'Spontaneous' Visual Perspective-Taking Mediated by Attention Orienting that is Voluntary and not Reflexive

Gardner, M., Taylor, D.A., Hull, Z. and Edmonds, C.J.

This is an Accepted Manuscript of an article published by Taylor & Francis in Quarterly Journal of Experimental Psychology, doi: 10.1080/17470218.2017.1307868

The final definitive version is available online:

https://dx.doi.org/10.1080/17470218.2017.1307868

Author Accepted Manuscript

Journal: *Quarterly Journal of Experimental Psychology*

Date accepted: 30 January 2017

'Spontaneous' Visual Perspective-Taking Mediated by Attention Orienting that is Voluntary and not Reflexive

Mark R. Gardner [a] , Zainabb Hull [a], Donna Taylor [a], and Caroline J. Edmonds [b]

[a] Department of Psychology, University of Westminster, 115 New Cavendish Street, London, W1W 6UW, United Kingdom

[b] School of Psychology, University of East London, Stratford Campus, Water Lane, London, E15 4LZ, United Kingdom

Running head: perspective-taking & attention orienting

Corresponding author:

Mark R. Gardner

Department of Psychology, University of Westminster, 115 New Cavendish Street, London, W1W 6UW

m.gardner@westminster.ac.uk

Tel: 020 3506 9020

Acknowledgements

This work was supported by the Experimental Psychology Society under a small research grants scheme awarded to the first author. We are grateful to Dana Samson for generously making available the materials used in the avatar task, and to two anonymous referees for constructive comments on a previous version of this paper.

Abstract

Experiments revealing ‘spontaneous’ visual perspective-taking are conventionally interpreted as demonstrating that adults have the capacity to track simple mental states in a fast and efficient manner (‘implicit mentalising’). A rival account suggests that these experiments can be explained by the general purpose mechanisms responsible for reflexive attentional orienting. Here, we report two experiments designed to distinguish between these competing accounts. In Experiment 1, we assessed whether reflexive attention orienting was sufficient to yield findings interpreted as spontaneous perspective taking in the ‘avatar task’ (Samson et al., 2010) when the protocol was adapted so that participants were unaware that they were taking part in a perspective-taking experiment. Results revealed no evidence for perspective-taking. In Experiment 2, we employed a Posner paradigm to investigate the attentional orienting properties of the avatar stimuli. This revealed cue-validity effects only for longer stimulus onset asynchronies, which indicates a voluntary rather than reflexive shift in spatial attention. Taken together, these findings suggest that attentional orienting does indeed contribute to performance in the Samson et al. avatar task. However, attention orienting appears to be voluntary rather than reflexive, indicating that the perspective-taking phenomenon measured may be less spontaneous than first reported.

keywords

social attention, attention orienting, perspective-taking, theory of mind, implicit mentalising

'Spontaneous' Visual Perspective-Taking Mediated by Attention Orienting that is Voluntary and not Reflexive

‘Implicit mentalising’ refers to the ability to track others’ mental states in a relatively automatic, fast and efficient manner, in contrast to ‘explicit mentalising’ which develops during childhood (e.g., Wellman, Cross, & Watson, 2001), and remains effortful even for adults (e.g., German & Hehman, 2006). The conventional view is that implicit mentalising is a specialised neurocognitive mechanism (e.g., Apperly, 2010). An alternative ‘submentalising’ hypothesis contends that implicit mentalising is mediated by domain general cognitive processes for memory and attention (Heyes, 2014). Resolving this issue is crucial to elucidate whether there are one or two dedicated systems for mentalising (Apperly & Butterfill, 2009; Heyes & Frith, 2014) and to understand the pattern of difficulties encountered in Autism Spectrum disorders (Schwarzkopf, Schilbach, Vogely, & Timmermans, 2014; Senju, Southgate, White, & Frith, 2009). This paper aims to address this issue by discriminating between the conventional and alternative accounts for one source of evidence of implicit mentalising: spontaneous visual perspective-taking revealed by the Samson avatar task (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010).

Evidence for spontaneous visual perspective-taking provided by the Samson avatar task has conventionally been interpreted as providing strong evidence for implicit mentalising. This task requires participants to judge the number of items seen from their own perspective while a visually represented human figure ‘sees’ either a matching or non-matching number of items (Samson et al., 2010). A robust correspondence effect, shown by poorer performance for non-matching compared to matching trials, is conventionally taken to be due to a specialised process that automatically computes what the figure sees. This process appears to be automatic because the correspondence effect occurs even though the avatar is formally task irrelevant and leads to a detriment in performance. Furthermore, this effect is not suppressed by dual-task demands (Qureshi, Apperly, & Samson, 2010). The magnitude of the correspondence effect has been found to be positively correlated to individual differences in perspective-taking and empathy measured by self-report (Nielsen, Slade, Levy, & Holmes, 2015). This effect has also been found to be absent under conditions in which the avatar was believed to be unable to see the items (dots attatched to a wall), arising from a "belief induction procedure" whereby the participants were given direct experience of opaque and transparent goggles worn by the avatar (Furlanetto, Becchio, Samson, & Apperly, 2016). These findings lend support to the view that mental states are automatically tracked in the Samson avatar task, although a recent study reports a failure to replicate the latter two findings (Conway, Lee, Ojaghi, Catmur & Bird, in press).

The rival submentalising account claims that the Samson avatar task does not measure implicit mentalising. Instead, the specific hypothesis has been proposed that the avatar simply serves to orientate attention automatically to a region of space containing either a matching or non-matching number of items (Heyes, 2014). In other words, the two accounts do not disagree on the automaticity claim but disagree about whether it is the directional or "agentive" properties of the avatar that drive the effect. *Prima facie* support for the submentalising account comes from comparable correspondence effects that occur when the figure is replaced by cues known to direct attention, such as arrows (Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014; Nielsen et al., 2015). Additional support is provided by the finding that the effect remains under experimental conditions in which the avatar's visual access to the dots is disrupted by a physical barrier (Cole, Atkinson, Le, & Smith, 2016), an optical 'cloaking device' (Conway et al., in press), and opaque goggles (Conway et al., in press).

