Seeing it both ways:

Using a double-cueing task to investigate the role of spatial cueing in level-1 visual perspective-taking

Forthcoming in: *Journal of Experimental Psychology: Human Perception and Performance*

John Michael[ab*], Thomas Wolf[b], Clément Letesson[b], Stephen Butterfill[a], Joshua Skewes[c], Jakob Hohwy[d]

Abstract

Previous research using the dot perspective task has produced evidence that humans may be equipped with a mechanism that spontaneously tracks others' gaze direction and thereby acquires information about what they can see. Other findings, however, support the alternative hypothesis that a spatial cuing mechanism underpins the effect observed in the dot perspective task. In order to adjudicate between these hypotheses, we developed a double-cuing version of Posner's (1980) spatial cueing paradigm to be implemented in the dot perspective task, and conducted three experiments in which we manipulated stimulus onset asynchrony as well as secondary task demands. Crucially, the two conflicting hypotheses generate divergent patterns of predictions across these experimental conditions. Our results support the hypothesis of an automatic perspective-taking mechanism.

[a] University of Warwick, UK

[b] Central European University, Hungary

[c] Aarhus University, Denmark

[d] Monash University, Australia

* Corresponding Author: j.michael.2@warwick.ac.uk

Keywords: level 1 visual perspective-taking, theory of mind, attention, spatial cueing, implicit processing

Public Significance Statement

Recent research has revealed evidence that humans are equipped with a perspective-taking mechanism that spontaneously tracks others' gaze direction and thereby acquires information about what others can see. This research has been controversial however, with critics arguing that the evidence in question can also be interpreted by appealing to general attentional mechanisms without postulating a perspective-taking mechanism. To adjudicate between these two competing theoretical positions, we conducted a series of experiments for which the two positions generate conflicting predictions. Our findings support the hypothesis of a perspective-taking mechanism that is distinct from general attentional mechanisms. This result implies that the general experimental paradigm used in this context can be used as a tool to investigate perspective-taking abilities – not only in neurotypical children and adults but also in individuals with disorders of social cognition such as autism and schizophrenia.

Introduction

The ability to track others' gaze direction and to infer what they can see (a process often referred to as level-1 visual perspective-taking (Flavell, Everett, Croft, & Flavell, 1981)) is an important component of human social cognition. It enables us to acquire information about others' mental states (e.g. what they want or intend), and thereby helps us to anticipate their actions, and to communicate and coordinate fluently with them. Previous research using the dot perspective task (Samson et al., 2010; Qureshi et al., 2010) has produced evidence that humans are equipped with a mechanism that spontaneously performs level-1 visual perspective-taking.

In the dot perspective task, participants view an image of a room with an avatar standing in the middle and facing either leftward or rightward (this is varied from one trial to the next). On each trial, anywhere from 0-3 red dots are displayed on the walls of the room – sometimes all on one side, sometimes distributed between both sides. On half of the trials, the avatar can see all of the dots (e.g. she is facing to the left; three dots appear on the left wall but none appear on the right wall), so the perspectives of the participant and the avatar are consistent (Consistent trials). On the other half of the trials, the avatar can see some but not all of the dots (e.g. she is facing to the left; one dot appears on the left and one on the right), or none of the dots (i.e. all of the dots are on the wall behind her). On these trials, the perspectives of the participant and the avatar are inconsistent (Inconsistent trials). Participants have the task of calculating either how many dots the avatar can see (Other trials), or how many dots they themselves see (Self trials). A main finding is that participants perform worse on Inconsistent Self trials than on Consistent Self trials. The authors conclude that participants calculate the avatar's perspective even on trials for which they need only calculate their own (namely, Self trials), and that computing the avatar's perspective interferes with reporting their own (Samson et al., 2010). This is an *altercentric interference effect*: another's task-irrelevant perspective impairs performance.

In a follow-up study using the same paradigm, Qureshi, Apperly, & Samson (2010) exposed participants to an additional cognitive load, and found that the interference from inconsistent perspectives increased. The authors interpret this as evidence that participants automatically calculated the avatar's perspective (level-1 perspective taking) in parallel to the calculation of what they themselves could see. In contrast, the selection of a perspective to draw upon in forming a response is a controlled process requiring executive resources, and was therefore impaired by the cognitive load manipulation. This pattern of results is suggestive of an automatic mechanism for level-1 visual perspective-taking.

