Eye-tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task

Heather J. Ferguson [1]

Ian Apperly [2]

James E. Cane [3]

[1] University of Kent, England, UK

[2] University of Birmingham, England, UK

[3] London South Bank University, England, UK

Correspondence to:

Heather Ferguson

School of Psychology, Keynes College

University of Kent, Canterbury, CT2 7NP

Email: h.ferguson@kent.ac.uk

Phone: +44 (0)1227 827120

Fax: +44 (0)1227 827030

Word length: 7061 (excluding abstract and references)

Author Notes.

This work was carried out with the support of a grant to HF and IA from the Leverhulme Trust (Ref: RPG-063).

Abstract

Previous studies have shown that while people can rapidly and accurately compute their own and other people's visual perspectives, they experience difficulty ignoring the irrelevant perspective when the two perspectives differ. We used the 'avatar' perspective-taking task to examine the mechanisms that underlie these egocentric (i.e. interference from their own perspective) and altercentric (i.e. interference from the other person's perspective) tendencies. Participants were eye-tracked as they verified the number of discs in a visual scene according to either their own or an on-screen avatar's perspective. Crucially in some trials the two perspectives were inconsistent (i.e. each saw a different number of discs), while in others they were consistent. To examine the effect of perspective switching, performance was compared for trials that were preceded with the same *versus* different perspective cue. We found that altercentric interference can be reduced or eliminated when participants stick with their own perspective across consecutive trials. Our eye-tracking analyses revealed distinct fixation patterns for self and other perspective-taking, suggesting that consistency effects in this paradigm are driven by implicit mentalising of what others can see, and not automatic directional cues from the avatar.

Keywords: Theory of Mind, visual perspective taking, perspective switching, self/other, eye-tracking.

Introduction

Visual perspective-taking is commonly seen as a key component of people's ability to understand the mental states of others (their meaning, intentions, desires and knowledge), which is often described as Theory of Mind (ToM), mentalising, or mindreading. Visual perspective-taking is typically examined along one of two dimensions; one that simply assesses *what* someone else can see (termed 'level 1' perspective-taking), and another that requires participants to adopt someone else's spatial point of view to judge *how* that person sees a particular visual stimulus (termed 'level 2' perspective-taking). These two types of visual perspective-taking can be differentiated according whether or not they require one to mentally rotate into the position of the other person (Michelon & Zacks, 2006; Surtees, Apperly, & Samson, 2013). The current study focuses on level 1 visual perspective-taking, examining the cognitive mechanisms that underlie judgments about 'self' and 'other' perspectives.

Research investigating level 1 visual perspective-taking has frequently employed a task in which participants are presented with images of a 3D room, and have to verify the number of discs in that visual scene according to either their own or an on-screen avatar's perspective. Crucially, in some of the trials the two perspectives are inconsistent (i.e. each sees a different number of discs), while in others they are consistent. In the first study of this kind, Samson, Apperly, Braithwaite, Andrews, and Scott (2010) found that healthy adults can rapidly and accurately compute other people's visual perspectives, or respond according to their own broader viewpoint (which may include objects that are hidden form the avatar's view). Nevertheless, participants experienced difficulty ignoring the irrelevant perspective (i.e. either what they saw or what the avatar saw) when the two perspectives differed; performance on the task was influenced by both egocentric and altercentric tendencies. That is, participants' responses were slower and less accurate when judging what the avatar could see required them to inhibit their own visual perspective, and when judging what they could see required them to inhibit the avatar's visual perspective.

An egocentric or reality bias (Mitchell, Robinson, Isaacs, & Nye, 1996) has been frequently documented in the ToM literature- even among healthy adults- and can have the effect of delaying or even cancelling perspective-taking all together (e.g. Barr & Keysar, 2002; Birch & Bloom, 2007; Dumontheil, Küster, Apperly, & Blakemore, 2010; Epley, Morewedge & Keysar, 2004; Keysar, & Barr, 2005; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003). Indeed, some researchers have proposed that perspective-taking is initially biased to one's own knowledge, and that integration of other peoples' perspectives operates only as a subsequent and controlled correction mechanism (Epley, Keysar, Van Boven, & Gilovich, 2004). Evidence of altercentric intrusions is more limited, and has so far been found only in tasks that tap visual perspective-taking (Capozzi, Cavallo, Furlanetto, & Becchio, 2014; Furlanetto, Becchio, Samson, & Apperly, 2016; Kovács, Téglás, & Endress, 2010; Nielsen, Slade, Levy, & Holmes, 2015; Ramsey, Hanson, Apperly, Samson, 2013; Samson et al., 2010; Surtees & Apperly, 2012; Qureshi, Apperly, & Samson, 2010). These studies clearly show that observers can rapidly and involuntarily compute other people's visual perspectives, though recent research has shown that this spontaneous 'other' perspective-tracking is impaired when observers must track multiple discrepant viewpoints (Capozzi et al., 2014).

Taken together, this research shows that perspective-taking can occur automatically, and can be rapidly integrated into subsequent processing. However, there are also good reasons for thinking that adopting someone else's perspective is more cognitively effortful than simply referring to the self-perspective (Apperly et al., 2009, 2010; Birch & Bloom, 2007). For example, studies that have applied working memory load manipulations have shown that higher cognitive load impedes one's ability to infer other peoples' mental states (Bull, Phillips, & Conway, 2008; Cane, Ferguson, & Apperly, submitted; Lin, Keysar, & Epley, 2010; McKinnon & Moscovitch, 2007; Schneider, Lam, Bayliss, & Dux, 2012), and recent research has shown that individual differences in executive function predicts perspective-taking ability in complex tasks (Bradford, Jentzsch, & Gomez, 2015; Brown-Schmidt, 2009; Grodner, Dalini, Pearlstein-Levy, & Ward, 2012; Lin et al., 2010). Interestingly, ability to compute level-1 visual perspectives is not disrupted by a simultaneous load on executive function (Qureshi et al., 2010). This suggests that inferring other peoples' visual perspectives may tap into an automatic, more cognitively efficient system than other more complex ToM processes, such as belief reasoning and inferences from language, which rely on a more flexible but cognitively demanding system (Apperly & Butterfill, 2009).

The current study adapted Samson et al's (2010) 'avatar' visual perspective-taking task, and set out to address two main aims. First, we aimed to use eye-tracking to examine the cognitive mechanisms that underlie visual perspective-taking. Eye-tracking has been applied to this visual perspective-taking paradigm once before (Nielsen et al., 2015), but with a fairly rudimentary measure of gaze duration collapsed across a large area of interest that included the avatar and areas in front and behind the avatar (including the discs). This study showed that gaze durations were elevated when the self and other perspectives were inconsistent, thus reflecting both egocentric and altercentric interference. The social basis of this altercentric effect was further supported by a correlation with self-reported perspective-taking ability and empathy that was not present in conditions where the avatar was replaced with a semisocial (an arrow) or nonsocial (a dual-coloured block) central stimulus. Our study employs much finer grained analyses of visual attention. Specifically, we recorded the number and location of participants' fixations in the level 1 visual perspective-taking task to compare how people allocate their visual attention between the avatar's gaze location and the wall behind the avatar's gaze (the 'no-gaze' location), and how this differs when they have been prompted to take the self or other perspective. Though no previous studies have examined these visual biases in the visual perspective-taking task, related research has demonstrated that gaze direction provides a strong attentional cue in guiding eye movements toward the location of an actor's gaze (Borji, Parks, & Itti, 2014; Castelhano, Wieth, & Henderson, 2007). Crucially, this methodology should also provide a means of disentangling traditional mentalising accounts for spontaneous visual perspective-taking from directional (or sub-mentalising) accounts, which suggest that attention is driven by domain-general processes based on directional features of the avatar (Heyes, 2014; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). Specifically, Santiesteban et al. (2014) found a comparable reaction time difference between consistent and inconsistent trials when the central avatar was replaced with an arrow (but see Schurz, Kronbichler, Weissengruber, Surtees, Samson, & Perner, 2015 and Nielsen et al., 2015). Since arrows provide directional, but not agentive, cues the authors interpret this as evidence against an implicit mentalising account of the altercentric effect, and instead supports the role of attentional processes. Here, we use eye-tracking to examine how visual attention is directed around the scenes during the visual perspective-taking task, specifically examining the location of the first fixation in each scene. The directional account would predict that early visual attention automatically shifts to the dots in the avatar's field of view regardless of the perspective cue condition. In contrast, if the consistency effect is driven by implicit mentalising of what the avatar can see, we would expect to see modulation of the gaze location bias depending on whether participants were cued to take their own/the avatar's visual perspective. Therefore, we tested the implicit mentalising prediction that participants would show a reduced bias to fixate the avatar's gaze location when they adopted their own perspective compared to when they took the avatar's, as participants should divide their attention between the two possible locations (i.e. the gaze location and no-gaze location).