The submentalising account reasonably assumes that the attentional orienting effect of the avatar is reflexive, in accordance with the apparent automaticity of the spontaneous visual perspective-taking phenomenon. In line with this account, perspective-taking in the Samson task appears to be (a) independent of task demands, occurring despite the avatar being formally task irrelevant (i.e., an uninformative cue), and (b) fast acting, with the avatar and dot presented simultaneously (i.e., SOA = 0ms). However, a challenge for the submentalising account is that attention orienting effects from gaze cues are not typically shown with an SOA of 0ms, as noted by Bukowski, Hietanen, & Samson (2016).

Bukowski et al. (2016) have provided a partial reconciliation of discrepancies between the spontaneous visual perspective-taking and gaze cueing literatures. In a series of experiments, they modified the classic gaze cueing paradigm to examine the attention orienting properties of stimuli identical to those used in the original avatar task (Samson et al., 2010). Although these modifications required various departures from the avatar task (e.g., instructions superimposed on the avatar), clear evidence was found that these avatars do indeed show attention orienting effects, lending support to the submentalising account. However, these effects were contingent upon either the target occurring after a delay (SOA = 300ms), or when instructions endow the avatar with high salience. These findings are difficult to reconcile with the original versions of either account. On the one hand, these conditions of occurrence would appear to be inconsistent with a reflexive mode of attention orienting, as specified by the submentalising account (Heyes, 2014; Santiesteban et al., 2014). On the other hand, these conditions of occurrence appear to limit the claim that visual perspective-taking is spontaneous. Specifically, if mental state ascription occurs, it appears not to be automatically and mandatorily triggered by the mere presence of an avatar (Bukowski et al., 2016). Instead, the authors accommodate these findings within an implicit mentalising account by proposing that visual perspective-taking is context-dependent and affected by the extent to which the avatar is attended to.

In the present study, two experiments were conducted with the aim of discriminating between the aforementioned submentalising and implicit mentalising accounts of performance in the Samson avatar task. Our approach was to assess the influence of reflexive attention orienting in two complementary ways. Experiment 1 adapted the avatar task to investigate whether reflexive attention orienting was sufficient to yield an avatar-consistent responding bias, even when the context was modified so that participants were unaware that they were taking part in a perspective-taking experiment. Experiment 2 employed a Posner paradigm to investigate whether the stimuli employed in the avatar task elicit reflexive attention orienting effects. In both experiments performance for avatars was compared to that for arrows, as a symbolic directional cue to which mental states would not be ascribed.

Experiment 1

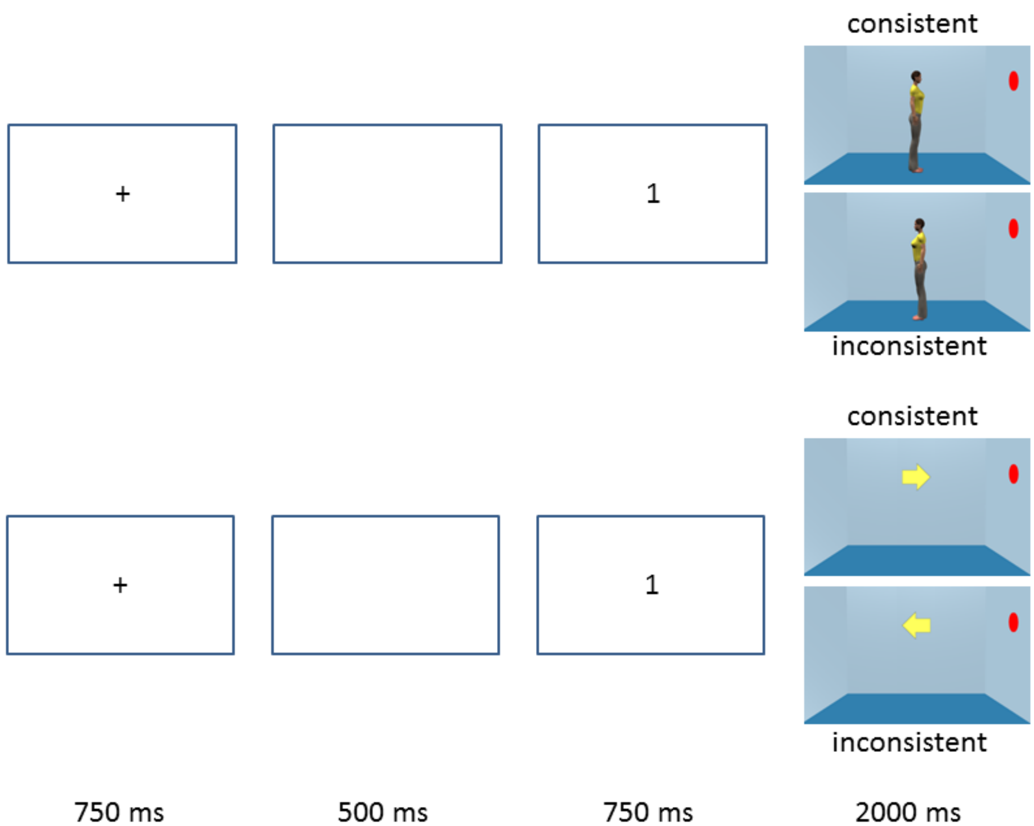
Experiment 1 assessed whether reflexive attention orienting was sufficient to yield a correspondence effect in the Samson avatar paradigm by modifying the context in which the experiment was conducted so that participants were unaware that they were taking part in a perspective-taking experiment. In this respect, the present experiment complements previous work examining the attention orienting properties of avatar stimuli in the absence of perspective-taking instructions within a gaze cueing paradigm (Bukowski et al., 2016). Our experiment differs by examining the effect of removing this expectation directly within the Samson avatar paradigm, with our procedure otherwise modelled on that employed by Samson et al. (2010, Experiment 3).

