Egocentric bias across mental and non-mental representations in the Sandbox Task.

ACCEPTED FOR PUBLICATION in QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 9TH OCTOBER 2017

Samuel, Steven[a]

Legg, Edward. W.[a]

Lurz, Robert[b]

Clayton, Nicola S.[a]

[a] Department of Psychology, University of Cambridge, U.K.

[b] Brooklyn College, City University New York, New York, U.S.A

Word count: 7901 (excluding abstract, acknowledgements, references and tables, including footnotes and captions)

Key words: Theory of Mind, Egocentric bias, Sandbox Task, False belief, False photo.

Address for correspondence: Steven Samuel. University of Cambridge, Department of Psychology, Downing Street, CB2 3EB. Tel: 01223 333563. Email: ss2391@cam.ac.uk

Abstract

In the Sandbox Task (e.g. Sommerville, Bernstein & Meltzoff, 2013), participants indicate where a protagonist who has a false belief about the location of an object will look for that object in a trough filled with a substrate that conceals the hidden object's location. Previous findings that participants tend to indicate a location closer to where they themselves know the object to be located have been interpreted as evidence of egocentric bias when attributing mental states to others. We tested the assumption that such biases occur as a result of reasoning about *mental states* specifically. We found that participants showed more egocentric bias when reasoning from a protagonist's false belief than from their own memory, but found equivalent levels of bias when they were asked to indicate where a false film would depict the object as when they were asked about a protagonist's false belief. Our findings suggest that that egocentric biases found in adult false belief tasks are more likely due to a general difficulty with reasoning about false representations than a specialised difficulty with reasoning about false mental states.

Introduction

Successfully comprehending what someone else thinks, knows, sees or believes, known as Theory of Mind (ToM) or 'mentalizing', is a critical ability in our day-to-day interactions with other people (e.g. Clark & Brennan, 1991; Grice, 1975; Sperber & Wilson, 2002), and has attracted a vast amount of research as a result (for reviews see Apperly, 2010; Apperly & Butterfill, 2009; Carruthers, 2015; Heyes, 2014a; 2014b). Historically, ToM was thought to develop at around four years of age because this is when typically developing children are capable of attributing false beliefs to others, showing that they are aware that what they believe and what another person believes can differ (Wellman, Cross, & Watson, 2001). However, even adults, who clearly understand that states of mind can vary from person to person, are susceptible to egocentric bias; that is, they struggle to disengage from their own current mental states when reasoning about the mental states of others. For instance, it has been shown that adults over-ascribe their own thoughts and feelings to others (e.g. Ross, Greene, & House, 1977; Van Boven & Loewenstein, 2003), that they fail to fully take into account that what others see may differ from what they see (e.g. Apperly, Carroll, Samson, Humphreys, Qureshi, & Moffitt, 2010; Keysar, Lin, & Barr, 2003), that what they believe may not be what others believe (e.g. Coburn, Bernstein, & Begeer, 2015), and that others may be less knowledgeable than they are (e.g. Bernstein, Atance, Loftus, & Meltzoff, 2004).

Crucially, the intrusion of our own knowledge, or an important proportion of such intrusions, is often argued to be specific to cases where we reason about others' mental states, and not when thinking about functionally equivalent information held in other formats, such as memories, rules, or non-mental representations such as photographs. A number of researchers have argued for a degree of domain-specificity for ToM processes (e.g. Baron-Cohen, 1995; Cohen, Sasaki & German, 2015; Leslie, 1987; Leslie, Friedman & German, 2004; Leslie & Thaiss, 1992), and such accounts would seem to support the possibility—at least indirectly—that levels of egocentric bias may vary depending on whether a mental representation is processed by a specialised cognitive mechanism.

Evidence for an important role for reasoning about mental states in the generation of egocentric biases comes from a number of studies where a naïve protagonist's mental state or visual

perspective is required to be 'read', such as in false belief tasks and perspective taking tasks (e.g. Apperly et al., 2010; Bernstein, Thornton, & Sommerville, 2011; Coburn et al., 2015; Dumontheil, Apperly & Blakemore, 2010). For example, in the Director's Task (e.g. Keysar, Barr, & Horton, 1998; Keysar et al., 2003; Apperly et al., 2010), participants are instructed to move objects within a 4x4 grid by a director. In doing this they need to take into account the director's perspective when interpreting their instructions because some slots of the grid are occluded by opaque inserts that prevent the director seeing into them. For instance, if asked by the director on the other side of the grid to 'move the small ball' when the participant sees three balls of differing size, the smallest of which is not in the director's view, the correct response is to move the medium-sized ball. Despite being made explicitly aware that the director's view is partial, even healthy adult participants make errors—that is, they sometimes select objects that only they can see. Crucially, when participants are asked to follow an abstract rule, according to which they should ignore all objects in occluded slots, they must ignore the same objects as when they have to take into account the director's perspective—that is, the information is functionally identical to the director-present condition. Yet in this abstract rule version of the task, participants make fewer egocentric errors than in the director-present version (Apperly et al., 2010; Dumontheil, Apperly & Blakemore, 2010; though see Dumontheil, Hildebrandt, Apperly, & Blakemore, 2012; Dumontheil, Küster, Apperly, & Blakemore, 2010). It is argued that this shows that egocentric bias is greater when reasoning about mental states specifically, since in the rule version the task requirement was functionally identical.

However, previous concerns within the developmental literature that such biases may generalise to *all* representations have led researchers to compare children's performance on false belief tasks with tasks requiring children to reason about other forms of false representations, such as false photographs (e.g. Leekam & Perner, 1991; Leslie & Thaiss, 1992; Slaughter, 1998; Zaitchik, 1990). Recently, the photograph control has also been used to investigate adults' biases in ToM tasks. When applied to the Director's Task, with participants required to follow instructions based on what appeared in a photograph rather than according to a director's instructions, similar error rates were found in the photograph and the 'ToM' version of the experiment (Santiesteban, Shah, White, Bird,

and Heyes, 2015). These results suggest that egocentric bias may be linked to the reasoning about representations *per se*, be they mental or non-mental, a view that is supported by a number of developmental studies (e.g. Iao & Leekam, 2014; Iao, Leekam, Perner, & McConachie, 2011; Leekam, Perner, Healey, & Sewell, 2008; Zaitchik, 1990).

The results of these studies suggest that a mental-state specific account of egocentric bias needs to simultaneously satisfy not one but two independent requirements in order to hold. We term these requirements the Representation Contrast and the Mentalizing Contrast (see Figure 1). The most basic requirement is the Representation Contrast, which contends that there is something unique to the processing of representations—be they mental or non-mental—that leads to one's own knowledge intruding more compared to when processing functionally identical but *semantic* content, such as memories or rules. For example, greater egocentric bias should be found when reasoning about another's belief (e.g. 'Where does she think the item is?') compared to reasoning from one's own memory (e.g. 'Where was the item originally hidden?') or a rule (e.g. 'Ignore objects in front of occluders'). The Mentalizing Contrast, on the other hand, states that there is something special about the processing of *mental* representations that is not captured when processing functionally identical information carried instead by *non-mental* representations, such as photos (e.g. 'Where will the photo show the object was hidden?'). Therefore, the Mentalizing Contrast predicts the presence of additional bias related to thinking about *beliefs* over and above any bias related to thinking about photos or similar non-mental representations. Evidence of a unique proportion of egocentric bias attributable to the mental representation tasks tested against *both* the Representation Contrast *and* the Mentalizing Contrasts would suggest particularly strong support for a mental-state specific account of egocentric bias, as it would be shown to differ from both semantic knowledge and other forms of representation. On the other hand, support for the Representation Contrast alone would instead support the view that egocentric bias is the result of processing representations *per se*. Since there is presently no theory in the literature which predicts greater bias for non-mental representations than both mental representations and semantic content in the simultaneous absence of a difference between the latter

formats, we take these two hypotheses to be hierarchical, such that we may find support for the Representation Contrast in the absence of support for the Mentalizing Contrast, but not vice versa.

