophisticated ToM reasoning, as characters make choices based not only on their basic preferences but also on how they wish to be seen in the eyes of others. Additionally, in the 'secret' condition (1b), there is a conflict between a choice of action based on the salient basic preference (*Elaine's love of junk food*) and the equally salient intention to keep this preference secret from other people. Thus, although the reasoning from the background information in (1) to a prediction about the choice in (2) is equally sophisticated for both conditions, there appears to be an additional hurdle to overcome in the secret condition in ignoring potential predictions based on the character's salient basic preference.

(2) a. Earlier, Elaine ate lunch in public and she intentionally chose to eat a burger.

b. Earlier, Elaine ate lunch in public and she intentionally chose to eat a salad.

As suggested above, with this design a fully warranted prediction in both cases would involve taking into account the character's intentions with respect to other people in addition to basic preferences. Specifically, it would involve making inferences about the perspective of other people if the simple preference is acted upon. This step applies to both cases, but is of particular interest in the secret condition since a correct prediction cannot be made without making both inferences. In contrast, information about basic preferences and intentions in the ‘open’ condition (1a) both point to the same target- so it is possible to correctly predict the appropriate referent without deriving the full warrant from the context.

Therefore, if comprehenders have established a representation of the character’s perspective based on contextual information about their basic preferences and higher-order intentions, then different anticipatory biases should become apparent within each of these contexts, regardless of the matched information about the character’s basic preferences. Specifically, we reasoned that such anticipation should emerge from the point at which comprehenders learn about the character’s overall goal or activity (e.g. *eating lunch*). Thus, we would expect proficient theory of mind reasoners’ visual biases to reflect this prediction some time before the offset of the disambiguating word (“burger”/ “salad”).

However, if higher-order inferences about others’ minds are more difficult to set up (as suggested by Apperly et al., 2008), we would expect to see a delay in generating expectations. Additionally, if inferences that involve conflicting mental states necessarily involve the operation of costly executive mechanisms, then we should expect to see a difference to emerge between the ‘open’ condition and the ‘secret’ condition, as only the latter involves inferences about conflicting intentions. We will examine the time-course of such anticipation with the aim of drawing conclusions as to whether increased cognitive effort, as reflected by the time required to set up expectations, is needed to generate inferences when they are based on contextually available information about the character’s intentions to keep basic preferences a secret compared to the character’s intention to openly act according to those basic preferences.

Experiment 1

Method

Participants

Thirty participants (15 female) from University College London were paid to take part in the study. All were native English speakers with normal or corrected to normal vision and had no prior exposure to the experimental items.

Stimuli and Design

Sixteen experimental pictures were paired with auditory passages in one of two conditions. Table 1 and Figure 1 provide an example of such experimental sentences and the associated visual displays. Visual displays were created using commercially available clip art collections and were presented on a 17 inch colour monitor running at 85 Hz refresh rate in 1024 x 768 pixels resolution. Each scene contained four images: Character, Open Referent (pink car), Secret Referent (green car), and a Distracter (weights), which was neither open- nor secret-congruent. To prevent any systematic viewing strategies, spatial arrangements of these four picture elements differed across items. Sound files consisted of two sentences. Sentence one introduced a fact about the story character (e.g. *Tom's favourite colour is pink*) within either an open or secret context ("Tom is always telling people that…" versus "Tom doesn't want anyone to know that…"). Sentence two then described an event (e.g. *buying a new car*) that drew reference to an open- or secret- relevant referent ("…he deliberately chose a *pink* versus *green* car"), resulting in a 1-factor within subjects design. The position of the critical word always occurred roughly mid-sentence.

One version of each item was assigned to one of two presentation lists, with each list containing sixteen experimental items, eight in each of the two conditions, blocked to ensure that they were evenly distributed. In addition, twenty-four unrelated filler items were added to each list[1]. They all consisted of correctly matched picture-sentence pairings and were interspersed randomly among the sixteen experimental trials to create a single random order. These filler items required the participant to make linguistic inferences (e.g. "If mosquitoes thrived in dry climates, they would quickly find a suitable habitat. Mosquitoes would be in abundance in certain locations and frustrate all the local residents."), but importantly they did not involve ToM reasoning, surprising events or unexpected endings. Each subject only saw each target sentence once, in one of the two conditions. At least one filler trial intervened between any two experimental trials.