The perspective-taking demands of the avatar task were reduced by three modifications to the context in which it was presented to participants. First, in contrast to the standard protocol, none of the recruitment materials, information sheets, or instructions to participants made any mention of perspective-taking. Instead, the task was presented to participants as examining the interpretation of a previously published cognitive test. Second, participants did not undertake any "other perspective" trials at any time within the current experiment. It has been acknowledged that other perspective trials explicitly invite participants to adopt the perspective of the avatar (Samson et al., 2010) and may induce a transfer of a volitional perspective-taking strategy to "self perspective" trials when presented within the same experiment (Samson et al., 2010; see also Santiesteban et al., 2014). In line with this view, the size of the correspondence effect has been found to be modulated by whether the previous trial was from the self- or other-perspective (Ferguson, Apperly, & Cane, 2016). Third, the trial structure was modified so that the now redundant cue to respond from the self-perspective (the word "YOU") was never presented. This change was also made in light of evidence that inducing a self-perspective (e.g., Stephensons & Wicklund, 1983), including through the use of personal pronouns (Brunye, Ditman, Mahoney, Augustyn, & Taylor, 2009), can lead to a focus on a third-party perspective.

With the perspective-taking demands of the avatar task minimised in these ways, the implicit mentalising account predicts no correspondence effect. This is because this account proposes that a correspondence effect is contingent upon increased salience to the cue induced by a perspective-taking context when, as is normally the case, the avatar is presented simultaneously with the dots (Bukowski et al., 2016). By contrast, the specific submentalising account of Heyes and colleagues (Heyes, 2014; Santiesteban et al., 2014) predicts that a correspondence effect should remain despite these modifications. This is because stimulus-driven or reflexive attention orienting should be impermeable to manipulation of the perspective-taking context.

Methods

Participants. In total, 34 first year undergraduate psychology students from the University of Westminster participated for course credit. Two participants that made high numbers of errors were excluded (see Results), resulting in a sample size of 32 (27 female), aged 18 to 36 (M = 20.0, SD = 3.38). This sample size exceeds the 24 participants required to detect the effect size obtained by Samson et al. (2010, Experiment 3), d = 0.61, with 80% power.



* Figure 1 about here *

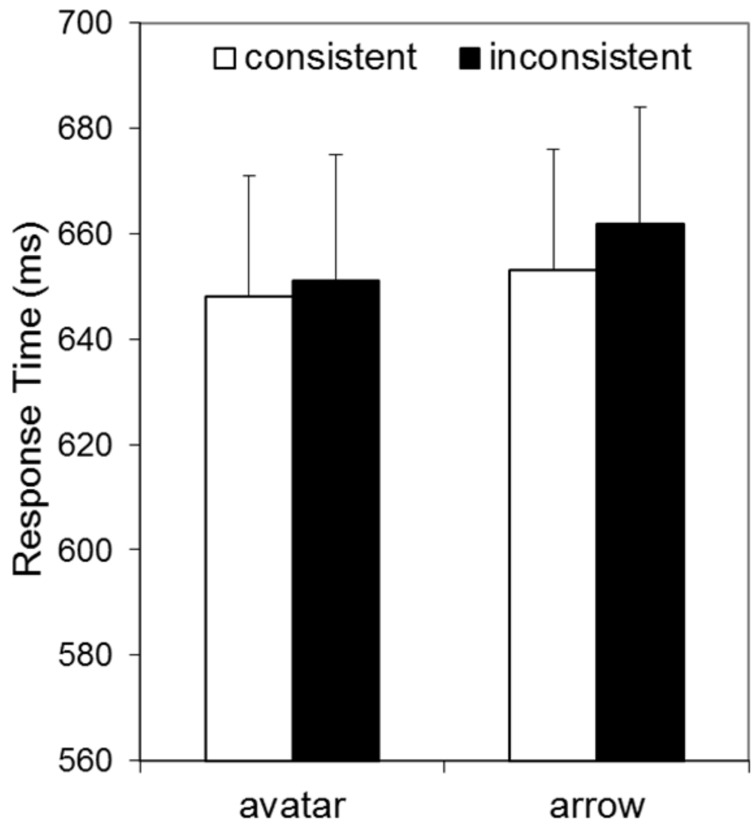
Materials. Stimuli depicted a simple virtual room within which a centrally located stimulus (an avatar or an arrow) was presented, always oriented either to the left or to the right, as illustrated in Figure 1. A variable number of 'dots' were also presented (0, 1, 2 or 3), that appeared as bright red ellipses situated on the left and/or right walls. Stimuli containing an avatar were the same images used by Samson et al. (2010) while those containing an arrow were digitally edited from these originals. Therefore, the distribution of dots in each trial were also identical to previous work (Samson et al., 2010; Santiesteban et al., 2014). "Yes" responses were made by pressing the "L" key on a QWERTY keyboard, and "No" responses were made by pressing the "A" key. Stimulus presentation and response collection were controlled by E-Prime (Schneider, Eschman, & Zuccolotto, 2002), running on a Dell PC with 22" screen.

Procedure. Trials commenced with a fixation cross presented for 750ms. After a 500ms interval, a digit (0-3) was presented for 750ms. Finally, the image of the room containing the central stimulus and between 0-3 dots was presented until a response was made, up to a maximum of 2000ms. The participant's task was to respond as quickly and accurately as possible indicating "Yes" if the "number of dots present" corresponded to the digit, and "No" if these quantities did not correspond. Thus the trial structure closely replicated that of Samson et al. (2010, Experiment 3), with the exception of the perspective prompt "YOU", and, consequentially, a reduction in inter-trial intervals of 1250ms.

Trials were organised into two consecutive blocks for the arrow stimulus and two consecutive blocks for the avatar stimulus, with order counterbalanced between participants. Each block comprised 52 trials, including 4 filler trials in which no dots were presented, 24 (50%) in which the dots were located such that the avatar/arrow was oriented to the same number as seen by the participant (consistent condition), and 24 (50%) in which the dots were located such that the avatar/arrow was oriented to fewer than those seen by the participant (inconsistent condition). For each condition, there were an equal number of trials in which the central stimulus was oriented to the left/right, and an equal number of trials in which the correct response was “yes” and “no”, using the trial compositions employed by Samson et al. (2010). Furthermore, trial order was pseudorandomised, constrained to avoid more than three consecutive trials of the same type, using the sequences employed by Samson et al. Thus, the procedure closely replicated the original avatar task, with the exception of features that might induce a perspective-taking context (explicit perspective-taking instructions, “other” trials, as well as the absence of a “You” prompt).