[Insert Figure 1 about here]

In the present study, we aimed to investigate the ToM Account by testing both the Representation Contrast and Mentalizing Contrast in healthy adult participants. One potential criticism of tasks used to test ToM accounts, such as the Director Task and the classic change-of-location false belief task, is that they may lack sensitivity to egocentric biases in adults because they rely largely on an analysis of dichotomous correct/incorrect data, which are not sensitive to the degree of bias that may be present in adults but that is not produced in cases where only one of two responses is permitted. A recent candidate for a task that avoids such issues is the 'Sandbox Task', adapted by Bernstein, Sommerville and colleagues (Bernstein et al., 2011; Sommerville, Bernstein, and Meltzoff, 2013) from a paradigm originally developed by Huttenlocher, Newcombe, and Sandberg (1994). In this task, the participant is told about a scenario in which a protagonist hides an object in a trough ('sandbox') while being watched by a second person. The protagonist then moves the object to another location within the same trough while the second person is absent. The trough has no distinctive features, and as such, participants' responses are not confined to two discrete locations, as is the case in more typical false belief tasks. Thus, the Sandbox Task provides a scalar metric (i.e. measured in centimetres) rather than a binary one and as such is more sensitive to subtle response biases. Typically, when participants are asked to indicate where the person with a false belief would search for the object, they tend to indicate locations closer to the object's actual location (where participants know the object to be) than would be expected by chance. Moreover, this egocentric bias is greater than when participants are asked to locate, based on their own memory, where the object was first hidden (Sommerville et al., 2013; Coburn et al., 2015). As with the results of the Director Tasks, these findings from the Sandbox Task have been seen as evidence of an increased tendency to

be biased when reconstructing information from another person's perspective or belief compared to when simply recalling the first location from personal memory.

That the results from the Director's Task may be explained by a bias when reasoning about any form of representation (not just beliefs and mental states) highlights the need to compare participants' performance when reasoning about false mental representations with their performance when reasoning about false non-mental representations. To date, performance on the Sandbox Task has indicated that adults (and children) show greater bias on trials in which they have to indicate where another person mistakenly believes the hidden object to be compared to trials in which they have to indicate where they themselves remember the hidden object to be. This result would appear to satisfy the Representation Contrast. However, performance on these trials has not yet been compared to performance on matched non-mental representation trials. That is, there is evidence for the Representation Contrast, but not yet evidence for the Mentalizing Contrast. Consequently, we sought to test whether or not bias would be specific to mental state reasoning alone. Initially, we aimed to replicate the effect of bias on mental state trials versus memory trials. To do so, we presented participants with a Sandbox Task similar to those used in previous studies (Sommerville et al., 2013; Coburn et al., 2015). We compared the levels of bias on false belief trials, in which participants were asked to indicate where a protagonist with a false belief about the object's location would search for the object, to memory trials, in which participants were asked to indicate from their own memory where the object was originally hidden. In both cases, the correct response would be the object's original location). Since we were also validating a version of the task for participants to perform online, it was also important to replicate this effect first. In Experiment 2 we tested whether equivalent bias occurs whenever participants have to reason about false representations. If bias is mental-state specific, as the Mentalizing Contrast holds, then bias should differ when participants are reasoning about another's belief compared to when they are reasoning about the content of a non-mental representation, such as a film.

Experiment 1

Method

Participants

Participants were recruited using Prolific Academic (www.prolific.ac) and were required to be between 18–40 years of age, English native-speakers and residents of the UK. All provided informed consent and were compensated for their time. We recruited a total of 81 participants but excluded a large number who did not meet a series of attention and motivation checks (see *Analysis* section) such that the final sample size was 40.

Materials

The images presented to participants were modelled on the pen and paper version of the Sandbox Task developed by Coburn et al. (2015). All images[1] were created in Powerpoint and converted into PNG files for presentation on Qualtrics (Qualtrics, Provo, UT, 2015). Participants saw the images in landscape (756 x 567 pixels) format, with a container ('trough') 584 pixels in width and 36 pixels in depth, centrally positioned on the horizontal axis, with text (red, bold font) above. Crosses ('x', 15 x 15 pixels) on the troughs marked the locations where the object was hidden first (Location A) and hidden the second time (Location B). The distance the object moved from Location A to Location B were derived from Coburn et al. (2015) by converting the absolute distances reported into proportions of the length of the trough. Thus the object was moved either 22.65% ('Short Move' trials) or 45.3% ('Long Move' trials) of the trough to either the left or right of Location A, counterbalanced for both distance and direction (see Table 1). Given that the actual locations within the trough were not reported in Coburn et al., the co-ordinates of Location A were derived from Sommerville et al.'s (2013) Experiment 2 by converting them into proportions for use with our stimuli. The texts used in our study were based on those used in Sommerville et al.'s study (see supplementary information for texts). After participants had finished reading the test scenarios, they were given a word search puzzle to complete before being asked the test question. The word search task was used, as in Coburn et al.

[1] Available from the author.

(2015) and Sommerville et al. (2013), to prevent participants from using a perceptual strategy to answer the test questions regarding the location of the hidden object. The word searches were created by inserting seven randomly-generated words into a word search puzzle using www.puzzlemaker.com.

[Insert Table 1 about here]

Procedure

The procedure for both experiments closely followed that used by Coburn et al. (2015). The test consisted of nine trials, eight experimental trials (four false belief trials and four memory trials) and one trial (true belief trial) designed to check participants' attention. Each trial was formed of four slides (see Figure 2). In the eight experimental trials a vignette described how one protagonist saw an object hidden in the location A (slide 1) but did not see it being moved and hidden at location B (slide 2). In the true belief trial the vignette described a protagonist who saw an object being hidden in both the first and the second location. All true belief trials were long distance trials. On the third slide, participants were presented with the word search, which remained on screen for 20 seconds. Finally, on the fourth slide participants were asked to indicate by clicking on a location with their mouse either where the object was first hidden (memory questions) or where the protagonist would search for the object (belief questions). We recorded where on the screen (in pixels) the participant clicked.

[Insert Figure 2 about here]

Following Coburn et al. (2015) the experimental trials and questions were presented in a fixed order with memory questions in trials 1, 2, 8 and 9 and false belief questions in trials 3, 4, 6 and 7. In all eight experimental trials the correct response was to indicate the first location in which the object

had been hidden. The true belief trial was always presented as the middle trial (trial 5) and participants were asked a belief question; unlike the 8 experimental trials the correct response was to indicate the second location in which the object had been hidden.

Analysis

For each trial, we converted the pixel on which the participant clicked into a measure of the participant's relative bias. This was calculated by taking the difference in pixels between the correct location (in experimental trials this was the first location in which the object was hidden) and the location the participants clicked on and then dividing this difference by the length of the trough in pixels. The bias was assigned a negative value if the clicked-on location was further from the first location the object was hidden in relative to the second location the object was hidden in and a positive value if the clicked-on location was closer to the second location the object was hidden in.

To allow for comparisons between studies we aimed to use some of the same statistical tests run by Coburn et al. (2015) and Sommerville et al. (2013), namely paired comparisons of the average bias between memory and false belief trials where the object was moved a 'long' distance and the average bias between memory and false belief trials where the object was moved a 'short' distance. As our data were not normally distributed, we report Wilcoxon signed-rank tests rather than the original t-tests. Following Coburn et al., we also compared the number of participants who showed greater bias on false belief trials than memory trials to the number of participants who did not.

To ensure that participants were paying attention to the task we excluded from the analyses all participants who failed to indicate a location closer to the second point than the first on the attention-check trial. Since this was the only trial type in which the second location was the correct response, any strategy that involved the consistent selection of the first location would lead to failure on this trial and exclusion from the results. Details of exclusions owing to attention checks are available in the Supplemental Information.

Results

Figure 3 displays the results for each trial type, and Figure 4 displays the distribution of responses for each trial type.

[Insert Figure 3 about here]

Overall, and consistent with previous research, bias was significantly larger when participants used the protagonist's false belief to indicate the first location in which the object was hidden (Mdn_{bias} = 3.7%) than when they used their own memory to indicate this location (Mdn_{bias} = 0.6%), Z = 2.137, p = .033, r = .17). Of the 40 participants, 25 (63%) showed a positive bias score (greater bias on false belief than memory trials), though this proportion was not significant, $\chi^2(1, N = 40) = 2.5$, p = .114. However, overall bias was significantly different from zero on false belief trials, Z = 4.274, p < .001, r = .48, but not on memory trials, Z = 1.465, p = .143, r = .16.

We next analysed the data split by length of move (short vs. long). In long trials, participants were significantly biased towards the object's second location when asked where the protagonist would search (Mdn_{bias} = 2.0%, Z = 2.272, p = .023, r = .25) but not biased when asked to remember the object's first location (Mdn_{bias} = 0.3%, Z = 0.941, p = .347, r = .11). Bias was marginally larger in the false belief trials (Z = 1.707, p = .088, r = .19). Of the 40 participants, 24 (60%) showed a positive bias score (greater bias on false belief than memory trials), though this proportion was not significant, $\chi^2(1, N = 40) = 1.6$, p = .206.

In short distance trials, participants were biased towards the second location that the object was hidden in when asked where the protagonist would search (Mdn_{bias} = 3.3%, Z = 2.782, p = .005, r = .31) but not when asked to remember the first location in which the object was hidden (Mdn_{bias} = -1.6%, Z = 0.349, p = .727, r = .04). The bias shown in the false belief trials was not significantly larger than the bias found in the memory trials (Z = 1.546, p = .122, r = .17). Of the 40 participants,

26 (66%) showed a positive bias score (greater bias on false belief than memory trials), a proportion of the total *N* that differed marginally from chance, $\chi^2(1, N = 40) = 3.6$, $p = .058$.