The second aim in our study was to test how switching between self and other perspectives across consecutive trials influences performance on a level 1 visual perspective-taking task. Neuroimaging research has demonstrated that mental state attributions about the self and others engage distinct brain mechanisms (e.g. Decety & Sommerville, 2003; Jeannerod & Anquetil, 2008; Samson, Apperly, Kathirgamanathan, & Humphreys, 2005; Saxe, Moran, Scholz, & Gabrieli, 2006; Vogeley et al., 2001), as well as specific regions that respond to conflicts between self and other states (e.g. McCleery, Surtees, Graham, Richards, & Apperly, 2011). Thus, switching between self and other perspectives is likely to require disengagement of one perspective to adopt the other. Nevertheless, rapidly switching between self and other perspectives depending on context is one of the key processing steps involved in successful everyday social cognition; children's development of ToM correlates with their cognitive flexibility (Hughes, 1998). Samson et al. (2010) examined perspective shifting indirectly in their paper by comparing performance across their experiments, which varied whether perspective shifting between self and other viewpoints was manipulated within (Experiment 1) or between blocks (Experiment 2), and when no shift of perspective was required (participants only attended to their own perspective throughout the task, Experiment 3). Results showed that egocentric and altercentric biases were present regardless of whether participants had to shift between perspectives within a block or not, though self-perspective judgments were faster when repeatedly judging one's own perspective. This supports the idea that intrusions from an irrelevant visual perspective occur automatically, but suggests that this interference may be greater when one needs to switch between the self and other perspective. Similarly, a recent study that manipulated self/other perspectives in a variant false belief task found that participants responded faster according to the self than other perspective, but only when a perspective shift had occurred within the narrative (Bradford, Jentzsch, & Gomez, 2015). Together these findings support a dissociation between self and other attributions, and suggest that switching perspectives is cognitively effortful. In the current study, we compared performance between consecutive trials that tapped the same perspective (i.e. (*you*-) *YOU* or (*they*-) *THEY*; referred to as 'stick' trials here) and trials that tapped different perspectives (i.e. (*you*-) *THEY* or (*they*-) *YOU*; referred to as 'switch' trials here).

In sum, the current study examined level 1 visual perspective-taking, and manipulated three key variables: whose visual perspective to take on a trial (self *vs.* other), the consistency of visual perspective between the participant and avatar (consistent *vs.* inconsistent), and the presence or absence of a perspective switch between consecutive trials (switch *vs.* stick). We recorded participants' response accuracy, reaction times and eye movements as a measure of their egocentric (i.e. interference from their own perspective) and altercentric (i.e. interference from the other person's perspective) tendencies. Behaviourally, we expected to replicate the pattern of effects found by Samson et al. (2010), with higher error rates and slower response times when the two perspectives differed. In addition, we predicted that switching perspectives between trials would lead to increased error rates and response times compared to when the perspective cues on two consecutive trials were the same. Moreover, we expected egocentric and altercentric interference to increase when participants had to switch perspectives across consecutive trials, due to difficulty disengaging from the previously activated perspective. It is also possible that interference would be reduced or eliminated on stick trials when no perspective switch is necessary. Analyses of eye movements were used to uncover the distinct processing strategies that underlie these effects (i.e. implicit mentalising *versus* directional account).

Method

Participants

A total of 26 participants from the University of Kent took part in the study. Two participants were removed due to poor overall accuracy on the task (<50% accuracy); all other participants achieved at least 95% accuracy. Therefore, the final sample included 24 participants (20 females; M_{age} = 19.04, SD_{age} = 1.17).

Materials

Participants took part in an eye-tracked version of the visual perspective-taking task (Samson et al., 2010), which will be described in full below. Visual stimuli included a picture of a room in 3D lateral view, where the left, back and right walls were visible. Red discs were displayed on one or two of the left/right walls. The number and position of discs changed on each trial. In addition, a male or female human avatar was standing in the centre of the room, facing either the left or right wall (an even split of trials facing each direction). On half the trials, the avatar's orientation meant that s/he saw the same number of discs as the participant (consistent condition), and on the other half, the avatar's orientation meant that s/he could not see some of the discs that were visible to the participant (since they were placed on the wall behind the avatar; inconsistent condition). See Figure 1 for examples of these visual stimuli. The avatar's gender always matched the participant's gender. Participants' task was to verify the number of discs that were visible either according to their own perspective (self perspective condition), or according to the avatar's perspective (other perspective condition).

Procedure

Participants were seated at a distance of 60cm (fixed by a chin-rest) in front of a 20.5 inch colour monitor in 1024 x 768 resolution, and gaze locations and movements from the right eye were recorded using an EyeLink 1000 eye-tracker running at 1000Hz (viewing was binocular). The experiment was controlled using Experiment Builder software and the experimental procedure is illustrated in Figure 1. Each trial began with a drift correction procedure, which was followed by a fixation cross in the centre of the screen for 750ms. The word "YOU" or "SHE/ HE" appeared 500ms later, and was presented for 750ms. This informed participants whether to respond to the current trial according to their own or the avatar's perspective. Following a blank screen lasting 500ms, a digit between 0 and 3 was shown in the centre of the screen for 750ms. This indicated the number of discs the participant needed to verify, according to the given perspective. Finally, the target image of the room, avatar and discs (650 x 480 pixels) appeared centrally on-screen. Thus, the target image subtended 24.5 x 18 degrees of visual angle, which extends beyond the central foveal window (5°), and should encourage participants to move their eyes to verify discs on each side of the room. Participants were instructed to select whether the number of discs in the target image matched the preceding digit according to the cued perspective, using keys "z" as yes and "m" as no (key associations were counterbalanced across participants). Participants were asked to respond as quickly and accurately as possible, and had a maximum of 2000ms to answer before the task moved to the next trial (see Figure 1). The screen advanced to the next trial once a keyboard response had been detected.