Results & Discussion

Following Samson et al. (2010), only data for trials in which the number of dots matched the digit were considered in the analysis (i.e., where “yes” was the correct response), excluding also “filler” trials (where the digit was 0 and no dots were present). Mean response times (RT) for correct responses, and the Percentage of Errors (PE) were computed as DVs. Two participants that made exceptionally high error rates were excluded from the analysis; one female participant in the avatar first condition (PE = 21%), and one female participant in the arrow first condition (PE = 17%). Across the remaining 32 participants, there were relatively low levels of errors (M = 3.0%, SD = 2.68%), and no response omissions due to timeout.



- - - - - - - - - - - - - - -

Figure 2 about here

- - - - - - - - - - - - - - -

Response time data are illustrated in Figure 2. These data show negligible difference between RTs for consistent and inconsistent conditions, both for the avatar and for the arrow stimuli. These data were analysed by a 2 x 2 repeated measures Analysis of Variance (ANOVA), with Correspondence (consistent *vs*. inconsistent) and Stimulus (avatar *vs*. arrow) as the factors. This revealed no effect of Correspondence, $F(1, 31) = 0.67$, $MSE = 1508$, $p = .421$. Furthermore, there was no effect of Stimulus, $F(1, 31) = 0.24$, $MSE = 8596$, $p = .631$, and no interaction, $F(1, 31) = 0.34$, $MSE = 1076$, $p = .567$. Related t-tests confirmed the absence of a correspondence effect for both stimuli; the elevation in response times for inconsistent compared to consistent trials was non significant, both for the avatar, 2.2ms, $t(31) = 0.282$, $p = .780$, and for the arrow, 9.0ms, $t(31) = 0.902$, $p = .374$.

Similar results were found for the equivalent analysis of PE. There was neither a main effect of Correspondence $F(1, 31) = 2.63$, $MSE = 10.0$, $p = .115$, nor of Stimulus, $F(1, 31) = 1.18$, $MSE = 10.35$, $p = .179$, and no interaction, $F(1, 31) = 0.20$, $MSE = 15.3$, $p = .654$. Related t-tests confirmed the absence of a correspondence effect both for avatars (inconsistent: M = 1.9, SD = 3.8; consistent: M = 3.1, SD = 3.5; $t(31) = 1.34$, $p = .189$), and arrows (inconsistent: M = 3.6, SD = 5.2; consistent: M = 3.0, SD = 3.2; $t(31) = 0.68$, $p = .500$).

Experiment 1 therefore revealed no evidence of a correspondence effect in the Samson avatar task when administered in the absence of a perspective-taking context. This was despite the methodology otherwise being a close replication of prior work (Samson et al., 2010), and the sample size providing adequate power for replicating previously reported effect sizes. To our knowledge, this is the first reported absence of a correspondence effect in this task. This absence of an effect helps to identify the boundary conditions of this effect by calling into question the degree to which it is automatic and mandatory. It therefore poses a challenge not only to original claims for *spontaneous* visual perspective-taking, but also to the submentalising account that suggests reflexive attention orienting may account for correspondence effects in this task (Heyes, 2014; Santiesteban et al., 2014).

By contrast, an absence of a correspondence effect is consistent with the implicit mentalising account. This account, as revised by Bukowski et al. (2016), predicts that visual perspective-taking is contingent upon the increased salience to the avatar cue that is offered by a perspective-taking context. However, given evidence for top down modulation of cueing effects (Vaquero, Fiacconi, & Milliken, 2010), our absence of a correspondence effect may be recast as suggesting that such effects are contingent upon participants' appraisal of task demands – a top-down evaluation that the avatar's perspective (or arrow's direction) is important to the task at hand. Thus, although questioning the spontaneous / reflexive nature of the phenomenon, the present results do not discriminate between mentalising and submentalising explanations.

Experiment 2

The absence of a correspondence effect in Experiment 1 raises doubts about whether the Samson avatar is measuring a spontaneous / reflexive mental process. Experiment 2 aims to examine directly whether the avatar stimuli used in the Samson task have attention orienting properties. The literature on social attention for cues such as body orientation and head position is relatively underdeveloped in comparison to that for gaze cueing and thus offers little to guide our predictions (Nummenmaa & Calder, 2008). For instance, work using a Posner cueing task has shown that photographs of left/right-facing bodies orient spatial attention in a direction-consistent manner, but only when the photographs implied that the actor was moving (Gervais, Reed, Beall, & Roberts, 2010). More recently, Bukowksi et al. (2016) have reported attention orienting effects for static avatars at an SOA of 300ms. Cue-validity effects for the same stimuli only emerged at 0ms SOA when the task had been adapted to make it more complex (with an increase in response times) and instructions altered to make the central stimulus more salient. Thus, it is unclear from results to date whether static avatars direct attention through a reflexive process or via relatively slow volitional attention orienting.