Table 1:

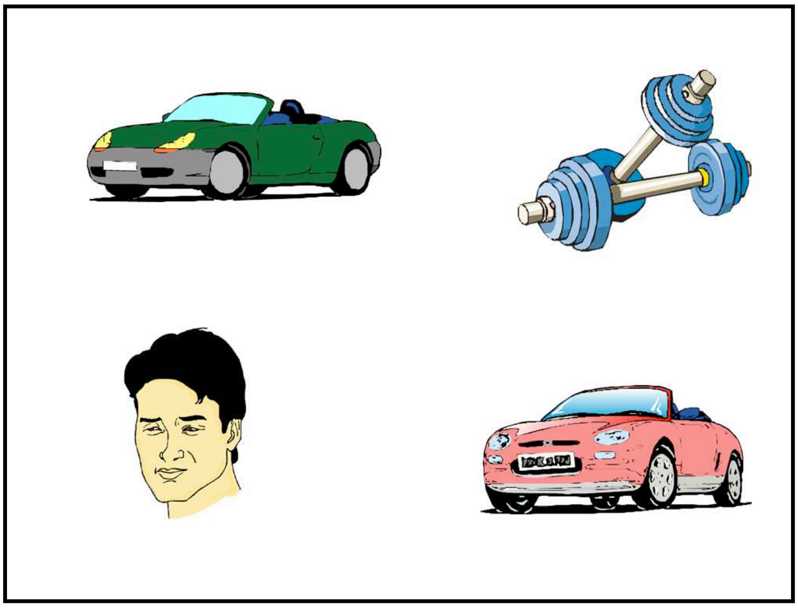
Examples of experimental sentences (Experiment 1).

Open
Tom is always telling people that his favourite colour is pink.
Last week Tom bought a new car and he deliberately chose a pink car.
Secret
Tom doesn't want anyone to know that his favourite colour is pink.
Last week Tom bought a new car and he deliberately chose a green car.

Figure 1:

Example visual stimulus used in Experiments 1 and 2. Participants heard the target sentence (see above) whilst viewing this picture

[1] Please contact the first author for a full list of filler items.



Sentences were recorded in one session from a female native British English speaker who was instructed to use a neutral intonation. The auditory files were presented as 44.1 KHz mono sound clips via headphones connected to the eye-tracker PC. The temporal onsets of critical words were hand-coded with millisecond resolution using the GoldWave sound-editing package.

Comprehension questions, relating to either the auditory or visual input, followed half of the experimental and half of the filler trials. Participants did not receive feedback for their responses to these questions. Only participants scoring at or above 80% accuracy on the comprehension questions were used in the data analysis.