[Insert Figure 4 about here]

Discussion

Consistent with the results of previous research with young adults using the Sandbox Task (e.g. Coburn et al., 2015; Sommerville et al., 2013), we found evidence that participants tended to be more biased by their own current knowledge when making judgements about where a protagonist with a false belief would search for an object than when asked to remember where an object was hidden. A separate analysis that compared the number of participants who showed more bias on false belief trials to those participants who displayed more bias on memory trials, while not reaching statistical significance, confirmed that this pattern of results was not the result of extreme biases in a minority of individuals. As in some previous studies (e.g. Sommerville et al., 2013), we found some evidence that bias was more reliably present for one distance than another—in this case long distance trials. This may be due to the loss of statistical power when splitting the trials by distance (creating only two per condition), since when both short and long trials were analysed together the effect of bias was significantly greater on false belief than memory trials. In a different analysis, we found that bias only reached statistically significant levels when tested against zero on false belief trials, but not on memory trials. This effect was true regardless of distance, and this kind of analysis has not always been reported in previous research. In sum, we replicated the finding that participants are biased by their own knowledge of the location of an object when asked to reason about another's false belief but not when merely asked to recall the location from memory. These results support the Representation Contrast, according to which bias towards one's own knowledge is stronger when this knowledge is contrasted with another's false belief than when it is contrasted with functionally identical but semantic content, such as in memory.

In Experiment 2 we sought to test whether such bias is specific to mental states or extends to other forms of representation (the Mentalizing Contrast). We chose to employ a 'false film' condition to test this hypothesis. Although photographs have been used as controls in previous studies (Leekam & Perner, 1991; Leslie & Thaiss, 1992; Slaughter, 1998; Zaitchik, 1990), we here use a film sequence. Unlike photographs, film is not restricted to capturing a single moment, and is thus more closely matched to the belief condition in which the protagonist witnesses a scenario unfold over a period of time. Film is also able to represent the location of an object that is eventually unseen, as it captures the action of hiding it at the location, whereas a photo would be required to represent the brief moment in which the object is half-buried and still in the process of being hidden in order for the relevant location to be rendered visually. Finally, alternative formats such as false notes, maps or signs are less suitable for scenarios in which responses are indicated along a continuum rather than in discrete locations.

Since the stories accompanying the slides in Experiment 1 were not created with the potential inclusion of non-mental representations in mind, we created new vignettes which were designed to make the scenarios more ecologically valid in the context of hiding objects, and more suitable for the inclusion of a condition in which the scene was filmed. Critically, the only difference between the false belief and false film trials was whether the question at the end of the trial asked about the content of a protagonist's belief or the content of a film (see Figure 5). Otherwise the text was identical across condition. In order to avoid increasing the length of the task, all trials in Experiment 2 were belief or film trials (i.e. we excluded memory trials).

Experiment 2

Method

Participants

Eighty-nine participants were recruited and were pseudo-randomly assigned to start with belief trials (N = 46) or film trials (N = 43). Thirty-four participants did not meet our series of attention and

motivation checks (see supplemental information for details) such that the final sample size was 55 including 26 participants assigned to the Belief-Film order and 29 participants in the Film-Belief order.

Procedure

In this experiment, participants were given a scenario and were instructed at the very end to indicate wither where a protagonist believed an item to be, or where a camera than had been used to film the scene would depict the item. In this way, we sought to contrast the degree of bias that would be elicited from mental (belief) and non-mental (film) representations, and hence test the Mentalizing Contrast. An example scenario (scenario 3, in this experiment) began as follows: "Rebecca and Steve are in the restaurant kitchen. Rebecca has the tips jar, and Steve is filming her with a camera. While Steve is watching and filming her, Rebecca buries the tips jar in the freezer here. Steve then goes outside with the camera to smoke a cigarette." On the second slide, the scenario continues: "While Steve is outside, Rebecca digs the tips jar out and hides it here. She puts everything else in the freezer back where it was so it looks undisturbed." Finally, after the intervening word search, the test screen appeared with the following introductory text over the empty trough: "After a while, Steve comes back, still holding the camera." This line was included regardless of condition to ensure that the context in which the test question was embedded was equivalent across both belief and film trials. Depending on the condition, the test question was either "Where does he think the tips jar is?" (belief trial) or "Where will the camera show the tips jar is?" (film trial). Each participant either received four false belief trials followed by a true belief trial or four film trials followed by a true belief trial, before completing four trials of the other condition. The order of the scenarios was always constant such that all scenarios that were followed by false belief questions for participants in one trial type order (e.g. Belief-Film) would be followed by film questions for participants assigned to the other trial type order (e.g. Film-Belief) (see Table 1). In all other respects, the experiment followed precisely the same procedure as that in experiment 1.

[Insert Figure 5 about here]

Results

Figure 6 displays the results for each trial type[2], and Figure 7 the distribution of mean responses for each trial type. One scenario (trial 3, short distance) was removed from both the belief and film analyses owing to an error in the wording that rendered this item confusing. Retaining this item would not have altered the balance of bias between film and belief trials.

[Insert Figure 6 about here]

The order in which participants completed the task (false belief trials first vs. false film trials first) did not cause aggregate bias scores on either the false belief ($U(55) = 277$, $p = .092$[3]) or false film ($U(55) = 281$, $p = .106$) trials to differ; so we analysed the data collapsed over this factor. Overall, bias on false belief trials ($Mdn_{bias} = 1.3\%$) did not differ significantly from bias on false film trials ($Mdn_{bias} = 6.1\%$), $Z = 1.341$, $p = .180$, $r = .13$). Of the 55 participants, 29 (53%) showed greater bias on false belief than false film trials, and 26 (47%) the inverse, meaning that the numbers of participants who showed more bias on false belief trials did not differ from the number of participants who showed more bias on false film trials, $\chi^2(1, N = 55) = .164$, $p = .686$. Additionally, bias was significantly different from zero on both false belief trials, $Z = 2.354$, $p = .019$, $r = .22$, and false film trials, $Z = 4.064$, $p < .001$, $r = .39$).

[2] The data were again not normally distributed, and we report the results of Wilcoxon Signed-Rank tests as a result.

[3] Bias was marginally larger on belief trials if they came before film trials, but since this was non-significant and crucially did not suggest the 'anthropomorphising' of the camera on film trials (since order only showed any effect on belief), we collapsed over order.

In long distance trials, participants were marginally biased towards the object's second location when asked about the protagonist's belief (Mdn_{bias} = 0.1%, Z = 1.944, p = .052, r = .19) and significantly biased when asked to indicate where the film would depict the object's first location (Mdn_{bias} = 1.2%, Z = 2.991, p = .003, r = .29). Bias in the false film condition was marginally larger than bias in the false belief condition (Z = 1.684, p = .092, r = .16). Of the 55 participants, 30 (55%) showed greater bias in the false film condition than false belief condition. The numbers of participants who showed more bias on false belief trials did not differ from the number of participants who showed more bias on false film trials, $\chi^2(1, N = 55) = .455$, p = .5.

In short distance trials, participants were not significantly biased towards the second location where the object was hidden when asked about the protagonist's belief (Mdn_{bias} = 0.4%, Z = 1.408, p = .159, r = .13), but were marginally biased when asked to indicate where the film would depict the object's first location (Mdn_{bias} = 0.8%, Z = 1.835, p = .067, r = .17). There was no significant difference in bias between these two conditions (Z = 0.670, p = .503, r = .06). Of the 55 participants, 31 (56%) showed greater bias on false film trials than false belief trials. Again, the numbers of participants who showed more bias on false belief trials did not differ from the number of participants who showed more bias on false film trials, $\chi^2(1, N = 55) = .891$, p = .345.

[Insert Figure 7 about here]

Discussion

The results of Experiment 2 support the main finding of Experiment 1, namely that participants' judgement about where the hidden object is represented in a belief is biased by participants' knowledge of the true location of the object. However, this bias extended to a condition in which participants were asked to judge where a film would show the object hidden, such that bias was significant in both conditions and did not differ between conditions. This suggests that this bias is not exclusive to reasoning about other people's mental states but is also evident when reconstructing a

location based on a non-mental representation. Since each text in Experiment 2 was matched until the critical question on the final slide, and since the same participants responded to trials of both mental and non-mental representations, we controlled for the possibility that artefacts related to a between-subjects design or differences in the wordings of the stories could account for the this finding. Overall, the results of Experiment 2 fail to support the Mentalizing Contrast; participants were as biased towards their own knowledge when reasoning about a film as when reasoning about another's mental state.