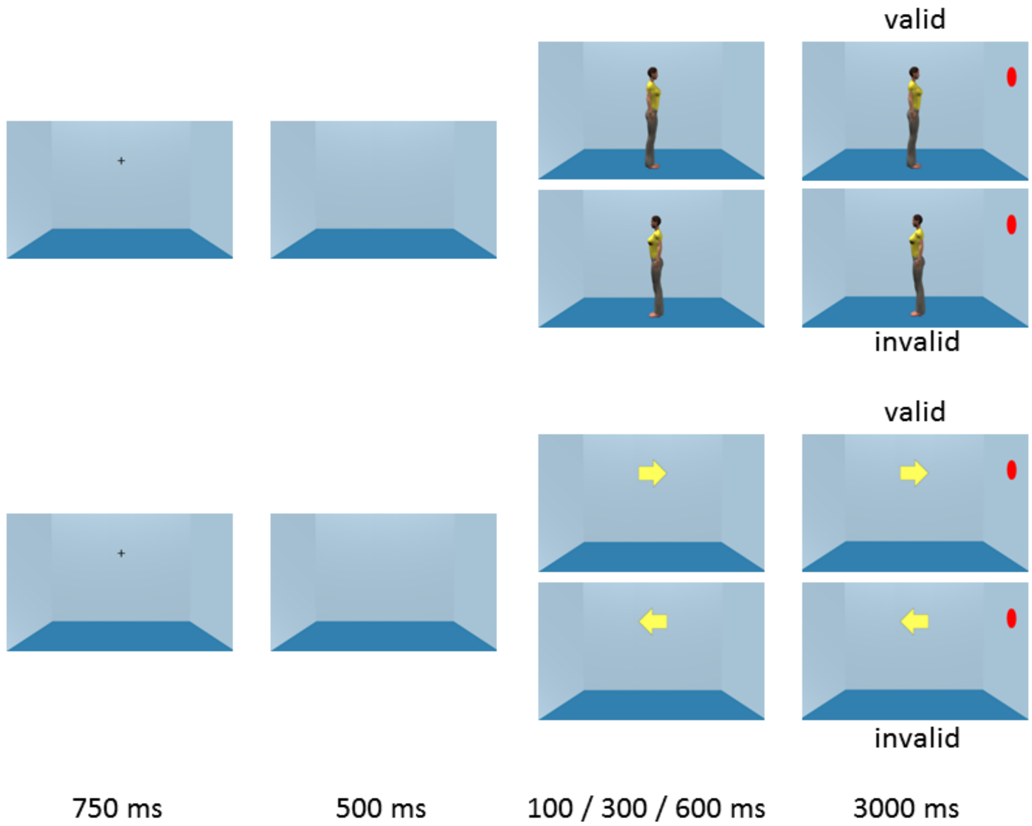
The objective of Experiment 2 was to address this issue using a Posner cueing paradigm and the same avatar stimuli as in Experiment 1 (as originally used by Samson et al., 2010).

Following Gervais et al. (2010), we selected a detection task to assess cue validity effects for both types of cue at a range SOAs that cover the range at which reflexive and volitional attention orienting effects occur (100, 300, 600 ms; Egeth & Yantis, 1997; Kingstone, Tipper, Ristic, & Ngan, 2003; Muller & Rabbit, 1989; Posner & Cohen, 1984). A simple detection task yielding short RTs provides a fairer test of fast-acting reflexive attention orienting than a task designed to provide longer RTs through superficial task matching. Participants were asked to detect the appearance of a single dot in the same virtual room, where this target was either validly or invalidly cued by the prior appearance of a central cue (avatar / arrow). In common with the Samson avatar task, the cue was non-informative (i.e., validly cued the target on 50% of trials). The reflexive attention orienting account predicts a cue-validity effect primarily at short SOAs. Similarly, the implicit mentalising account also predicts an effect at short SOAs, given that this account predicts that unintentional and efficient visual perspective taking may occur in the absence of a perspective-taking context if the avatar stimulus is presented alone, prior to the target (Bukowski et al., 2016). By contrast, a cue-validity effect at longer SOAs would be indicative of volitional attention orienting.

Methods

Participants. A fresh sample of student volunteers, drawn from the same population, participated for course credit. In total, 33 students participated. However, 1 participant was excluded on the basis of a high error rate (see Results), resulting in a sample, N = 32 (27 female) aged 18 to 36 (M = 20.3, SD = 3.21).

Materials. Stimuli used in Experiment 1 were re-employed in this experiment. Images containing a central stimulus (avatar/arrow) oriented to the left or to the right and a single dot located on the left or right wall served as valid and invalid cued targets. In addition, a new image, whereby a fixation cross was presented within the context of the virtual room, was digitally edited from the originals. The same PC as used in Experiment 1 and testing room was used for this experiment.



- - - - - - - - - - - - - - -

Figure 3 about here

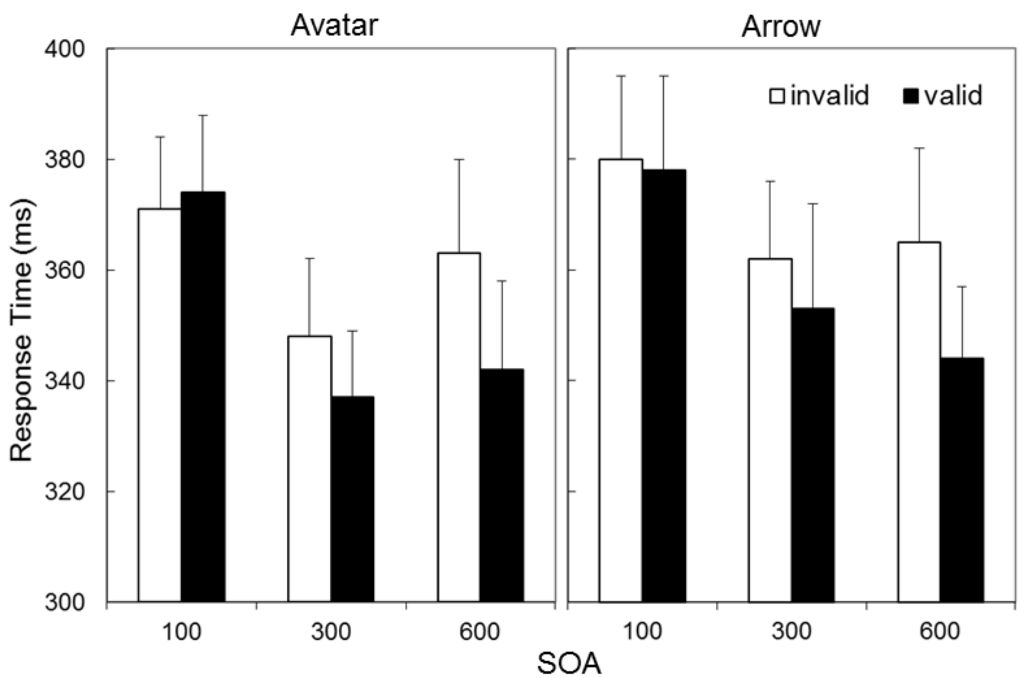
- - - - - - - - - - - - - - -

Procedure. As illustrated in Figure 3, trials commenced with a fixation cross presented within the context of the virtual room, which was displayed for 750 ms, followed 500 ms later by a central cue oriented either to the left or to the right. After a variable delay (SOA = 100, 300, 600 ms), the target (dot) appeared on the left or the right wall. The cue and target remained displayed until a keypress was made up to a maximum of 3000 ms, followed by an inter-trial interval of 500 ms when the room continued to be presented.