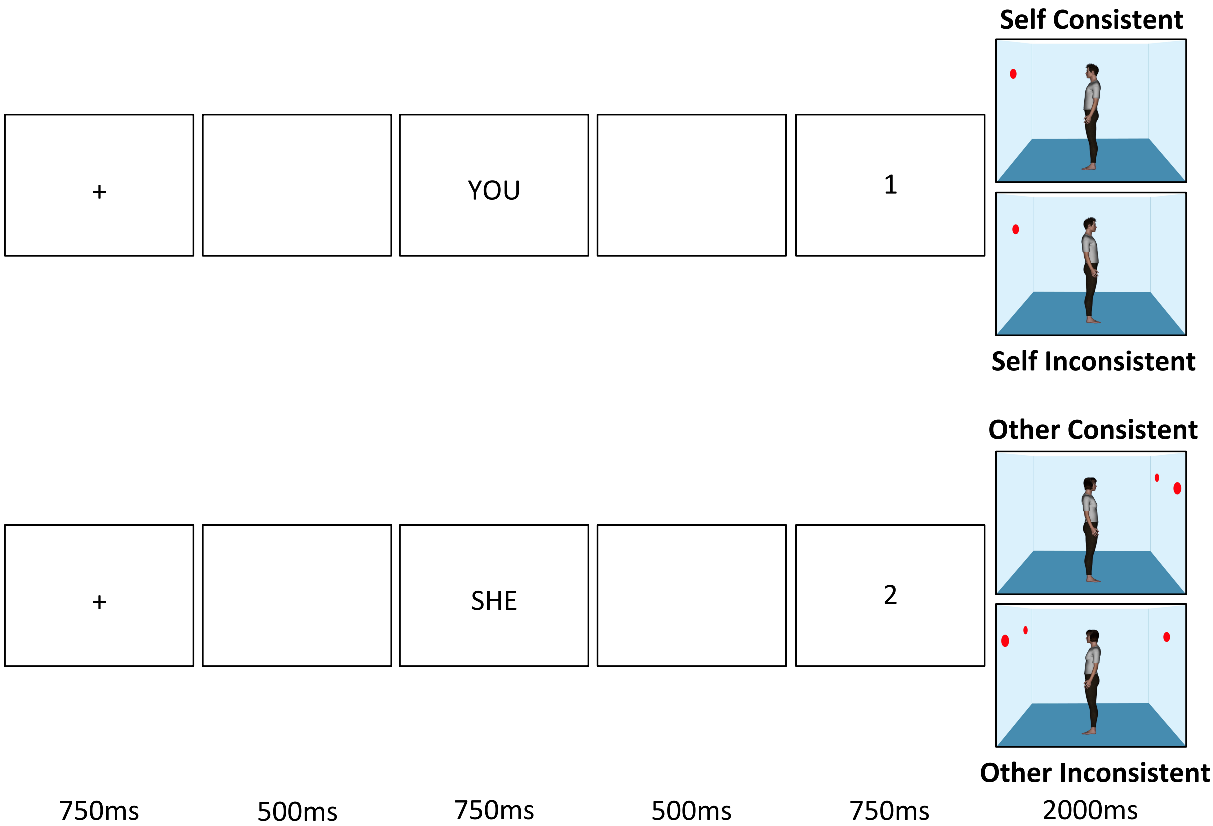

Figure 1: Schematic trial sequence of visual displays presented to participants in the visual perspective-taking task.

Trials could either be matching or mismatching. On matching trials the specified digit correctly corresponded to the number of discs that could be seen from the stated perspective in the target image. On mismatching trials the specified digit did not correctly correspond to the number of discs that could be seen from the stated perspective. Following Samson et al.'s (2010) procedure, and because mismatching trials require different processing, only matching trials were analysed.

Participants completed a practice block of 26 trials, followed by the main task, which consisted of four blocks, each with 52 trials. In total there were 96 matching trials, 96 mismatching trials and 16 ‘filler’ trials (where no discs were displayed on either wall so that the disc number 0 was sometimes correct for self perspective trials). Participants were asked to respond according to their own perspective on 48 matching trials, and to respond according to the avatar’s perspective on the other 48 matching trials. Of these, exactly half were consistent trials, where the avatar and participant’s visual perspectives were the same, and half were inconsistent trials, where the avatar and participant’s visual perspectives were different. Trials were presented in a fixed pseudo-random order so that half the time, self and other perspective trials were preceded by a trial that tapped the same perspective (‘stick’ trial type) and half the time self and other perspective trials were preceded by a different perspective trial, thus requiring participants to switch perspectives between trials (‘switch’ trial type). Half of each trial type (stick/switch) included an avatar that faced the same direction as the preceding trial, and the other half included an avatar that faced a different direction as the preceding trial. Overall, trials were equally likely to have been preceded by a same perspective trial as a different perspective trial. The maximum number of consecutive trials that tapped the same perspective was four, and there were no more than three consecutive trials of the same perspective-consistency condition. No complete stimulus repetitions (i.e. same perspective cue and image) were included.

At the beginning of the experiment, and once every 26 trials thereafter, the eye-tracker was calibrated and validated against nine fixation points, using the standard EyeLink calibration procedure. This procedure took about half a minute and the entire main experiment lasted for about 25 minutes.

In sum, the main experiment observed the effects of three independent variables, in a 2(trial type: stick *vs*. switch) x 2(consistency: consistent *vs*. inconsistent) x 2(perspective: self *vs*. other) repeated-measures design. Effects were analysed on four dependent variables: accuracy of responses, response time, number of fixations, and the location of the first fixation.

Results

Behavioural Data Preparation

Behavioural analyses focused on participants' accuracy and response times when verifying the target image against the given perspective and number of discs. As in Samson et al. (2010) and Qureshi et al. (2010), only matching trials were included in the statistical analyses; mismatching trials, where the correct answer was "no" due to the target image not matching the specified number of discs, were eliminated prior to analysis. Incorrect picture verification responses and trials where the participant did not respond to the image in the given 2000ms were excluded from the response time analysis (5.54%), which was measured from the onset of the picture.

Accuracy and reaction time data were analysed using separate 2x2x2 analyses of variance (ANOVA), with trial type (switch *vs.* stick), perspective (self *vs.* other), and consistency (consistent *vs*. inconsistent) as the within-subjects variables. Bonferonni adjustments were used to correct for multiple comparisons when necessary.

Response Accuracy

Mean accuracy in each condition is shown in Figure 2.

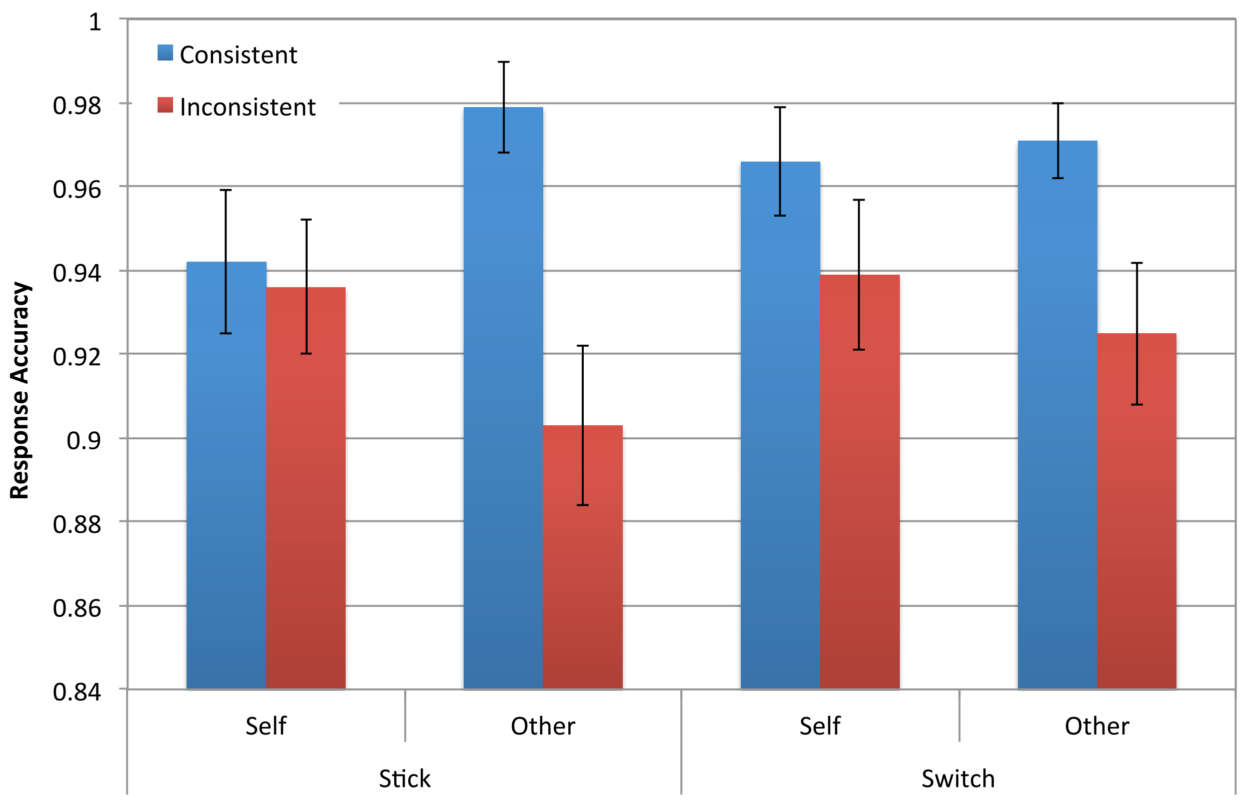

Figure 2: Mean accuracy for each experimental condition. Errors bars show standard errors.