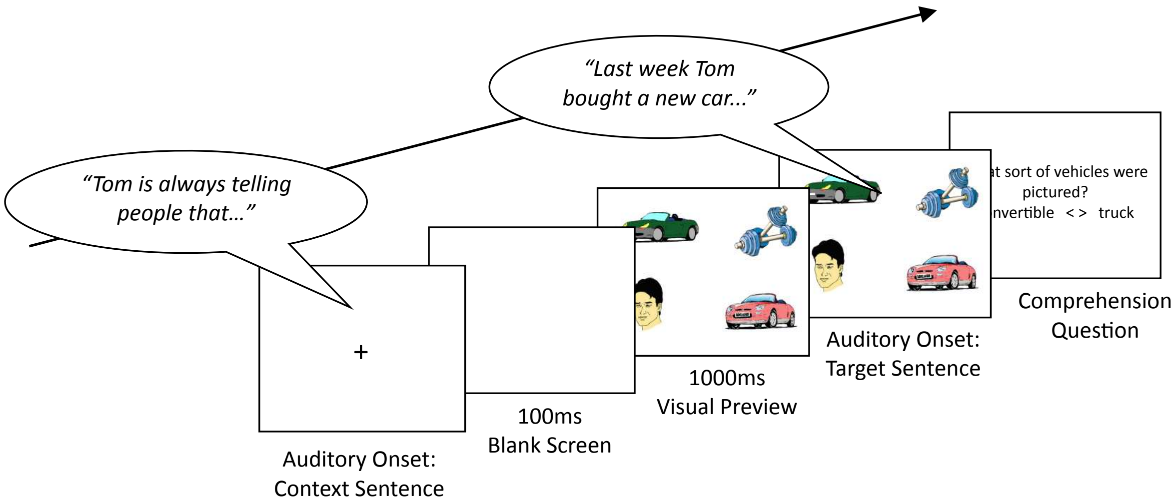
Procedure

Participants were seated in front of a 17 inch colour monitor with integrated eye tracking system (Tobii 1750) running at 50 Hz sampling rate. Viewing was binocular and eye movements were recorded from both eyes simultaneously. Participants were given the following instruction: “In this experiment you will hear short spoken passages and during the second sentence a picture will also be displayed. We are interested in how the pictures help you understand the spoken passages".

The experiment was controlled using e-Prime (Schneider, Eschmann, & Zuccolotto, 2002). As illustrated in Figure 2, each trial began with the presentation of a single centrally-located cross and participants were asked to fixate it for 1500msec before the trial was automatically initiated. The fixation cross remained on screen while participants heard Sentence 1 (open or secret context, no picture presentation). They were asked to continue looking at the fixation cross during this time. Then a 100ms blank screen was presented, followed by the target picture combined with Sentence 2. The onset of the target picture preceded the onset of the corresponding spoken sentence by 1000ms. The picture stayed on the screen for nine seconds, and the corresponding sentence typically ended 1-2 seconds before the end of the trial.

Figure 2:

Illustration of the experimental procedure in Experiments 1 and 2.



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

Results and Discussion

Data Processing

Eye movements that were initiated during Sentence 2 were processed according to the relevant picture and sound onsets for the purpose of aggregating the location and duration of each 50Hz (i.e. 20ms) sample from the eye tracker. For analysis, any sample that was deemed 'invalid' due to blinks or head movements was removed from the data. The spatial coordinates of the eye movement samples (in pixels) were then mapped onto the appropriate object regions using colour-coded bitmap templates; if a fixation was located within 20 pixels around an object's perimeter, it was coded as belonging to that object, otherwise, it was coded as background. All consecutive samples within one object region before the eyes moved to a different region were pooled into a single *gaze*. Finally, temporal onsets and offsets of the gazes were recalculated relative to the corresponding picture onset by subtracting the picture onset from the relative gaze onsets and offsets.

Probabilities of gazes to the critical open and secret referents as a function of time were analysed as described by Arai, van Gompel & Scheepers (2007), using the following log-ratio measure:

(Eq. 1) $$\log(\mathrm{Open}/\mathrm{Secret}) = ln(P_{(\mathrm{Open})} / P_{(\mathrm{Secret})}),$$

where $P_{(\mathrm{Open})}$ refers to the probability of gazes on the open referent (pink car) and $P_{(\mathrm{Secret})}$ to the probability of gazes on the secret referent (green car); *ln* refers to the natural logarithm.

The output is therefore symmetrical around zero such that a positive score reflects higher proportions of gazes on the open referent and a negative score reflects higher proportions of gazes on the secret referent. Note that a score of zero indicates equal bias towards the two referents.[2]