Following a short practice block (14 trials), participants completed 4 blocks in the experiment proper in which the type of central stimulus alternated, with order counterbalanced across participants. Each block comprised 66 trials, including 60 target present trials and 6 catch trials in which no target was present. Trials were randomly ordered, with all combinations of SOA, cue direction, and cue validity occurring on 10 trials within each block. Participants were asked to press a single response key ("H") as fast as possible on detecting the target, but to withdraw from responding when the target was not presented. They were also instructed to keep their eyes focused on the centre of the screen throughout, and informed that cues did not predict the location of the target.

Results and Discussion

Rates of false alarms during catch trials were low (M = 4.2%, SD = 4.8), after the data for one participant was excluded whose performance was more than two standard deviations away from the mean. Mean RT was computed for each participant and condition, excluding catch trials and response omissions through timeout (0.13% of the data).



- - - - - - - - - - - - - - -

Figure 4 about here

- - - - - - - - - - - - - - -

Figure 4 presents Mean RT as a function of validity, SOA, and cue-type. These data appear to indicate a cue validity effect only at longer SOAs, with similar performance for an avatar to an arrow cue. These impressions were confirmed by a 2 x 3 x 2 repeated measures Analysis of Variance (ANOVA), with Validity (valid *vs*. invalid), SOA (100 *vs*. 300 *vs*. 600ms), and Cue-type (avatar *vs*. arrow) as the factors. This revealed main effects of Validity, $F(1, 31) = 6.54$, $MSE = 1495$, $p = .016$, and SOA, $F(2, 62) = 19.7$, $MSE = 1230$, $p < .001$, but not of Cue-type, $F(1, 31) = 0.95$, $MSE = 6502$, $p = .338$. Critically, Validity was found to interact with SOA, *F*(2, 62) = 4.52, *MSE* = 744, *p* =.015, but not with Cue-type (*F*s < 0.132, *p*s > .877). Related-t tests revealed a statistically significant cue validity effect to be restricted to 600ms SOA, *M* ± *SD* = 20.3 ± 33.7 ms, *t*(31) = 3.40, *p* = .002 (300: 10.2 ± 31.3, *t*(31) = 1.85, *p* = .075; 100: 0.2 ±29.4, *t*(31) = 0.04, *p* = .968).

These results revealed cue-validity effects indicating that the characters employed in the Samson avatar task do indeed have attention orienting properties that were similar to arrows. However, these effects were apparent only at longer SOAs (600ms), indicating a voluntary rather than reflexive shift in spatial attention (Egeth & Yantis, 1997; Muller & Rabbit, 1989). This finding is inconsistent with accounts predicting cue-validity effects at short SOAs, based upon implicit mentalising or reflexive attention orienting, and implies that top-down processes contribute to correspondence effects in the Samson avatar task.

General Discussion

The present study examined whether spontaneous visual perspective-taking could be explained by reflexive attention orienting, as proposed by the 'submentalising' account (Heyes, 2014; Santiesteban et al., 2014). Two experiments examined predictions of this account. Experiment 1 found no evidence of a correspondence effect in a Samson avatar task modified to be administered in the absence of a perspective-taking context. This suggests that reflexive attention orienting alone is not sufficient to drive this effect. Crucially, Experiment 2 revealed that the avatar stimuli, like arrows, do indeed direct spatial attention in a Posner cueing task but only at longer SOAs. This aspect indicates a volitional rather than reflexive process. Therefore, while domain general attention orienting appears to play a role in the Samson avatar task, our findings suggest that shifts in attention are under voluntary rather than reflexive control. The demands of the Samson avatar task (e.g., working memory load) yield response times that would allow slow acting volitional shifts in attention to influence performance despite simultaneous presentation of the target and avatar, in line with reports of larger cue validity effects with increasing RT (Bukowski et al., 2016). Taken together, the results of both experiments provide converging evidence that the mere presence of an avatar does not exert an automatic and mandatory influence on cognition. This finding implies that the visual perspective-taking phenomenon measured by this task has limited ‘spontaneity’, casting doubt on the original versions of *both* the submentalising *and* implicit mentalising accounts of performance in the Samson avatar task.

Limitations to the spontaneity of the visual perspective-taking phenomenon measured by the Samson task have already been acknowledged (Bukowski et al., 2016; Ferguson et al., 2016). This led Bukowski et al. to revise the implicit mentalising account, suggesting that visual perspective-taking is context dependent and not mandatorily triggered by the avatar. Their view is that initiation of an implicit mentalising process (e.g., line-of-sight computation) is contingent upon sufficient attention being directed towards the avatar, either by the avatar appearing alone or by the task instructions providing a context that makes this stimulus more salient. Although this account is consistent with their finding that a cue validity effect was present at an SOA of 300ms and not at 0ms, it does not explain the absence of a cue validity effect at 100ms in Experiment 2. Consequently, we believe that the present findings are better accommodated by a revised submentalising account. By this account, the direction that the avatar faces influences performance in the Samson avatar task through fast voluntary shifts in spatial attention and top-down modulation by participants’ assessment of the task relevance of the central cue (Vaquero et al., 2010).