General Discussion

The use of non-mental representations as controls has previously revealed evidence that the role of specific ToM processes in eliciting egocentric biases may have been overestimated (e.g. Santiesteban et al., 2015; Zaitchik, 1990). In the present study, we argued that for a mental-state specific account of egocentric bias to hold, the bias must be manifested not only when contrasted against reasoning from semantic content (the Representation Contrast) but also when reasoning about non-mental representations (the Mentalizing Contrast). The results from the two experiments presented here show that, in the Sandbox Task, participants are biased towards where they know an object to be regardless of whether they are attempting to indicate the object's location as represented by a mental or a non-mental representation but are not biased when using their own memory to indicate the object's location. In other words, we found evidence for the Representation Contrast, but not for the Mentalizing Contrast. Thus, the levels of bias found in the original Sandbox Task and the present study would appear not to be linked to a domain-specific process, such as reasoning about other people's belief states, but to constraints on representations in general (e.g. Iao et al., 2011; Zaitchik, 1990).

The suggestion that the bias in the Sandbox Task is not specific to the attribution of beliefs also receives support from recent evidence that children with Autism Spectrum Disorder (ASD) show a similar degree of bias in the Sandbox Task to typically developing children (Begeer et al., 2016).

According to ToM-based accounts, children with impaired mentalizing abilities (such as those with ASD) should exhibit a greater bias than typically developing controls. Thus, together with the current experiment it would appear that the source of bias in the Sandbox Task is unlikely to be the result of mentalizing processes.

Importantly, this account of participants' bias in the Sandbox Task holds up to a number of counter-arguments that have been applied to similar claims. For instance, it has been argued that in certain cases stimuli used as non-social controls may be treated as agents ('anthropomorphised') and end up being processed by the Theory of Mind system (e.g. Furlanetto, Becchio, Samson, & Apperly, 2016). Should this be the case, this process is most likely to have influenced the results of Experiment 2, where half the participants first received trials where they had to reason about a protagonist's belief and then received trials where they had to reason about a film's representation of a hidden object's location. This could have induced a carry-over effect whereby participants continued to treat the film trials as if they were belief trials. However, the order of presentation (belief followed by film or film followed by belief) showed no influence on levels of bias on film trials.

A different, though related, argument might suggest that the film condition did not consist of a truly non-mental state, since the camera was operated by the protagonist, and the film might be thought of as 'watched' by the protagonist in order to answer the question "Where will the camera show…". This is a common issue in non-mental control conditions in ToM tasks, as whatever the format (whether false film, false photo or false belief) it is not possible to conceive of a photo, note or film without imagining that a human, whether the protagonist or the participant, is 'seeing' the item in question, and hence non-mental formats can be seen in some sense as being 'filtered' through human perception. A corollary to this potential criticism is that having the protagonist operate the camera introduces an extra degree of such 'filtering'. While we would very much take on board these objections at a theoretical level, seeing them as an inevitable consequence of having non-mental controls, there are a number of *empirical* reasons to believe that this is not how participants treated this condition. Firstly, it is a common finding in the theory of mind literature that participants follow instructions to reason from a specific representation, even when following this processing route is not

the most efficient strategy. For example, Cohen et al. (2015) found that participants were slower when instructed to reason from false notes than false beliefs, and Cohen and German (2010) found that participants were slower when instructed to reason from false maps and 'false' arrows, than false beliefs. Crucially, the outcome of the participants' reasoning was identical regardless of whether they were asked to reason from a mental or non-mental state, yet a behavioural dissociation was clear. Evidence from neuroimaging experiments also supports differential processing of mental states and non-mental states as conveyed through purely textual instructions (such as ours), such as when contrasting "On Peter's holiday photo…?" (false photo trials) with "The tourists now think…?" (false belief trials) (Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006). In the case of the holiday photo, 'Peter' was clearly identified as the one who took the photo, yet participants showed both a behavioural and neurological dissociation between false photo trials and false belief trials, suggesting participants did not apply mental state reasoning to false photos, but rather made their responses according to the specific wording of the instructions (see Schurz, Radua, Aichhorn, Richlan, & Perner, 2014, for a recent meta-analysis showing clear differences between the way participants respond to mental and non-mental states in tasks employing text-based vignettes such as ours). Most importantly, there is direct evidence for precisely such a behavioural dissociation from the difference in bias we found between memory trials and belief trials in experiment 1. The only reason for this must be that participants were working according to the wording of the critical question to either reason from a memory or the protagonist's belief. Our decision to exclude from the analyses data from participants who did not show the requisite attention to the wording of the true belief trial was motivated precisely by the desire to ensure that participants were indeed following instructions on a trial-by-trial basis rather than adopting a strategy of 'skipping' the text and continually indicating the first location in which the objects was hidden. It is particularly convincing, in our view, that participants across all these studies did not appear to take the opportunity to *ignore* the belief element of the belief trials and instead follow a strategy of basing their decisions on their memory alone (in the case of our experiment) or beliefs alone (in the case of Cohen and colleagues' and Perner and colleagues' work), because doing so would actually have led to more accurate/faster responses. Our results, together with findings of other research in the literature, strongly suggest that when

participants were instructed to indicate where the camera will show the item, this is very likely how they arrived at their response."

An alternative account for the apparent difficulty of reasoning from others' beliefs as opposed to one's own memories (our experiment 1) comes from the literature on the role of self-other differentiation on perspective taking[4]. For example, it has been shown that participants are less likely to be biased by their own knowledge when judging an out-group member's belief than an in-group member's belief, or when they have been primed to a distinction between themselves and other people (Santiesteban, White, Cook, Gilbert, & Heyes, 2012; Todd, Hanko, Galinsky, & Mussweiler, 2011). Although we did not manipulate in-group and out-group membership when designing our experiment, there are three ways in which the self-other distinction might be primed in our studies, and might therefore have contributed to egocentric bias. Firstly, in every case, and whether the trial was a memory trial, belief trial or a film trial, two 'others' were described in the description, usually with accompanying proper names (e.g. Sally and Ann, Judy and Judy's dad). This makes clear that the events are happening to 'named' others external to the participant's own direct experience, and hence more bias may be produced as a result. Alternatively, others' beliefs are not our own, but our memories are, and this distinction in experiment 1 may itself prime a self-other differentiation. By this account, more bias is introduced because the specific instruction that the participant is asked to respond to regards another individual (the other) rather than the participant (the self). Thirdly, when participants are asked to reason from another's mental state specifically, the self-other distinction is primed more than when they are asked to reason from a non-mental state or a memory. However, three arguments militate against the idea of the self-other distinction *alone* accounting for bias concerns the scenarios themselves. Firstly, if the self-other distinction is primed by references to the protagonists in the text more generally, then we should not expect any difference in the levels of bias between any condition, since the self-other distinction is equivalent in each case. This is contradicted by the results of experiment 1, which found more bias on belief trials than memory trials, despite the scenarios always referring to two others who experienced the events in question. Secondly, if the self-

[4] We are grateful to an anonymous reviewer for raising this point.

other distinction is made most salient when the reference to the protagonist occurs not just in the scenario preamble but in the critical question on the final slide, then we should see less bias on those occasions where the protagonist was *not* referred to in the test question itself, namely on the film trials in experiment 2. Although these trials mentioned the protagonist in the final slide (e.g. "Then Steve comes back, still holding the camera"), they omitted the protagonist from the critical test question (e.g. "Where will the camera show…?"). Nevertheless, bias did not differ between belief questions which did include the protagonist in the question, and these film trials which did not. Further evidence comes from experiment 1, in which the test questions *always* referred to the protagonist, regardless of whether that trial was a memory trial (e.g. 'Where did *she* put the red toy dog…?") or a belief trial (e.g. "Where is *he* going to look for the ice cream?') (italics added). Despite the presence of the protagonist in each case, participants showed more bias on belief trials than memory trials. Since in both cases the participant is explicitly instructed to take into account either what the protagonist did, or what the protagonist thinks, any remaining difference in bias between memory and belief trials is therefore more likely to be attributable to the *type* of information being reasoned from, rather than the individual character per se. Finally, if reasoning from another's belief is what primes the self-other distinction, then we should expect a difference between belief trials and film trials in experiment 2. No such difference was evident. It would appear therefore that it is not the self-other distinction but rather the type of the information being reasoned from that is most important in generating differences in egocentric bias. Crucially for present purposes, it would therefore appear that the primary outcomes of the present research, namely that reasoning from representations elicits more bias than reasoning from semantic content, and that non-mental representations are capable of inducing bias, and evidently *as much* bias, as mental representations, are not undermined. Nevertheless, the self-other distinction remains an interesting avenue of research, and it would be useful to clarify its role (or otherwise) in generating egocentric bias via more direct testing. One possibility would be to ask participants to indicate where the object was according to their *own* memory, and contrast this with trials in which they were instructed to indicate where the object was in the *protagonist's* memory. Presumably, any distinction between the degrees of bias on one trial type or the other would therefore

be specific to a self-other distinction, and this could then form a baseline for further contrasts of interest.