However, as Heyes and colleagues have pointed out (Heyes, 2014; Santiesteban et al., 2014), it is possible that this task does not tap a mechanism for level 1 visual perspective-taking but, rather, spatial cueing, with the avatars serving as cues to facilitate attentional processing either on the left or the right side. Indeed, this interpretation is supported by the findings of Santiesteban et al. (2014), who replicated Samson et al. (2010)'s effect using arrows instead of avatars.

While Santiesteban and colleagues' findings are consistent with the hypothesis that a spatial cueing mechanism underpins the effects observed in the dot perspective task, their results are not decisive. First, it is possible that participants' prior experiences with arrows lead them to interpret the arrows as indicating an implied perspective (e.g. the perspective of an agent who places an arrow and/or the perspective of an agent who looks in the direction indicated by an arrow). Secondly, Santiesteban and colleagues' results do not rule out the possibility that the version of the paradigm with arrows taps a *different* underlying mechanism that leads to a similar pattern of findings in the particular circumstances of these experiments. To rule this out, it would be important to specify experimental conditions in which the two hypothesized mechanisms should lead to dissimilar patterns of findings (and indeed that is what the present study accomplishes, as we explain below).

There are at least two reasons why it is important to adjudicate between these two competing hypotheses. The first is that they lead to conflicting views of the relevance of the dot perspective task for disorders of social cognition such as autism. For example, if the task indeed taps a mechanism for automatically calculating others' perspectives, then one may expect autistic participants not to exhibit the altercentric interference effect, at least insofar as one is persuaded by evidence from previous research suggesting that autistic persons tend not to spontaneously engage in perspective-taking or other forms of mindreading, and must instead mobilize conscious cognitive effort in order to do so (Schneider et al., 2013; Senju et al., 2009; Hamilton et al., 2009). And yet, interestingly, Schwarzkopf and colleagues (2014) did in fact observe the altercentric interference effect in autistic participants. Thus, if the perspective-taking hypothesis is correct, then this finding indicates a need to reconsider our understanding of perspective-taking in autistic persons. If, however, the task does not tap a perspective-taking mechanism, then Schwarzkopf and colleagues' results may not bear so directly on differences between autistic and non-autistic persons' spontaneous perspective-taking. Instead, Schwarzkopf and colleagues' result may in this case be expected to generalize to spatial cueing paradigms.

A second reason why it is important to adjudicate between these two competing hypotheses is that the dot perspective paradigm is increasingly being relied upon as a tool for investigating the nature and limitations of fast and efficient mindreading processes (Qureshi, Apperly, & Samson, 2010; Furlanetto et al., 2016; Schwarzkopf et al., 2014), and thus also as a means of testing theories of the cognitive architecture of mindreading (Butterfill & Apperly, 2013; Christensen & Michael, 2015, Westra, 2016). If, however, the paradigm does not in fact tap a perspective-taking process, then it may be necessary to re-evaluate these uses of it.

To investigate whether the effects observed in the dot perspective task are due, at least in part, to spatial cueing, we adapted Maylor's (1985)'s 'double-cueing' task (see also Posner & Cohen, 1984). In the basic spatial cueing task (Posner, 1980), participants are instructed to detect the appearance of a target either on the left or the right side of a screen. In the double-cueing version of this task, the appearance of the target is preceded by two simultaneous peripheral cues, one on the left and one on the right. The main finding is that target detection on either side is facilitated: regardless of which side the target appears on, target detection is faster than in a baseline condition with no cue (Maylor, 1985, exp. 3; Posner & Cohen, 1984). The authors conclude that attention can be concurrently facilitated at two locations.