The ANOVA revealed a significant main effect of consistency, $F(1, 23) = 13.32$, $p < .001$, ${}_{p}\eta^2 = .37$, reflecting higher accuracy when participants shared the same visual perspective with the avatar ($M = 97\%$), compared to when the two perspectives were inconsistent ($M = 93\%$). In addition, consistency interacted significantly with perspective, $F(1, 23) = 4.46$, $p < .05$, ${}_{p}\eta^2 = .16$. Follow-up analyses with paired t-tests revealed a significant egocentric intrusion effect, with significantly reduced accuracy on inconsistent versus consistent trials when taking the avatar's perspective, $t(23) = 4.44$, $p < .001$, but no altercentric intrusion effect when taking the self perspective, $t < 1$. Neither the main effect of perspective or trial type, or the remaining interactions were significant, *F*s < 1.

Response Times

Figure 3 presents the mean response times for each condition.

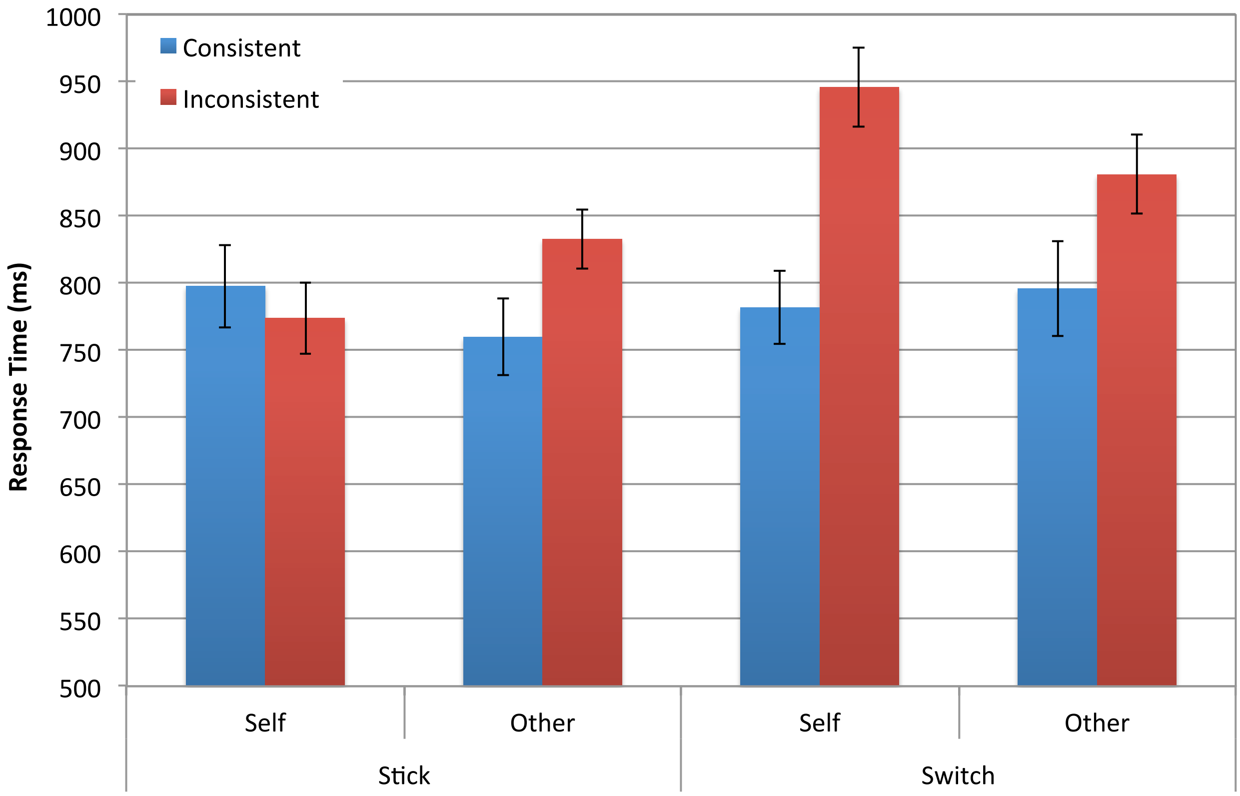

Figure 3: Mean response time for each experimental condition. Errors bars show standard errors.

The ANOVA showed a significant main effect of consistency, $F(1, 23) = 21.89$, $p < .001$, ${}_{p}\eta^2 = .49$, with responses being slower when perspectives were inconsistent ($M = 859$ms) compared to when perspectives were consistent ($M = 784$ms). There was also a main effect of trial type, $F(1, 23) = 51.51$, $p < .001$, ${}_{p}\eta^2 = .69$, which reflected faster overall responses when the previous trial probed the same perspective as the current trial (i.e. stick, $M = 791$ms), compared to when participants had to switch perspective between the previous and current trial ($M = 851$ms). Trial type also interacted with consistency, $F(1, 23) = 21.96$, $p < .001$, ${}_{p}\eta^2 = .49$, and as part of a three-way interaction between perspective*consistency*trial type, $F(1, 23) = 27.39$, $p < .001$, ${}_{p}\eta^2 = .54$. None of the remaining main effects or interactions were significant, Fs < 2.3.

The three-way interaction was examined by running separate ANOVAs for each trial type. When the previous trial probed the same perspective as the current trial (i.e. stick trials), the perspective*consistency interaction was significant, $F(1, 23) = 17.38$, $p < .001$, ${}_{p}\eta^2 = .43$. Paired-sample t-tests revealed egocentric interference when participants judged the avatar's perspective, $t(23) = 3.45$, $p < .005$, with slower responses on inconsistent (M = 833ms) than consistent trials (M = 760ms), but no altercentric interference when participants judged their own perspective, $t < 1.3$. In contrast, when participants had to switch perspective from the previous trial the significant perspective*consistency interaction, $F(1, 23) = 7.17$, $p < .01$, ${}_{p}\eta^2 = .24$, reflected both egocentric and altercentric interference (inconsistent > consistent in both perspective conditions; other, $t(23) = 3.44$, $p < .005$; self, $t(23) = 5.91$, $p < .001$), though here the altercentric intrusion effect (946ms *vs*. 782ms) was larger than the egocentric intrusion effect (881ms *vs*. 796ms).

Eye Movement Analysis

First, we visualized fixations around the scenes by fitting a Gaussian distribution to all fixations for each trial to determine the onscreen distribution of these fixations for each of the experimental conditions (for similar analyses see, e.g., Bindemann, 2010; Bindemann, Scheepers, Ferguson, & Burton, 2010). To correct for avatar gaze direction (i.e. avatar could face left or right) all images were flipped so that the avatar faced left, and fixation locations were adjusted accordingly on the horizontal axis. The resulting distributions were converted to z-scores and are displayed as 'heatmaps' in Figure 4.