The current results also go some way towards elucidating the mechanism that may be causing adults' biases when attributing beliefs. There are two accounts that the results of the present study do not appear to support. Firstly, within the developmental psychology literature, children's difficulty with the false belief task has sometimes been ascribed to a difficulty in understanding that another person may have a different belief to themselves, and sometimes it has been ascribed to a difficulty understanding that beliefs, which are intended to be true representations, may sometimes be false (e.g. Bloom & German, 2000). Comparing performance when participants reason about beliefs and film allows for these theoretical positions to be distinguished because beliefs and films differ in whether or not they are intended to be true. Beliefs are true if they accurately represent reality, or false if they do not. In contrast, films that do not represent a current reality may be considered outdated, but because they are not intended to represent reality they are not false (e.g. a photograph depicting a person on a beach is not considered false because that person is now at work, cf. Perner & Leekam, 2008). Our results suggest that the bias found in the Sandbox Task is not the result of adults finding it difficult to process representations that are intended to be true to reality, such as beliefs.

Additionally, the results presented here would appear at least on first view not to support the 'curse of knowledge' account (e.g. Birch & Bloom, 2004; 2007), which predicts that our judgments of other people's naive states are contaminated by what we ourselves know. Firstly, in some instantiations the account is also associated with bias when reasoning not only about another's mental state but also about one's own naïve or uninformed past state (e.g. Birch & Bloom, 2004). Given that memory trials are linked to an uninformed past state (in the context of present reality), the curse of knowledge account could predict bias for those trials. Since we found no evidence of bias on memory trials, this account would need to treat these trials as *informed* (as opposed to *un*informed) or *non*-naïve past states. This may well be the case, since the memory trials in the present study did not refer to information that is later revealed to be untrue, but simply becomes outdated after the item is moved. However, if we accept this characterisation of memory trials, it is unclear how the curse of knowledge

account could explain the clear evidence of bias on film trials, which clearly shares the property of being outdated.

Instead, our results appear to support the explanation that it is representations *per se*, mental or non-mental, that are difficult to process when they differ from reality. The results of the first false photo study by Zaitchik (1990) led her to suggest this as a candidate explanation for children's performance on false belief tasks. Nevertheless, we do not claim that both films and beliefs are therefore processed by one and the same mechanism. Behavioural dissociations between false photos and false beliefs have been demonstrated in related tasks in both children and adults (Perner et al., 2006; Saxe & Kanwisher, 2003; Slaughter, 1998; Zaitchik, 1990), and are further backed up by evidence of dissociations in the regions of the brain recruited for each format (see Schurz et al, 2014, for a review). Cohen and colleagues (2015) recently suggested that there may indeed be a domain-specific process for mental representations that, though privileged for this format, can be extended to the processing of non-mental representations (the 'by-product account'; see also Cohen & German, 2010, and Bowler, Briskman, Gurvidi, & Fornells-Ambrojo, 2005). They posit that such a mechanism may be co-opted by non-mental representations such as photographs, but at a cost. In Experiment 2 we found some —albeit weak — evidence of greater bias on false film than false belief trials which, though limited in strength, is consistent in direction with some previous research (e.g. Cohen & German, 2010; Cohen et al., 2015; Perner et al., 2006; Saxe & Kanwisher, 2003; Zaitchik, 1990). Our evidence would appear at a minimum not to contradict the relative difficulty of non-mental representations that the account predicts. Overall, we take the findings of the present study only to support the hypothesis that egocentric bias is unlikely to be unique to the processing of mental state representations.

In sum, our results not only corroborate previous findings that adults are egocentrically biased when indicating the location of objects on the basis of another's false belief, but go on to show that this egocentric bias occurs even when adults indicate the location of objects on the basis of false non–mental representations. Consequently, the current experiments provide evidence that adults'

egocentric difficulties when reasoning about others' false beliefs are not specific to processes requiring Theory of Mind.

Acknowledgements

We are grateful for comments on a draft of this paper by Ljerka Ostojic and Idalmis Santiesteban. This work was supported by an ESRC grant (ES/M008460/1) (SS, RL and NC) and by a Leverhulme grant (RPG-2014-353) (EL and NC).

Disclosure of interest

The authors report no conflicts of interest.

References

Apperly, I. (2010). *Mindreaders: the cognitive basis of "theory of mind"*. New York: Psychology Press.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953.

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: The MIT Press.

Begeer, S., Bernstein, D. M., Aßfalg, A., Azdad, H., Glasbergen, T., Wierda, M., & Koot, H. M. (2016). Equal egocentric bias in school-aged children with and without autism spectrum disorders. *Journal of Experimental Child Psychology*, *144*, 15–26.

Bernstein, D. M., Atance, C., Loftus, G. R., & Meltzoff, A. (2004). We saw it all along visual hindsight bias in children and adults. *Psychological Science*,*15*(4), 264–267.

Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of mind through the ages: Older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Experimental Aging Research*, *37*(5), 481–502.

Birch, S. A., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences*, *8*(6), 255–260.

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*(5), 382–386.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*(1), B25–B31.

Bowler, D. M., Briskman, J., Gurvidi, N., & Fornells-Ambrojo, M. (2005). Understanding the mind or predicting signal-dependent action? Performance of children with and without autism on analogues of the false-belief task. *Journal of Cognition and Development*, *6*(2), 259–283.

Carruthers, P. (2015). Mindreading in adults: Evaluating two-systems views. *Synthese*, 1–16.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives On Socially Shared Cognition*, *13*(1991), 127–149.

Coburn, P. I., Bernstein, D. M., & Begeer, S. (2015). A new paper and pencil task reveals adult false belief reasoning bias. *Psychological Research*, *79*(5), 739–749.

Cohen, A. S., & German, T. C. (2010). A reaction time advantage for calculating beliefs over public representations signals domain specificity for 'theory of mind'. *Cognition*, *115*(3), 417–425.

Cohen, A. S., Sasaki, J. Y., & German, T. C. (2015). Specialized mechanisms for theory of mind: Are mental representations special because they are mental or because they are representations? *Cognition*, *136*, 49–63.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental science*, *13*(2), 331–338.

Dumontheil, I., Hillebrandt, H., Apperly, I. A., & Blakemore, S. J. (2012). Developmental differences in the control of action selection by social information. *Journal of Cognitive Neuroscience*, *24*(10), 2080–2095.

Dumontheil, I., Küster, O., Apperly, I. A., & Blakemore, S. J. (2010). Taking perspective into account in a communicative task. *Neuroimage*, *52*(4), 1574–1583.

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 158–163.

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science*, *317*(5843), 1351–1354.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (eds.) *Syntax and semantics volume 3: Speech acts* (pp. 41–58). New York: Academic Press.

Heyes, C. (2014a). False belief in infancy: a fresh look. *Developmental Science*, *17*(5), 647–659.

Heyes, C. (2014b). Submentalizing I am not really reading your mind. *Perspectives on Psychological Science*, *9*(2), 131–143.

Huttenlocher, J., Newcombe, N., & Sandberg, E. H. (1994). The coding of spatial location in young children. *Cognitive Psychology*, *27*(2), 115–147.

Iao, L. S., & Leekam, S. R. (2014). Nonspecificity and theory of mind: new evidence from a nonverbal false-sign task and children with autism spectrum disorders. *Journal of Experimental Child Psychology*, *122*, 1–20.

Iao, L. S., Leekam, S., Perner, J., & McConachie, H. (2011). Further evidence for nonspecificity of theory of mind in preschoolers: Training and transferability in the understanding of false beliefs and false signs. *Journal of Cognition and Development*, *12*(1), 56–79.

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 179.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, *7*(2), 46–50.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

Leekam, S. R., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition*, *40*(3), 203–218.

Leekam, S., Perner, J., Healey, L., & Sewell, C. (2008). False signs and the non-specificity of theory of mind: Evidence that preschoolers have general difficulties in understanding representations. *British Journal of Developmental Psychology*, *26*(4), 485–497.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review*, *94*(4), 412.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, *8*(12), 528–533.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, *43*(3), 225–251.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience*, *1*(3–4), 245–258.

Perner, J., & Leekam, S. (2008). The curious incident of the photo that was accused of being false: Issues of domain specificity in development, autism, and brain imaging. *The Quarterly Journal of Experimental Psychology*, *61*(1), 76–89.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*(3), 279–301.

Santiesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic Bulletin & Review*, *22*(3), 844–849.

Santiesteban, I., White, S., Cook, J., Gilbert, S. J., Heyes, C., & Bird, G. (2012). Training social cognition: from imitation to theory of mind. *Cognition*, *122*(2), 228-235.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". *Neuroimage*, *19*(4), 1835–1842.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34.

Slaughter, V. (1998). Children's understanding of pictorial and mental representations. *Child Development*, *69*(2), 321–332.