Building upon this result, we reasoned that if a spatial cueing mechanism underlies the effect observed in the dot perspective task, then it should also be possible for two avatars facing in different directions (left and right) to cue attention to two different locations simultaneously, and thereby to facilitate attentional processing at both locations. Hence, on a version of the dot perspective task in which all test trials involve dots on both walls, participants should perform better on trials with two avatars (one facing in each direction) than on trials with just one avatar or with no avatar. This is because, if spatial cueing underlies the effect, the leftward and rightward orientations of two avatars should facilitate attentional processing at both locations (left and right), whereas the directional orientation of one avatar would only facilitate processing of objects at one location (left or right), and a room with no avatar would provide no facilitation. (This hypothesis does not generate a clear prediction about whether performance should be better on trials with 1 or 0 avatars. This is because, when there is 1 avatar, the discs on one side of the room are cued, but the discs on the other side are uncued, which may inhibit processing of the uncued side. If so, it is possible that these two influences may cancel each other out. It is therefore unclear what the net effect of these two influences may be.)

In contrast, if the effect observed in the dot perspective task is driven by a mechanism for perspective-taking, then one should predict a different pattern of findings. On this account, the observed effect results from the interference of an inconsistent perspective rather than from facilitation of attentional processing. Thus, given that all trials involve dots on both walls, and assuming that the perspective-taking mechanism can compute two avatars' perspectives, performance should be no better on trials in which there are two avatars (each with a perspective that is inconsistent with the perspective of the participant) than on trials with one avatar (with just one perspective that is inconsistent with that of the participant) or on trials with no avatar (i.e. no inconsistent perspective). (This hypothesis also predicts that performance should be worse on trials with 1 avatar than on trials with 0 avatars: even just one inconsistent perspective should interfere with performance.)

In order to test these conflicting predictions, we ran three separate experiments with different participants. In each experiment we asked the same question: When there are two avatars rather than just one in Self Inconsistent trials, is performance better or worse? In Experiment 1, we created a situation most likely to reveal a cueing effect by including a delay of 800 ms after the avatar(s) appeared and before the discs appeared on the walls. The inclusion of such a stimulus onset asynchony (SOA) has been shown to be necessary in some gaze cueing paradigms (Friesen & Kingstone, 1998; Driver, 1999; Frischen et al., 2007; Xu, Tanaka, & Mineault, 2012), including a gaze cueing paradigm that used the stimulus material from the dot perspective task and in which the overall latencies were matched with the dot perspective task (Bukowski, Hietanen, & Samson, 2016). The inclusion of an SOA may therefore be necessary in order to observe the facilitative effect of a spatial cueing mechanism.

However, the standard dot perspective task (i.e. as in Samson et al., 2010) does not include an SOA. Further, the inclusion of an SOA, as in Experiment 1, may mask the effects of a perspective-taking mechanism by providing participants with extra time to allocate their attention in accordance with the directional information extracted from the stimuli. Thus, if the results from Experiment 1 were to reveal that performance was better with two avatars than with one (as we would expect if a spatial cueing mechanism were at work), this would indicate that the stimuli and the (number verification) task used in the dot perspective paradigm can indeed elicit a spatial cueing mechanism, but it would not settle the question of whether a spatial cueing mechanism is responsible for the altercentric effect observed in the standard dot perspective task with no SOA. We therefore carried out Experiment 2, which differed from Experiment 1 only in that there was no SOA. We reasoned that if cueing is responsible for how the avatar affects performance in the standard dot perspective task, then since there is no SOA in that task, this cueing mechanism should also be activated in this version of our task with no SOA. In that case, we should observe better performance with two avatars than with one. If, on the other hand, we were to observe worse performance with two avatars rather than one, this would be consistent with the hypothesis of a perspective-taking mechanism.

Experiment 3 was designed to provide a more stringent test of the hypothesis of an automatic perspective-taking mechanism. We reasoned as follows. Suppose perspective taking really does occur when there is an SOA (as in Experiment 1) but its effects are masked because the SOA provides participants with extra time to allocate their attention. In that case, it should be possible to unmask its effects by instructing participants to perform a secondary task designed to tax the executive and thereby to interfere with the operation of a spatial cueing mechanism. That is, combining the SOA with a secondary task should mean that performance is worse with two avatars than one if the perspective-taking hypothesis is correct. Hence, the pattern of results from Experiment 3 should more closely resemble the pattern from Experiment 2 than the pattern from Experiment 1 (Tables 1a and 1b summarize the different patterns of predictions generated by the perspective-taking hypothesis and the spatial cueing hypothesis).