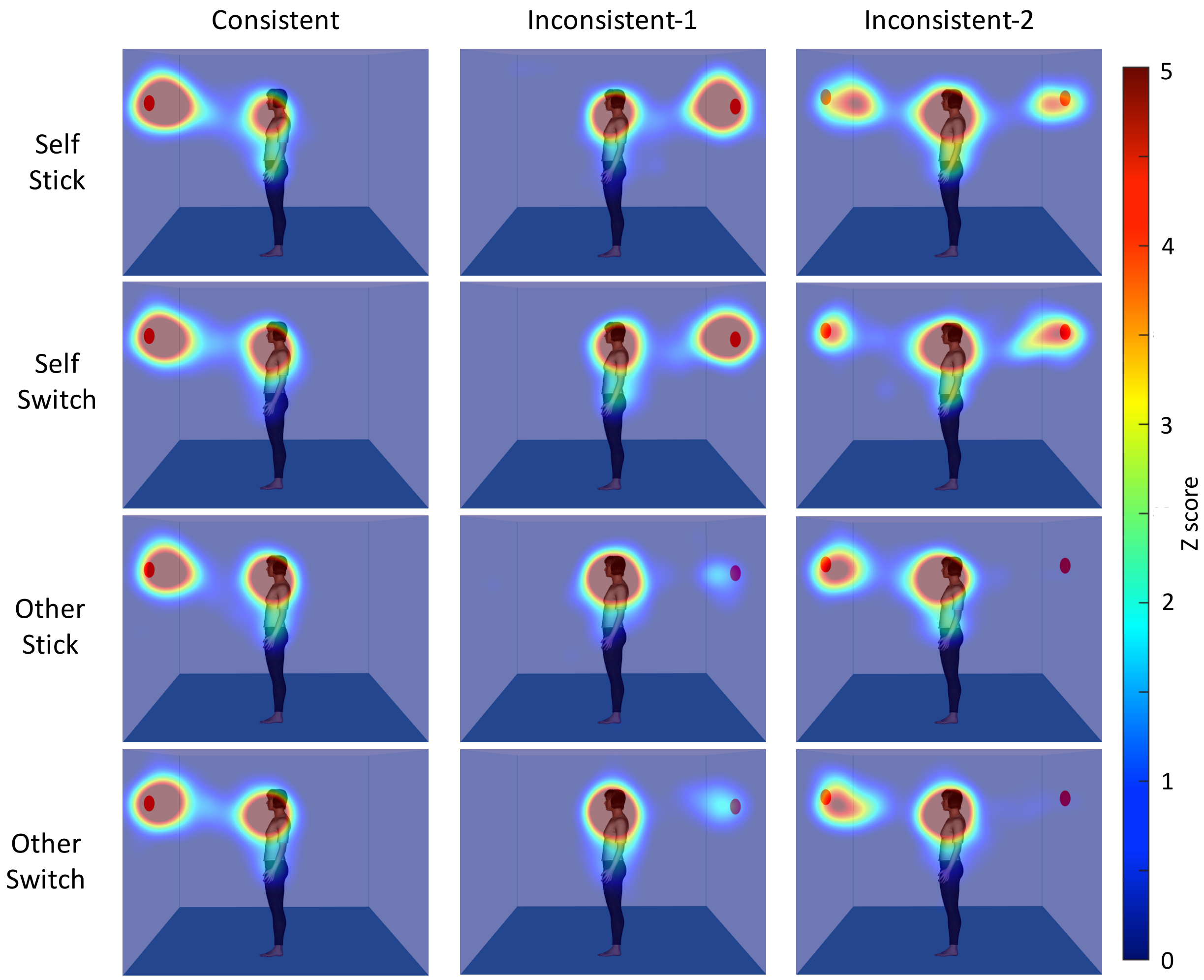

Figure 4: Distribution of all fixations around the scene for each experimental condition. Inconsistent-1 refers to trials where discs were only visible on the wall behind the avatar and Inconsistent-2 refers to trials where discs were visible on both walls. Error bars show standard errors.

Note that on consistent trials the wall behind the avatar never contained any discs, and on half of the inconsistent trials the wall facing the avatar never contained any discs (see Figure 1). The absence of discs on one wall meant that participants had no reason to fixate that location, and thus gaze was biased to the areas containing visual input on these trials, as seen in the fixation distributions in Figure 4. In contrast, on half the inconsistent trials discs were visible on both walls, meaning that the gaze/no-gaze locations contained an equal number of discs across trials and therefore were not affected by low-level visual biases. Due to these inherent visual biases across stimuli and hence fixation distributions, eye movement were analysed in two ways. First, analyses compared consistent trials where discs were only visible on the wall in front of the avatar (i.e. in front 1, behind 0; in front 2, behind 0; in front 3, behind 0) with inconsistent trials where discs were only visible on the wall behind the avatar (i.e. in front 0, behind 1; in front 0, behind 2; in front 0, behind 3). Then inconsistent trials where discs were visible on both walls (i.e. in front 1, behind 1; in front 1, behind 2; in front 2, behind 1) were examined in a separate analysis.

ANOVAs that compared consistent and inconsistent trials (where discs were only visible on the 'no gaze' wall) crossed trial type (switch *vs.* stick), perspective (self *vs.* other), and consistency (consistent *vs*. inconsistent) as the within-subjects variables. ANOVAs on inconsistent trials where discs were visible on both walls crossed trial type (switch *vs.* stick), perspective (self *vs.* other) as the within-subjects variables. Bonferonni adjustments were used to correct for multiple comparisons when necessary.

Number of Fixations First we examined the number of fixations that participants made around the scene during the decision period from target image onset until a keyboard response had been made. The mean number of fixations per trial for each condition is shown in Figure 5.

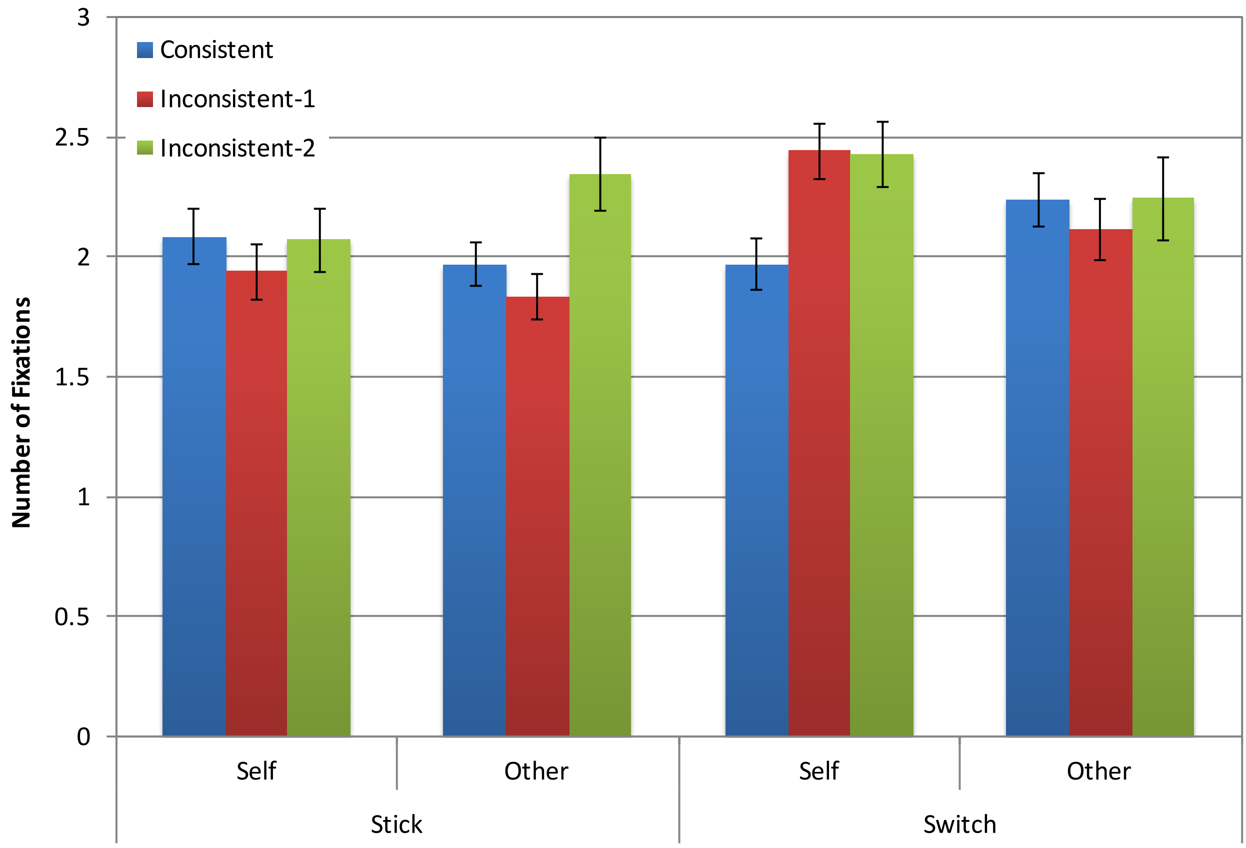

Figure 5: Mean number of fixations in each condition. Inconsistent-1 refers to trials where discs were only visible on the wall behind the avatar and Inconsistent-2 refers to trials where discs were visible on both walls. Error bars show standard errors.