Sommerville, J. A., Bernstein, D. M., & Meltzoff, A. N. (2013). Measuring beliefs in centimeters: Private knowledge biases preschoolers' and adults' representation of others' beliefs. *Child Development*, *84*(6), 1846–1854.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, *17*(1–2), 3–23.

Todd, A. R., Hanko, K., Galinsky, A. D., & Mussweiler, T. (2011). When focusing on differences leads to similar perspectives. *Psychological Science*, *22*(1), 134-141.

Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, *29*(9), 1159–1168.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, *72*(3), 655–684.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition*, *35*(1), 41–68.

Tables.

Table 1. Order and type of trials.

	Exp. 1	*Exp. 2*		*Direction*	*Distance*	*Location A*	*Location B*
Trial	*Belief*	*Belief-Film*	*Film-Belief*	*Direction*	*Distance*	*(% from left)*	*(% from left)*
1	Memory	False Belief	False Film	right	short	50.2 (6)	72.8
2	Memory	False Belief	False Film	left	long	80.2 (7)	34.9
3	False Belief	False Belief[5]	False Film[5]	left	short	40.1 (1)	17.5
4	False Belief	False Belief	False Film	right	long	30.1 (2)	75.4
5	True Belief	True Belief	True Belief	right	long	33.4 (5)	78.7

[5] Removed from analyses owing to wording error (see supplemental material for details).

6	False Belief	False Film	False Belief	left	long	70.2 (3)	24.9
7	False Belief	False Film	False Belief	right	short	60.1 (4)	82.8
8	Memory	False Film	False Belief	right	long	43.4 (8)	88.8
9	Memory	False Film	False Belief	left	short	73.5 (9)	50.9

Figures

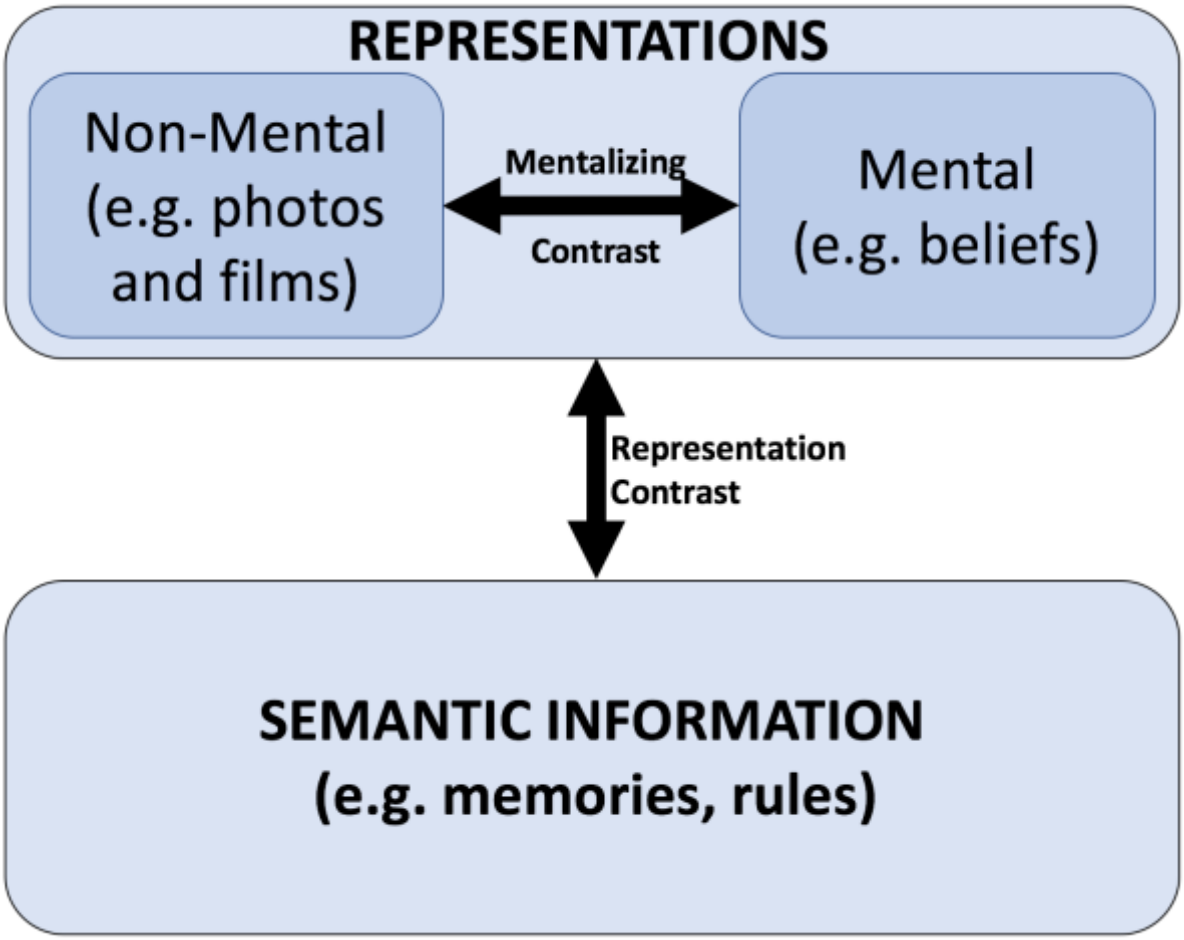

Figure 1. Schematic of the two contrasts that are required to support a mental-state specific account of egocentric bias. See main text for description.

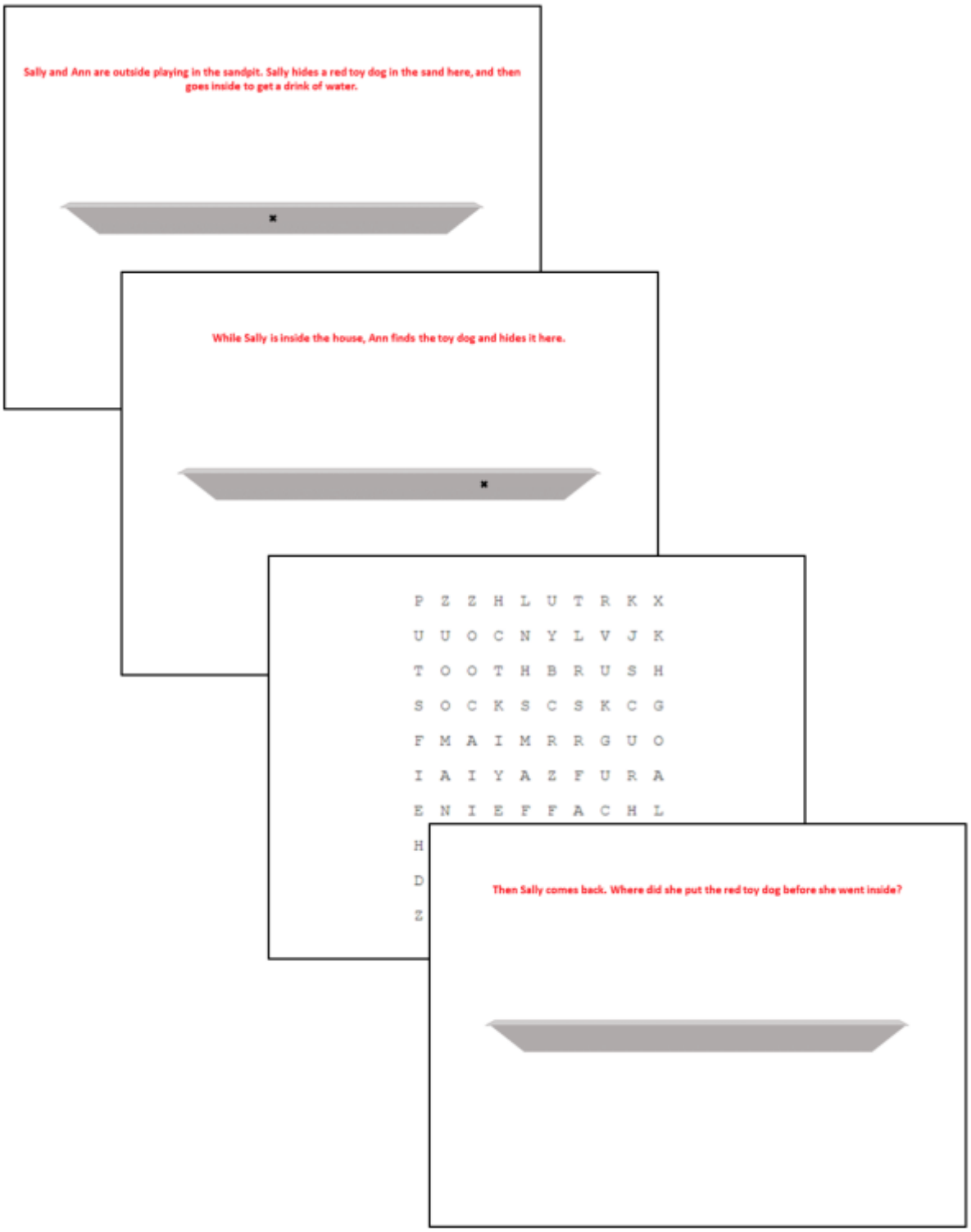

Figure 2. Slides from an example memory trial from Experiment 1. See text for details of procedure.