Experiment 1

Method

Participants

In determining the appropriate sample size, we used Samson et al. (2010) as our starting point. In Samson and colleagues' study, each of the three experiments included a sample size of 16. Their Experiment 3 most closely resembles the design we used here, since (i) participants in that experiment (as in our design) were only ever asked to calculate how many dots they themselves could see, and to ignore the distractor in the middle of the room; and (ii) since they varied the type of distractor in the middle of the room (a rectangle or an avatar). Based on the effect sizes observed by Samson and colleagues in their Experiment 3 ($\eta_p 2$ =.186 and $\eta_p 2$=.285), we determined that for 80% statistical power and with an alpha level of .05, the appropriate sample size for our study would be twenty. We therefore recruited twenty participants (11 females; age range: 18-24, M = 20.75, SD = 1.71) from student organizations in the Budapest area, all of whom received gift vouchers for their participation. All were naïve to the purpose of the study, reported normal or corrected to normal vision, and signed informed consent prior to the experiment. The experiment was conducted in accordance with the Declaration of Helsinki and was approved by the (EPKEB) United Ethical Review Board for Research in Psychology.

Procedure

'PsychoPy' software was used to control the stimulus presentation and data collection (Peirce, 2007). As in Samson et al. (2010), the stimuli consisted of a picture showing a lateral view into a room with the left, back and right walls visible and with red dots displayed on one or two walls. Images of a human avatar were produced from the image files used by Samson et al. (2010), and were positioned in the center of the room. On one-third of all trials, there was one avatar facing either the left or the right wall. On one-third of all trials, there were two avatars, one facing left and one facing right. On one-third of all trials, there was no avatar present (See Figure 1).

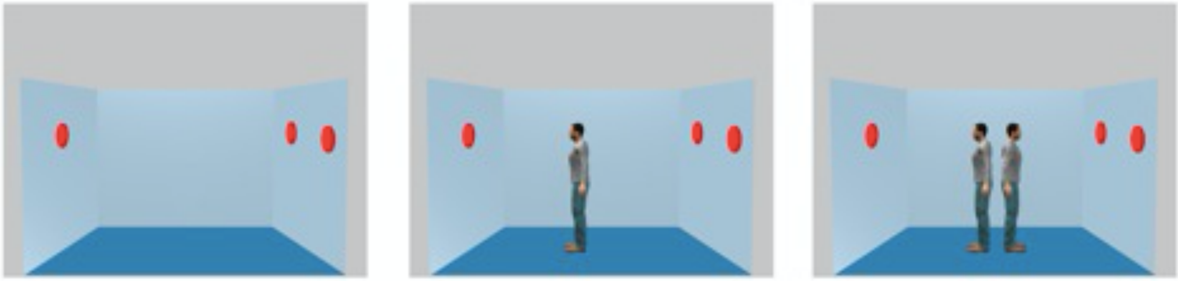


Figure 1: Examples of stimuli used in the three conditions (0, 1 or 2 avatars)

Following the display of a fixation cross for 500 ms, a digit (0-4) appeared for 750 ms, which specified a target number of dots for the participant to verify. The image of the room then appeared with the dots on the walls and with 0, 1 or 2 avatars in the center, followed by an 800 ms stimulus onset interval (SOA) before the dots appeared. This image remained until a response was given (or until 2000 ms passed). On matching trials ('yes' response), the digit specifying the target number corresponded to the number of dots on the walls. On mismatching trials ('no' response), the digit specified a number either one higher than or one lower than the number of dots on the walls (See Figure 2).

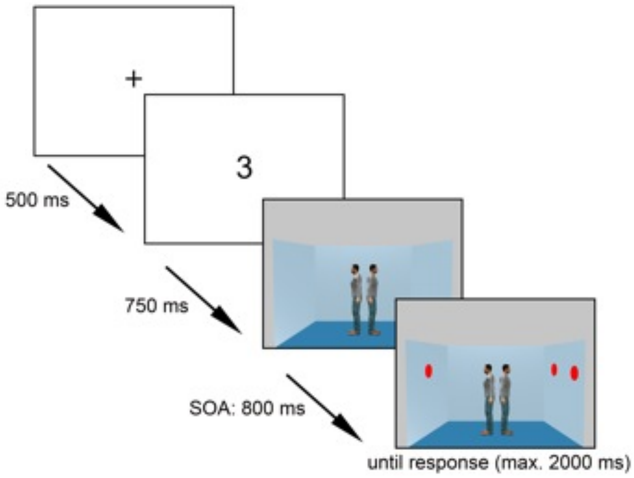


Figure 2: Trial structure. For this example, the correct response was 'yes'.