The ANOVA comparing consistent and inconsistent trials (where discs were only visible on the 'no gaze' wall) showed a significant main effect of trial type, $F(1, 23) = 17.45$, $p < .001$, ${}_{p}\eta^2 = .43$, with more fixations being made around the scene when participants had to switch perspective between the previous and current trial ($M = 2.19$), compared to when the previous trial probed the same perspective as the current trial (i.e. stick, $M = 1.95$). In addition, consistency interacted with trial type, $F(1, 23) = 10.55$, $p < .005$, ${}_{p}\eta^2 = .31$, perspective, $F(1, 23) = 5.86$, $p < .05$, ${}_{p}\eta^2 = .2$, and as part of a three-way interaction between trial type*perspective*consistency, $F(1, 23) = 7.31$, $p < .01$, ${}_{p}\eta^2 = .24$. The three-way interaction was examined by running separate ANOVAs for each trial type. These revealed no significant effects on stick trials (all Fs < 3), but a main effect of consistency (consistent < inconsistent), $F(1, 23) = 4.37$, $p < .05$, $_p\eta^2 = .16$, and a significant perspective*consistency interaction, $F(1, 23) = 9.91$, $p < .005$, $_p\eta^2 = .3$, on switch trials. Paired-sample t-tests showed that participants only made more fixations on inconsistent versus consistent trials when responding according to their own perspective, $t(23) = 4.05$, $p < .001$, but not when taking the avatar's perspective, $t < .9$.

The ANOVA on inconsistent trials where discs were visible on both walls showed a significant trial type*perspective interaction, $F(1, 23) = 5.23$, $p < .05$, $_p\eta^2 = .19$. Paired-sample t-tests showed that participants made more fixations on switch versus stick trials when responding according to their own perspective, $t(23) = 2.52$, $p < .05$, but no difference when taking the avatar's perspective, $t < .6$. They also made more fixations when taking the avatar's perspective than the self perspective on stick trials, $t(23) = 2.33$, $p < .05$, but no difference when a perspective switch was required between trials, $t < 1.1$.

First Fixation Location We examined the location of the first fixation following image onset since this provides the most accurate measure of automatic processes[1]. Eye movements were processed by mapping the spatial coordinates, in pixels, of fixations onto appropriate regions of analysis. These regions of analysis corresponded to the two sides of the room (excluding the avatar in the centre), that is the area in front of the avatar's gaze (referred to here as the 'gaze location') and the area behind the avatar's gaze ('no-gaze location') for each image. If a fixation was located within either of these two areas, it was coded as belonging to that gaze/no-gaze location, otherwise it was coded as background.

Visual preferences between the gaze and no-gaze locations were examined for each

---

[1] Participants' starting fixation (i.e. at the point of image onset) was discarded since this was always on the avatar in the centre of the screen (due to the centrally-located number on the preceding screen).

condition by calculating a gaze location bias score (i.e. the probability of first fixating the gaze location *minus* the probability of first fixating the no-gaze location). This measure is symmetrical around zero such that higher proportions of fixations on the gaze location result in a positive score, whereas higher proportions of fixations on the no-gaze location result in a negative score. The maximum score in either direction is therefore 1 or -1. Statistical analyses were carried out on this gaze location bias score. Figure 5 displays the mean bias scores for each experimental condition.

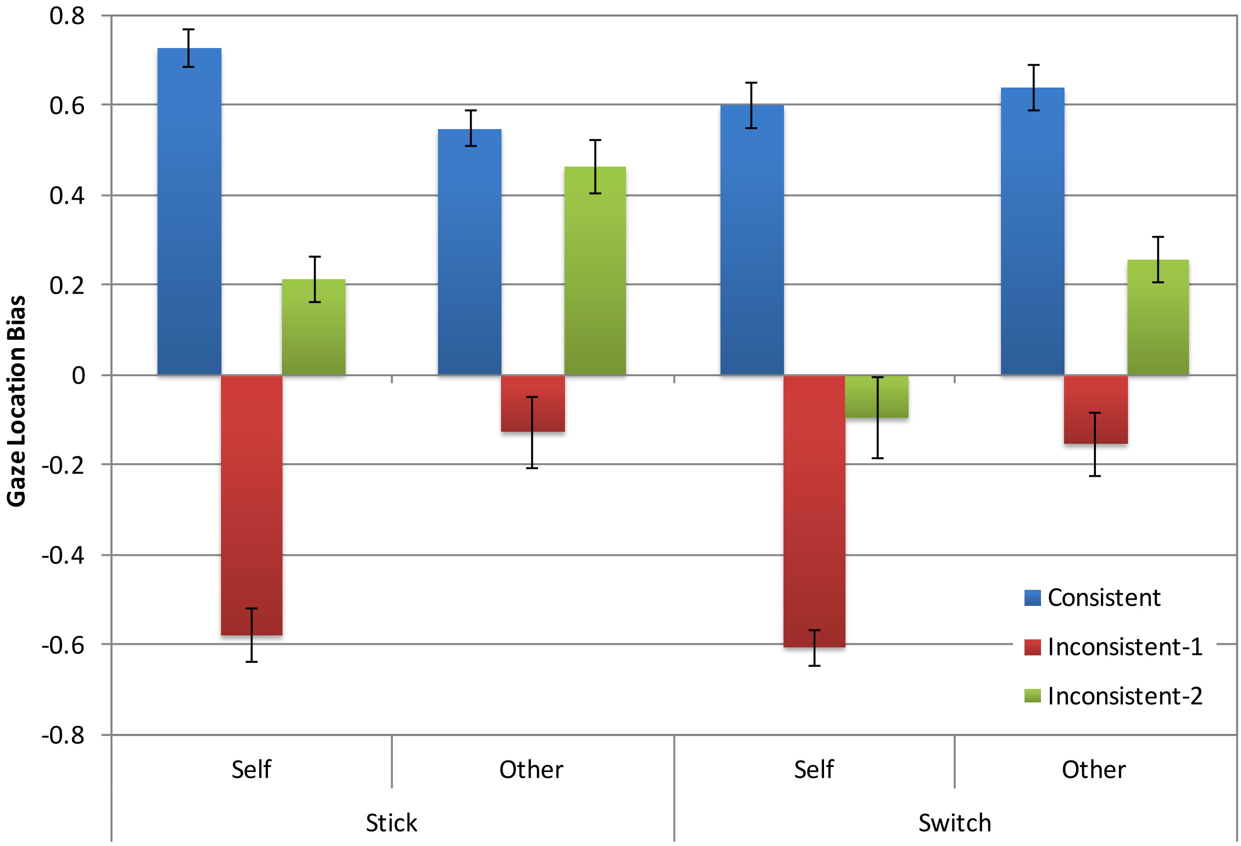

Figure 6: Mean probability of fixating either the gaze location (reflected in a positive value) or no-gaze location (reflected in a negative value) in each condition. Inconsistent-1 refers to trials where discs were only visible on the wall behind the avatar and Inconsistent-2 refers to trials where discs were visible on both walls. Error bars show standard errors.

The ANOVA comparing consistent and inconsistent trials (where discs were only visible on the 'no gaze' wall) showed a significant main effect of consistency, $F(1, 23) = 422.83$, $p < .001$, $_p\eta^2 = .95$, reflecting a bias to fixate the gaze location when the participant and avatar's perspectives were consistent ($M = .63$; 1-sample t-test to zero, $t(23) = 20.48$, $p < .001$), but a weaker bias to fixate the no-gaze location when the participant and avatar's perspectives were inconsistent ($M = -.37$; 1-sample t-test to zero, $t(23) = -10.95$, $p < .001$). The main effect of perspective was also significant, $F(1, 23) = 28.78$, $p < .001$, $_p\eta^2 = .56$, showing a bias to fixate the gaze location when participants were taking the avatar's perspective ($M = .23$; 1-sample t-test to zero, $t(23) = 7.13$, $p < .001$), but no significant bias to either the gaze or no-gaze location when responding according to their own perspective ($M = .04$; 1-sample t-test to zero, $t<1.6$). In addition, the interaction between consistency and perspective was significant, $F(1, 23) = 36.83$, $p < .001$, $_p\eta^2 = .62$. Paired-sample t-tests showed that while the consistency effect was significant for both 'you' and 'they' trials, it was larger when participants were responding according to their own perspective, $t(23) = 18.19$, $p < .001$, than when taking the avatar's perspective, $t(23) = 12.19$, $p < .001$. Moreover, the perspective effect was significant for both consistent and inconsistent trials, but reflected a reduced gaze location bias for other than self perspective trials when the participant and avatar's perspectives were inconsistent, $t(23) = 6.07$, $p < .001$, but the opposite pattern (other < self) when the participant and avatar's perspectives were consistent, $t(23) = 2.69$, $p < .05$. None of the remaining effects reached significance, Fs < 2.4.