A full list of the texts and locations used are available in the supplemental information.

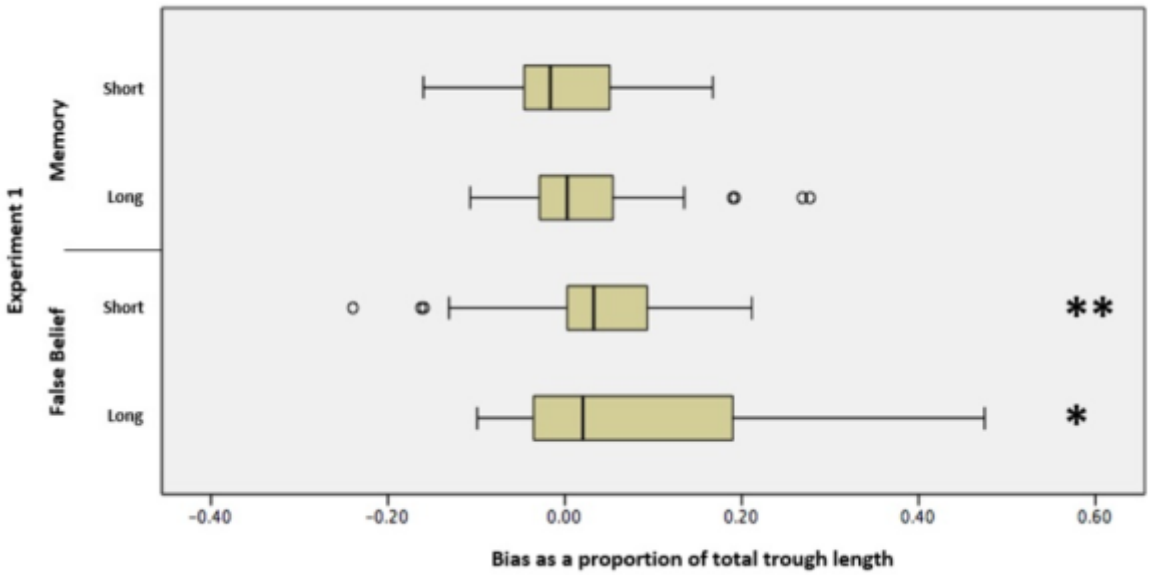

Figure 3. Bias as a proportion of the total trough size for Experiment 1. Outliers beyond 1.5 x IQ but within 3 x IQ indicated by circles. From a baseline of zero, * = bias significant at $p < .05$ level; ** = bias significant at $p < .01$ level.

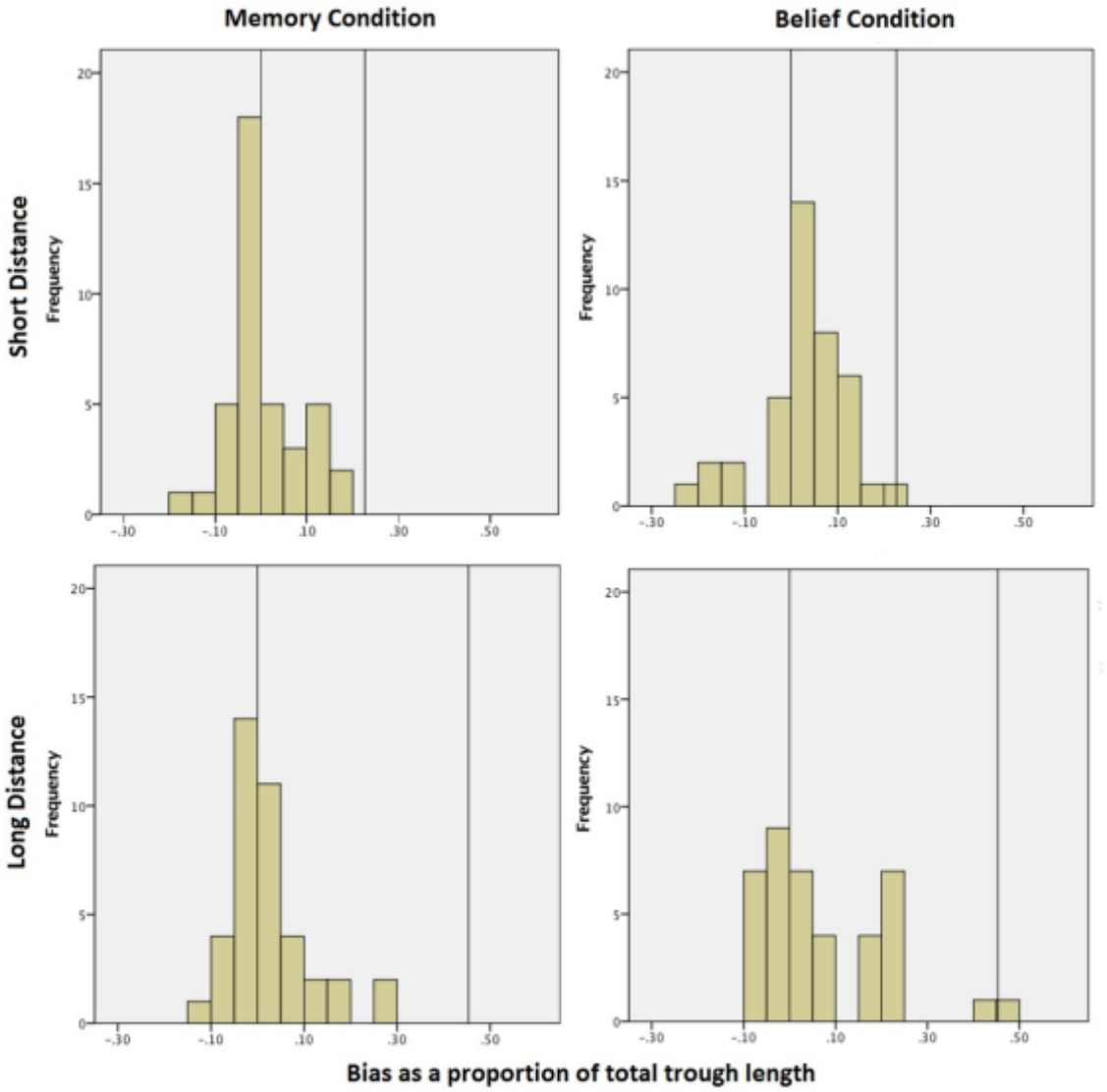

Figure 4. Distribution of mean locations as indicated by participants for each trial type in Experiment 1. The two reference lines refer to the first (false) location (always at point zero) and second (true) location.