We propose that the schema theory of gaze cueing (Cole, Smith, & Atkinson, 2015) provides a useful theoretical framework for understanding performance in the Samson avatar task, given that it offers an explanation of social attention effects in terms of fast voluntary shifts in attention, modulated by top-down control of activation thresholds. This theory assumes that social attention occurs because social cues (gaze, head position, body orientation, etc.) trigger schemas in a bottom-up manner. Once activated, these schemas automatically execute shifts of spatial attention to the attended-to location. These schemas are considered to originate from repeated learned associations between the social cue and the attended-to location, rather than an innate reaction to biologically relevant stimuli (cf. Brignani, Guzzon, Marzi, & Miniussi, 2009). In this respect, social cues are taken to be no different from any arbitrary stimulus that has become associated with an attended-to location. This is consistent with the comparable results for arrows found in Experiments 1 and 2. This theory generates several interesting predictions, including that social attention should be strongest, with the appearance of being reflexive, in cases where the cue is most unambiguous (e.g., gaze aversion stimuli, see Cole et al., 2015). Where cues are more ambiguous, either by being impoverished or being less overlearned in the first place, schema triggering is weaker and more susceptible to top-down moderation.

The schema theory may thus reconcile apparently contradictory findings when applied to the research literature employing the Samson avatar task. It predicts that the Samson avatar task should be fairly susceptible to top-down moderation through task context, given that the avatar stimuli depict whole body orientation and thus provide more ambiguous cues to social attention than the averted eyes used in classic gaze aversion paradigms (Driver et al., 1999). In the experiment by Cole et al. (2016), a correspondence effect was found irrespective of whether line of sight was occluded by a barrier. This may be because the barrier manipulation, and explicit instructions about what the avatar could see, may have induced a perspective-taking context in this experiment, exciting activation values across both conditions. In the experiment by Conway et al. (in press), a correspondence effect was found irrespective of whether visual access was made available by an optical cloaking device. This may be because the cloaking device manipulation was sufficiently novel and outside normal experience that it did not exert an influence on activation thresholds. Finally, in the experiment by Furlanetto et al. (2016), a correspondence effect was found for an avatar wearing transparent but not opaque goggles, in line with an implicit mentalising interpretation. However, a goggle wearing avatar is potentially the most ambiguous social cue employed to date; by obscuring the eyes, both goggle conditions no longer provide an overlearned cue to attended-to direction. Schema theory would predict that such testing conditions would make social cueing highly susceptible to top-down modulation. Therefore, participants may have been motivated to follow the orientation of the avatar in the "Seeing" (transparent goggles) condition, given the strong social context induced, while participants may have been motivated to suppress influence of the cue in the "Non-seeing" (opaque goggles) condition, given the strong and explicit belief induction that the avatar was an irrelevant cue.

While the schema theory may offer a promising framework to understand putative evidence of implicit mentalising provided by the Samson avatar task, it is important to acknowledge a couple of caveats to the foregoing account. First, Conway et al. (in press) propose that the original results suggesting that correspondence effects occur for an avatar wearing transparent but not opaque goggles (Furlanetto et al., 2016) were a false positive. This claim was based upon a failure to replicate this original finding, and offers a different way to accommodate these results within a submentalising framework. Second, it is important to acknowledge a new version of the implicit mentalising theory that proposes that "cue-driven shifts in attention are integral to the fast and efficient calculation of what the avatar sees" (Furlanetto et al., 2016, p. 163). In fact, there is much common ground between schema theory and this more nuanced version of the implicit mentalising theory: For both theories, shifts in attention are integral that may be moderated by top-down expectations. What distinguishes these positions is that the schema theory is a submentalising hypothesis; it does not evoke mental state attribution, in contrast to Furlanetto et al. (2016). Furthermore, the schema theory may also be distinguished by the specific claim that social stimuli are not a special category of stimulus, and that "any overlearned cue–target association can become encoded as a schema" (Cole et al., 2015, p. 1113). Further research would be required to discriminate between these accounts, perhaps by examining dissociations between perspective-taking and attention orienting performance (see Todd, Cameron, & Simpson, 2017).

A more general implication of the results of Experiment 2 is that the spatial orientation of an observed whole body may have attention orienting properties, even when that body is in a passive stance. This finding is noteworthy, given the somewhat inconsistent findings of previous research. For instance, prior work appears to indicate that such a phenomenon may be contingent on being able to see the body in action (Gervais et al., 2010) or may be dependent upon the interaction between body and head position (Pomianowska, Germeys, Verfaillie, & Newell, 2012). Nonetheless, the cue-validity effects reported here for Samson avatar stimuli are consistent with similar effects recently reported by Bukowski et al. (2016). While our effect was restricted to 600ms SOA, Bukowski et al. found that 300ms SOA was sufficient. It is possible that this discrepancy was due to between-task differences (we employed a detection task, they employed a discrimination task), given that task influences the size of attention orienting effects (Chica, Martin-Arévalo, Botta, & Lupiánez, 2014). These findings encourage further work examining the attention orienting properties of whole body orientation; research employing an adaptation paradigm indicates that direction codes derived from body orientation are independent of codes derived from gaze (Lawson, Clifford, & Calder, 2009), and a review of the literature has suggested that body orientation is a relatively under-investigated social attention cue (Nummenmaa & Calder, 2008).

Finally, our account of perspective-taking in the Samson avatar task is consistent with the broader ‘submentalising’ re-interpretation of data purporting to show implicit mentalising in adults (Heyes, 2014). According to the broad submentalising position, reported demonstrations of ‘implicit mentalising’ may arise from the operation of general purpose cognitive mechanisms, rather than dedicated mechanisms specialized for ascribing mental states. Thus, our evidence that voluntary attention orienting contributes to performance in the Samson avatar task lends support to this proposal, albeit not the specific hypothesis that this phenomenon could be accounted for by reflexive attention orienting (Heyes, 2014; Santiesteban et al., 2014). The great value of adopting a submentalising stance is to move on from considering general purpose cognitive mechanisms as merely an experimental artifact. Instead, it invites further work that explores the extent to which domain general mechanisms subserve the fast and efficient behavioural coordination that are the essence of much everyday social cognition.