The ANOVA on inconsistent trials where discs were visible on both walls showed a main effect of perspective, $F(1, 23) = 24.49$, $p < .001$, $_p\eta^2 = .52$, with a strong gaze location bias when participants were responding according to the avatar's perspective ($M = .36$; 1-sample t-test to zero, $t(23) = 8.22$, $p < .001$), but no bias to either the gaze or no-gaze location when taking their own perspective ($M = .06$; 1-sample t-test to zero, $t<1.3$). In addition, the main effect of trial type was significant, $F(1, 23) = 18.93$, $p < .001$, $_p\eta^2 = .45$, reflecting a bias to the gaze location on stick trials ($M = .34$; 1-sample t-test to zero, $t(23) = 8.43$, $p < .001$) but no bias to either location on switch trials (M = .08; 1-sample t-test to zero, t<1.7). The trial type*perspective interaction was not significant, F < 1.

Discussion

In this paper we used eye-tracking to examine the cognitive mechanisms that underlie performance on a level 1 visual perspective-taking task, and sought to investigate how switching between self and other perspectives influences this ability. Participants completed a version of Samson et al's (2010) dot-probe visual perspective-taking task. They were presented with images of a 3D room, and had to verify the number of discs in that visual scene according to either their own or an on-screen avatar's visual perspective. Crucially, in some of the trials the two perspectives were inconsistent (i.e. each saw a different number of discs), while in others they were consistent. To examine the influence of perspective switching, the order of trials was pseudorandomised such that half the trials were preceded by a trial with the same perspective cue (i.e. (*you*-) *YOU* or (*they*-) *THEY*; referred to as 'stick' trials here) and half the trials were preceded by a trial with a different perspective cue (i.e. (*you*-) *THEY* or (*they*-) *YOU*; referred to as 'switch' trials here). Participants' behavioural responses on the task (i.e. response accuracy and reaction times) were complemented by analyses of eye movements that examined the number and location of fixations (the avatar's gaze location *versus* the wall behind the avatar's gaze) in each condition.

Behavioural results largely replicated previous studies of this kind (Samson et al., 2010; Surtees & Apperly, 2012; Qureshi et al., 2010; Nielsen et al., 2015), showing higher accuracy and faster response times when participants shared the same visual perspective as the on-screen avatar, compared to when the two perspectives were different. However in our study the reduced accuracy on inconsistent versus consistent trials was only significant when participants were prompted to take the avatar's perspective, and not when taking the self perspective. Thus in line with previous studies, our participants experienced both egocentric (i.e. interference from their own perspective) and altercentric (i.e. interference from the avatar's perspective) intrusions during perspective-taking, though the effect of the egocentric bias was greater than the altercentric bias.

Interestingly, while perspective switching did not influence accuracy it did modulate the speed of responses, with participants responding slower overall when the previous and current trials tapped different visual perspectives compared to when no perspective switch was required. Analysis of the number of fixations corroborated this pattern by showing more fixations on switch trials compared to stick trials, suggesting that a broader visual search underlies the increased reaction times on switch trials. Moreover, while egocentric intrusions were apparent on both stick and switch trial types, altercentric interference appeared to be eliminated when participants stuck with their own perspective across consecutive trials. This suggests that inferences based on the other perspective might be weaker when consistently focusing on one's own perspective, which leads to reduced interference between self and other states. In contrast, altercentric interference was significantly increased when participants had to switch from the other to self perspective between trials, suggesting that recent activation of the avatar's perspective enhanced the saliency of the other perspective, leading to increased conflict between self and other states. This pattern was also reflected on the number of fixations measure, where an altercentric effect was indicated by more fixations on inconsistent versus consistent trials, only on switch trials, and not on stick trials. That the altercentric intrusion effect was modulated by perspective switching supports previous neuroimaging and patient studies in showing that distinct cognitive mechanisms underlie attributions about the self and others (e.g. Decety & Sommerville, 2003; Jeannerod & Anquetil, 2008; Samson et al., 2005; Saxe et al., 2006; Vogeley et al., 2001). Further, the fact egocentric interference was comparable on both stick and switch trials is consistent with previous research that has shown automatic computation of the self perspective (e.g. Apperly, Back, Samson, & France, 2008; Keysar et al., 2000, 2003) and pervasive effects of own knowledge (e.g. Birch & Bloom, 2007; Ferguson, Apperly, Ahmad, Bindemann, & Cane, 2015; Ferguson & Breheny, 2012; Mitchell et al., 1996).

It is interesting to find evidence that other perspective-taking is harder than self perspective-taking in the current task, even when a perspective-switch was not required. A previous attempt to load executive functions while participants computed level-1 visual perspectives found that self and other trials were equally affected by a secondary task (egocentric and altercentric intrusions increased to a similar extent; Qureshi et al., 2010). The authors take this as evidence that the executive load impacted the *selection* of the perspective necessary for that trial, after the self and other perspectives had been calculated in a relatively automatic manner. However, the absence of an altercentric effect in our behavioural data on stick trials means we have no grounds to believe that the other perspective was being calculated on these trials. This contrasts with previous reports that altercentric effects (due to the other perspective calculation) arise even when participants only have to judge self perspective and so are never required to switch perspectives (Samson et al., 2010, Experiment 3). One possibility is that participants in the present study did not calculate the other's perspective on the second successive self trial, which would suggest that other perspective calculation was less automatic than has previously been claimed. Another possibility, consistent with the observations of Samson et al., is that the altercentric effect tends to be relatively small compared with the egocentric effect, and so was not detected on behavioural measures in the stick condition of the present experiment.

Examining the location of participants' eye movements around the visual scenes provided further explanation for these effects. Analyses of the location of the first fixation following image onset compared consistent trials, where discs were only visible on the wall in front of the avatar, with inconsistent trials, where discs were only visible on the wall behind the avatar. As was expected given the salient visual differences between these conditions (i.e. presence/absence of discs on one wall), participants were biased to fixate the gaze location when the participant and avatar's perspectives were consistent, but were biased to fixate the no-gaze location when the perspectives were inconsistent. More interesting, however, was the finding that participants showed an overall bias to first fixate the gaze location when taking the avatar's perspective, but were equally likely to make their first fixation on the gaze and no-gaze location when responding according to their own perspective. Crucially, this effect of perspective reflected a reduced bias to first fixate the disc location (i.e. the gaze location on consistent trials and the no-gaze location on inconsistent trials) when referring to the other perspective than when referring to the self perspective. This pattern suggests that when taking the avatar's perspective people are less influenced by the salient presence/absence of discs. The fact that the no-gaze bias on inconsistent trials was reduced when taking the avatar's perspective merely supports the participants' task demands in this condition- to follow the avatar's gaze and verify the number of discs they see. However, the fact that the gaze location bias was also reduced when taking the avatar's perspective on consistent trials (when both the avatar and participant saw the same thing) is indicative of a tight link between visual attention and confidence in one's own interpretation of events. In addition, we found evidence of both egocentric and altercentric effects that were not modulated by trial type. When taking the avatar's perspective, egocentric interference was visible on inconsistent trials as a significant bias to first fixate the disc location (M = -.14), even though this location was in conflict with the avatar's visual perspective. When taking the self perspective, participants showed a stronger bias to first fixate discs when their location was reinforced by the avatar's gaze direction (i.e. the gaze location on self-consistent trials, M = .66) compared to when the discs' location was in conflict with the avatar's gaze direction (i.e. the no-gaze location on self-inconsistent trials, M = -.59). This finding supports the assertion that the avatar's visual perspective elicited altercentric interference on self trials, and demonstrates that participants did experience altercentric interference on self-stick trials even though this was not large enough to be manifest in behavioural responses.