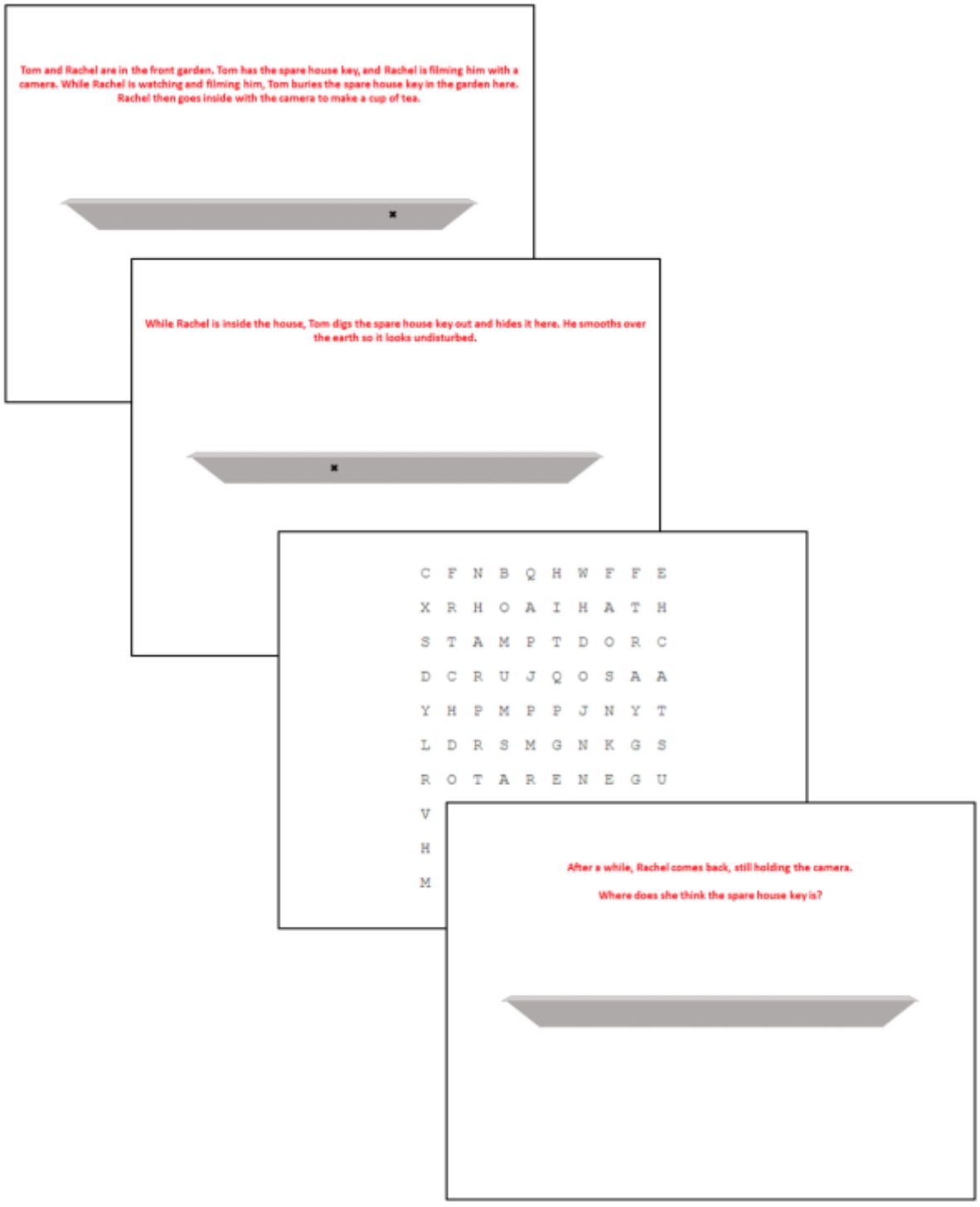

Figure 5. Slides from an example false belief trial from Experiment 2. In the false belief condition, the trial would differ only in that the critical question on the final slide (in this example, it read "Where will the camera show the spare house key is?". See text for details of procedure. A full list of the texts and locations used are available in the supplemental information.

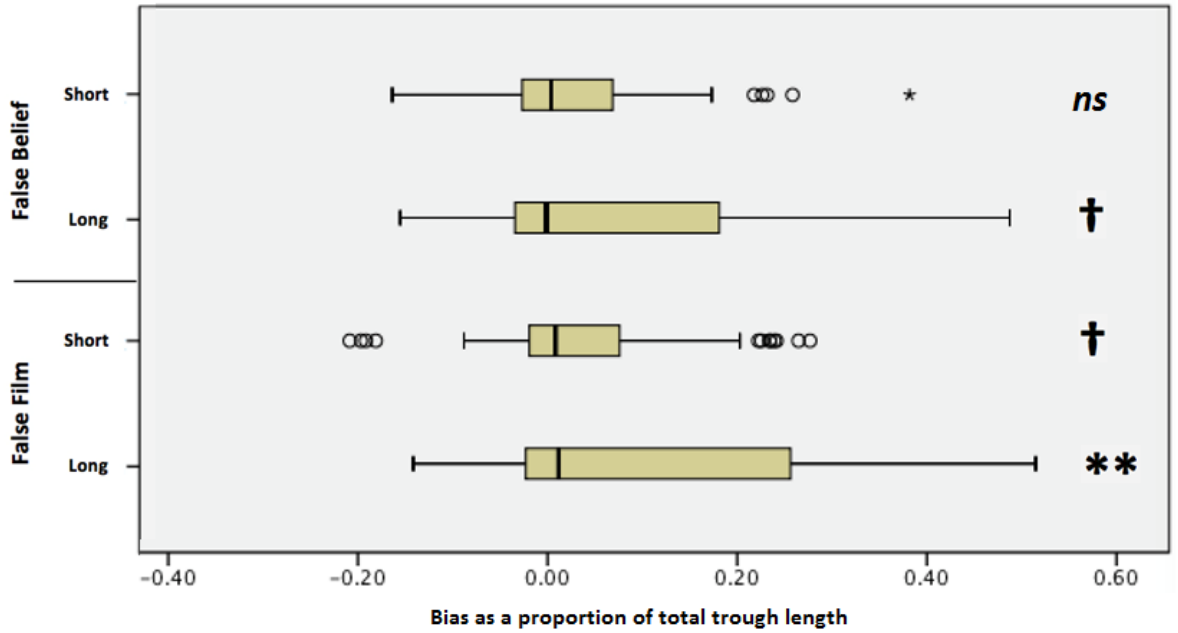

Figure 6. Bias as a proportion of the total trough size for Experiment 2. Outliers beyond 1.5 x IQ and 3 x IQ indicated by circles and stars respectively. From a baseline of zero, † = bias marginally significant ($p < .1$); ** = bias significant at $p < .01$ level.

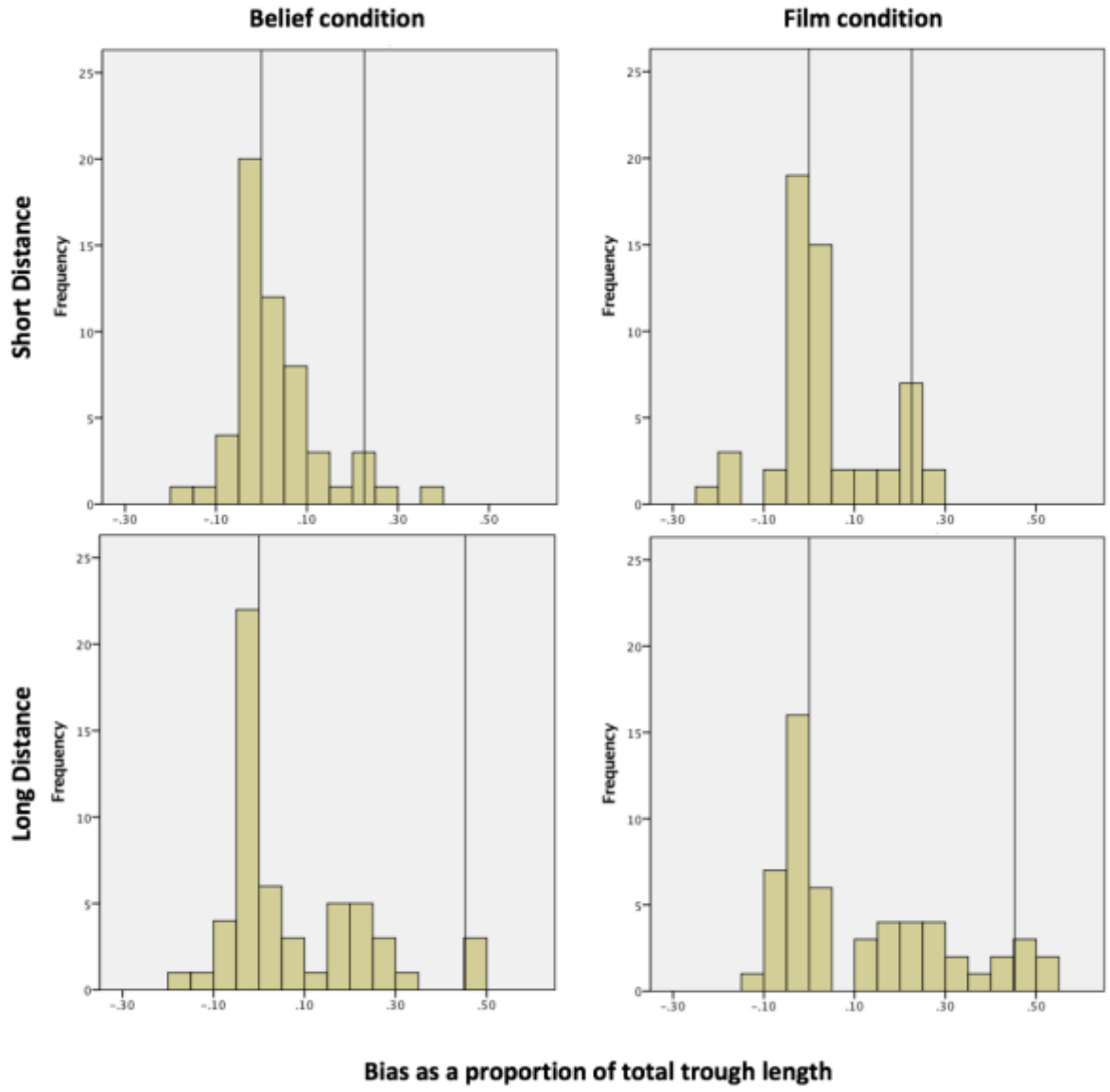

Figure 7. Distribution of mean locations as indicated by participants for each trial type in Experiment 2. The two reference lines refer to the first (false) location (always at point zero) and second (true) location.