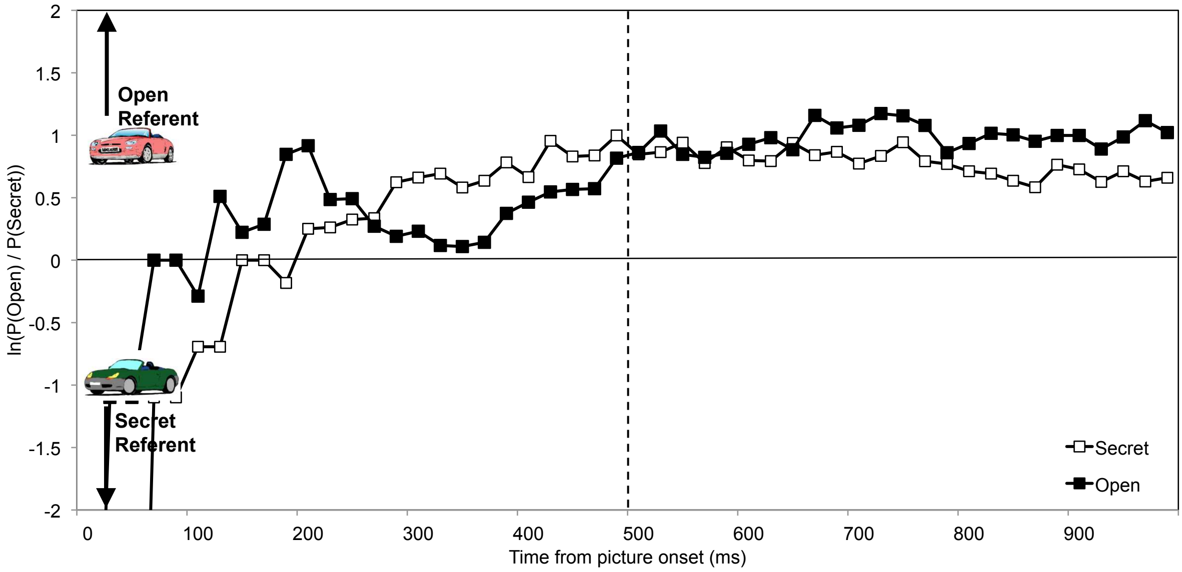
Preview Region Analyses

In order to examine very early anticipatory effects prior to language input, we analysed distributions of fixations during the picture preview. Recall that the onset of the picture preceded the onset of the corresponding target sentence by 1000ms. So, for each condition, we calculated the average log(Open/Secret) scores for each condition from 500ms to 1000ms post-picture onset (prior to that time period, participants were likely to fixate the area around the previously presented fixation cross, meaning insufficient numbers of observations for log-ratio analyses). The data were synchronized on a by-trial basis, relative to 500ms post-picture onset in the appropriate item-condition combination. Figure 3 shows the relevant descriptive data. Inferential analyses revealed no significant difference between conditions [all $Fs < 1.3$], and therefore indicate that context does not influence the distribution of looks between the two target referents prior to language input. Follow-up *t*-tests comparing the dependent measure to chance revealed an early preview-bias to the open referent for both secret ($t_1(29) = 3.61$, $p < 0.001$; $t_2(15) = 2.95$, $p < 0.01$) and open ($t_1(29) = 5.66$, $p < 0.001$; $t_2(15) = 2.99$, $p < 0.01$) conditions.

Figure 3:

The average log(Open/Secret) scores for each condition during the Preview in Experiment 1.

[2] Separate analyses of the proportions of gazes on the Character and Distracter referents revealed no significant cross-condition effects whatsoever (contact authors for a graphical representation of these data).



These results show that prior to linguistic input, participants' visual biases do not initially differ following open and secret contexts. Further, they show an initial visual preference for the open referent (*pink car*) over the secret referent (*green car*), which is likely to reflect the fact that the open referent was explicitly mentioned within both contexts.

Main Analyses

For the main analyses, we were interested in whether contextual information about a character's intention to keep a secret (open *vs.* secret) affected proportions of gazes on the open referent (the *pink car* in Figure 1) relative to the secret referent (the *green car* in Figure 1), specifically in time periods preceding the onset of the disambiguating word ("pink" or "green"). Thus, we analysed a time period ranging from the auditory onset of the target sentence (e.g. "Last week…"), until the offset of the disambiguating target noun ("pink" or "green", respectively). Here, the data were synchronized on a by-trial basis, relative to the actual onsets and offsets of relevant words in the appropriate item-condition combination. Figure 4 plots the observed average log(Open/Secret) data in each condition, for every 20 ms time-slot within the selected time period. Note that eye movements and auditory input have been resynchronized according to the onset of critical regions (see Altmann & Kamide, 2009), and as such represent more accurate plots of evolving visual biases around the scene. Thus, the dashed vertical lines in Figure 3 indicate the absolute onsets and average offsets of words in the target sentence, as labelled.