Conclusion

Taken together, our results have three main implications. First, they suggest that the body orientation of simple avatar characters has attention orienting properties comparable to directional arrow cues, implying that body orientation may be a hitherto neglected social attention cue. Second, our findings suggest that attentional orienting plays a role in the Samson avatar task. However, in contrast to the specific account proposed by Santiesteban *et al.* (2014)

and Heyes (2014), our data indicate that this attention orienting process appears to be voluntary rather than reflexive. This implies that the phenomenon measured by the Samson avatar task may be less spontaneous than first reported and moderated by top-down appraisal of the task context. Third, these results are generally consistent with the theory that implicit mentalising results from general purpose cognitive processes rather than a specialised mechanism for fast mental state tracking (Heyes, 2014). The schema theory of gaze-cueing (Cole et al., 2015) offers a potentially useful theoretical framework for future developments in this area.

Acknowledgments

This work was supported by the Experimental Psychology Society under a small research grant awarded to the first author. We are grateful to Dana Samson for generously making available the materials used in the avatar task, and to two anonymous referees for constructive comments on a previous version of this paper.

References

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

Brignani, D., Guzzon, D., Marzi, C. A., & Miniussi, C. (2009). Attentional orienting induced by arrows and eye-gaze compared with an endogenous cue. *Neuropsychologia*, *47*(2), 370-381.

Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. A. (2009). When you and I share perspectives: Pronouns modulate perspective taking during narrative comprehension. *Psychological Science*, *20*(1), 27-32

Bukowski, H., Hietanen, J. K., & Samson, D. (2016). From gaze cueing to perspective taking: Revisiting the claim that we automatically compute where or what other people are looking at. *Visual Cognition*, *23*(8), 1020-1042.

Chica, A. B., Martín-Arévalo, E., Botta, F., & Lupiánez, J. (2014). The Spatial Orienting paradigm: How to design and interpret spatial attention experiments. *Neuroscience & Biobehavioral Reviews*, *40*, 35-51.

Cole, G. G., Smith, D. T., & Atkinson, M. A. (2015). Mental state attribution and the gaze cueing effect. *Attention, Perception, & Psychophysics*, *77*(4), 1105-1115.

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

Conway, J.R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (in press). Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception & Performance*

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6*(5), 509-540.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual review of psychology*, *48*(1), 269-297.

Ferguson, H. J., Apperly, I., & Cane, J. E. (2016). Eye-tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *Quarterly Journal of Experimental Psychology*. DOI: 10.1080/17470218.2016.1199716

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.

Heyes, C. (2014). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science, 9(2)*, 131-143.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading.*Science*, *344*(6190), 1243091.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in ‘theory of mind’: Evidence from compromised belief-desire reasoning in old age. *Cognition*, *101*(1), 129-152.

Gervais, W. M., Reed, C. L., Beall, P. M., & Roberts, R. J. (2010). Implied body action directs spatial attention. *Attention, Perception, & Psychophysics*, *72*(6), 1437-1443.

Kingstone, A., Tipper, C., Ristic, J., & Ngan, E. (2004). The eyes have it!: An fMRI investigation. *Brain and cognition*, *55*(2), 269-271.

Lawson, R. P., Clifford, C. W., & Calder, A. J. (2009). About turn: The visual representation of human body orientation revealed by adaptation. *Psychological Science*, *20*(3), 363-371.

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption.*Journal of Experimental psychology: Human perception and performance*,*15*(2), 315.

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?. *The Quarterly Journal of Experimental Psychology*, *68*(10), 1931-1951.

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in cognitive sciences*, *13*(3), 135-143.

Pomianowska, I., Germeys, F., Verfaillie, K., & Newell, F. N. (2012). The role of social cues in the deployment of spatial attention: head-body relationships automatically activate directional spatial codes in a Simon task. *Frontiers in integrative neuroscience*, *6*, 4.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and performance X: Control of language processes*, *32*, 531-556.

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*(2), 230-236.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36(5)*, 1255-1266.

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing?. *Journal of Experimental Psychology: Human Perception and Performance, 40(3)*, 929-937.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Pittsburgh: Psychology Software Incorporated.

Schwarzkopf, S., Schilbach, L., Vogeley, K., & Timmermans, B. (2014). "Making it explicit" makes a difference: evidence for a dissociation of spontaneous and intentional level 1 perspective taking in high-functioning autism. *Cognition*, *131*(3), 345-354.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science*,*325*(5942), 883-885.

Stephenson, B., & Wicklund, R. A. (1983). Self-directed attention and taking the other's perspective. *Journal of Experimental Social Psychology*, *19*(1), 58-77.

Todd, A.R., Cameron, C.D., & Simpson, A.J. (2017). Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition, 159*, 97-101.

Vaquero, J. M., Fiacconi, C., & Milliken, B. (2010). Attention, awareness of contingencies, and control in spatial localization: A qualitative difference approach. *Journal of experimental* *psychology: human perception and performance*, *36*(6), 1342.

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.

Figure Captions

Figure 1: Schematic illustration of the trial structure employed in Experiment 1. The examples depict trials in which correct response was "yes", to confirm that the number of dots visible to the participant corresponded to the digit. For consistent trials, the avatar / arrow is oriented towards the same number of dots as those visible to the participant. Whereas for inconsistent trials, the avatar / arrow is oriented towards a different number (here, zero).

Figure 2: Data from Experiment 1. Response times in an adapted Samson avatar task as a function of whether the number of dots visible to the avatar and participant were consistent. In a control condition, the avatar was replaced by an arrow.

Figure 3: Schematic illustration of the trial structure employed in Experiment 2. The examples depict trials in which the target dot appeared on the right side (left side trials, and catch trials are not illustrated). For valid trials, the avatar / arrow is oriented towards the target. Whereas for invalid trials, the avatar / arrow is oriented away from the target.

Figure 4: Data from Experiment 2. Response times in a Posner task as a function of whether a central cue (avatar or arrow) was a valid or invalid predictor of a target at three cue-target onset asynchronies.