In order to understand the effects of perspective and switching in the absence of low-level cues that might guide visual attention, we ran further analyses on inconsistent trials where discs were visible on both walls, meaning that inherent visual biases (i.e. the mere presence/absence of discs) to one wall or the other were eliminated. When stimuli were equated, we found that participants were still more likely to first fixate the wall holding the avatar's gaze when cued to take the avatar's perspective compared to when referring to the self perspective (where they divided their attention equally between the gaze location and the location behind the avatar's view). This shows that participants were rapidly able to direct visual attention to appropriate locations in the scene according to the perspective cue, and not simply directional properties of the avatar. Biases to the avatar's gaze location were also eliminated when a perspective switch was required between trials, but not when sticking with the same perspective over consecutive trials. This suggests that switching perspectives between trials prompts participants to maintain a more global viewing strategy, thus facilitating self-perspective judgments. Finally, eye-tracking analysis of these visually balanced stimuli supports a weak altercentric effect on self stick trials, with a bias to first fixate the avatar's gaze location on inconsistent trials, though no bias was evident on self switch inconsistent trials. Taken together, these eye-tracking results show that the gaze location bias is highest in conditions that show egocentrism (i.e. the 'other' trials), and lowest of all in the self-switch condition that shows altercentrism. Considered alongside the clear effects of perspective seen in the 1-sided consistent/inconsistent trials, this pattern therefore provides further evidence against the suggestion that the avatar task (and the altercentric effect in particular) is purely driven by spatial cueing of attention (e.g. Heyes, 2014; Santiesteban et al., 2014). Directional accounts assert that responses on inconsistent trials are delayed because directional, not agentive, features of the avatar (i.e. forehead, eyes, nose, etc) automatically shift attention to the discs on one side of the screen, causing a conflict with the total number of discs on the screen. We specifically examined the location of the first fixation participants made in our task since this should provide an indirect measure of automatic processing (but see Duc, Bays, &

Husain, 2008; Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995). Therefore, the fact that our participants' first fixations showed different biases when cued to adopt the self or other perspective, specifically dividing their attention between the two sides on self trials, supports the mentalising explanation. It also fits with recent evidence showing attenuated altercentrism when the avatar's view is blocked by opaque goggles (Furlanetto et al., 2016, but see Cole, Atkinson, Le, & Smith, 2016), reduced altercentric effects for non-social versus social agents (Nielsen et al., 2015; Schurz et al., 2015), and correlations between social skills and altercentric intrusions (Nielsen et al., 2015).

In summary, we have shown that while altercentric interference from the avatar's perspective can be reduced or eliminated when participants stick with their own perspective across consecutive trials, egocentric effects are more pervasive and are not modulated by perspective switching. This behavioural effect was complemented by eye-tracking measures that revealed distinct fixation patterns for self and other perspective-taking. Participants showed a stronger bias to the avatar's gaze location when adopting the avatar's perspective, but divided their attention between both possible locations (i.e. the gaze and no-gaze locations) when referring to their own perspective. This finding goes against recent accounts that have attributed inconsistency effects in this task to directional, not agentive, cues from the avatar that automatically orient attention to one side of the screen.

References

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

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

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

Apperly, I.A., Samson, D., & Humphreys, G.W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology, 45*, 190-201.

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

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, *50*, 2577-2587.

Bindemann, M., Scheepers, C., Ferguson, H.J., & Burton, A. (2010). Face, body and centre of gravity mediate person detection in natural scenes. *Journal Of Experimental Psychology: Human Perception And Performance*, *36*, 1477-1485.

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

Borji, A., Parks, D., & Itti, L. (2014). Complementary effects of gaze duration and early saliency in guiding fixations during free viewing. *Journal of Vision, 14(13)*, 1-32.

Bradford, E.E.F., Jentzsch, I., & Gomez, J-C. (2015). From self to social cognition: theory of mind mechanisms and their relation to executive functioning. *Cognition, 138*, 21-34.

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

Bull, R., Phillips, L.H., & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of Theory of Mind and executive function. *Cognition*, *107*, 663-672.

Cane, J.E., Ferguson, H.J., & Apperly, I. (Submitted). Examining the influence of working memory and motivation on the time course of eye-movements in a perspective taking task. *Journal of Experimenal Psychology: Language, Memory and Cognition*.

Capozzi F., Cavallo A., Furlanetto T., & Becchio C. (2014). Altercentric intrusions from multiple perspectives: Beyond dyads. *PLoS One, 9*, e114210.

Castelhano, M. S., Wieth, M., & Henderson, J. M. (2007). I see what you see: Eye movements in realworld scenes are affected by perceived direction of gaze. *Attention in cognitive systems: Theories and systems from an interdisciplinary viewpoint* (pp. 251-262). Berlin: Springer.

Cole, G.C., Atkinson, M, Le, A,T.D., Smith, D.T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica, 164*, 165-168.

Decety, J., Sommerville, J.A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in Cognitive Science, 7*, 527-33.

Deubel, H. & Schneider, W.X. (1996) Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Research, 36*, 1827-1837.

Duc, A.H., Bays, P., & Husain, M. (2008). Eye movements as a probe of attention. Progress in *Brain Research, 171*, 403-411.

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

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*, 327-339.

Epley, N., Morewedge, C.K., & Keysar, B. (2004). Perspective Taking in Children and Adults: Equivalent Egocentrism but Differential Correction. *Journal of Experimental Social Psychology, 40*, 760-768.

Ferguson, H.J., Apperly, I.A., Ahmad, J., Bindemann, M., & Cane, J.E. (In press). Task constraints distinguish perspective inferences from perspective use during discourse interpretation. *Cognition*, *139*, 50-70.

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

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

Grodner, D., Dalini, M., Pearlstein-Levy, S., & Ward, A. (2012). Factors that contribute to the use of perspective in referent identification. Presented at the *CUNY Conference on Human Sentence Processing*, New York.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives in Psychological Science, 9*, 131-143.

Hoffman, J.E., & Subramaniam, B. (1995) The role of visual attention in saccadic eye movements. *Perceptual Psychophysiology, 57*, 787-795.

Hughes, C. (1998). Finding your marbles: Does preschoolers_ strategic behavior predict later understanding of mind? *Developmental Psychology, 34*, 1326-1339.

Jeannerod, M., Anquetil, T. (2008). Putting oneself in the perspective of the other: a framework for self-other differentiation. *Social Neuroscience, 3*, 356-367.

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

Keysar, B., Barr, D.J., Balin, J.A., & Brauner, J.S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*, 32-38.

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

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

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

McCleery, J.P., Surtees, A., Graham, K.A., Richards, J., & Apperly, I.A. (2011). The neural and cognitive time-course of theory of mind. *Journal of Neuroscience, 31*, 12849-12854.

McKinnon, M.C. & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*, 179-218.

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

Michelon, P. & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics, 68*, 327-337.

Nielsen, M.K., Slade, L., Levy, J.P., & Holmes, A. (2015). Inclined to see it your way: Do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process? *Quarterly Journal of Experimental Psychology, 68,* 1931-1951.

Qureshi, A.W., Apperly, I.A., & Samson, D. (2010). Executive function is necessary for perspective selection, not level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition, 117*, 230-236.

Ramsey, R., Hansen, P., Apperly, I.A., & Samson, D. (2013). Seeing it my way or your way: Frontoparietal brain areas sustain viewpoint-independent perspective selection processes. *Journal of Cognitive Neuroscience, 25*, 670-684.

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

Samson, D., Apperly, I.A., Kathirgamanathan, U., & Humphreys, G.W. (2005). Seeing it my way: A case of selective deficit in inhibiting self-perspective. *Brain, 128*, 1102-1111.

Santiesteban, I., Catmur, C., Hopkins, S., Bird, G. & Heyes, C.M. (2014). Avatars and arrows: Implicit mentalizing or domain general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40*, 929-937.

Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience, 1*, 229-234.

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

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

Surtees, A.D.R., & Apperly, I.A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development, 83*, 452-460.

Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective taking. *Cognition, 129*, 426-438.

Vogeley, K., Bussfeld, P., Newen, A., Hermann, S., Happe, F., Falkai, P., Maier, W., Shah, N., Fink, G., & Zilles, K. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *NeuroImage, 14*, 170-181.