Female participants were presented with female avatars and male participants with male avatars. On all test trials, there was at least one dot on each wall, and never more than three on any wall, or more than four in total. As a result, the participant's perspective was always inconsistent with that of the avatar(s): for every avatar that appeared, there was always at least one dot behind her/him and one in front of her/him.

There were 48 matching ('yes' response) and 48 mismatching ('no' response) trials for each condition (0-avatars, 1-avatar, 2-avatars). On mismatching trials, the digit presented at the beginning of the trial sometimes corresponds to the number of discs visible from an avatar's perspective (i.e. in the 1-avatar condition and in the 2-avatar condition), making such trials particularly difficult. Since the frequency of such trials differs among the three conditions, we followed Samson and colleagues in treating mismatching trials as fillers, and analyzed only matching trials. Thus, there were 144 test trials, 48 per condition. We also included 27 additional filler trials where there were dots on only one wall so that '1' would sometimes be the correct response, and so that participants could not reliably anticipate whether there would be dots on both walls. These additional filler trials included an equal number of 0-avatar, 1-avatar and 2-avatar trials. The trials were divided into 3 blocks of 105 trials (48 test trials and 57 filler trials) and were preceded by a block of 26 practice trials. The order of the trials within each block was pseudo-randomized and then fixed across participants so that there were no more than 3 consecutive trials of the same type.

Results and Discussion

To control for speed-accuracy tradeoffs, reaction time (RT) for correct responses and hit rates (HR) were also merged into inverse efficiency scores (IES), a combined measure which homogenizes different patterns of speed-accuracy trade-offs within a group (IES=RT/HR; Townsend & Ashby, 1978). Since the calculation of IES entails that RTs are quasiexponentially multiplied as the HR decreases, Bruyer & Brysbaert (2011) have recommend not using the IES unless the mean HR within a group is above 90%. In our sample, the mean HR was above 90% in all three conditions, indicating that it was appropriate to use IES for the primary analysis. Further below, we also include analyses of the RTs and HRs.

In calculating mean reaction times (RTs), response omissions due to the timeout procedure (0.31% of the data) and erroneous responses (3.94% of the data) were eliminated from the data set. We also removed trials with responses that were more than 2.5 *SD*s greater or less than the mean for each participant for each condition (2.83% of the data).

Inverse Efficiency Score Analysis

We performed a three-way ANOVA for IES, which revealed a significant main effect of number of avatars, with performance being better in the 2-avatar condition (M = 661.65, SD = 128.38) than in the 0-avatar condition (M= 678.05, SD= 130.91), and the 1-avatar condition (M= 681.20 SD= 134.75) ($F(2,18) = 5.30$, $p = .009$, $\eta_p^2 = 0.218$). This is consistent with the operation of a spatial cueing mechanism. Planned contrast analyses revealed a significant difference between the 2-avatar condition and the 1-avatar condition, with performance being significantly better in the former than in the latter (t (19) = 3.51, p = .002, d = 0.149), and also between the 2-avatar condition and the 0-avatar condition, with performance being significantly better in the former than in the latter (t (19) = 2.63, p = .016, d = 0.127). Both of those results are consistent with the hypothesis of a spatial cueing mechanism and not with the hypothesis of a perspective-taking mechanism. There was no significant difference between the 1-avatar and the 0-avatar conditions (t (19) = 0.424, p = .676, d= 0.024) (See Figure 3).

Reaction Time Analysis

We performed a three-way ANOVA for reaction times (RTs), which revealed a significant main effect of number of avatars, with performance being better in the 2-avatar condition (*M* = 634.35, *SD* = 117.26) than in the 0-avatar condition (*M*= 652.11, *SD*= 122.94), and the 1-avatar condition (649.20 *SD*= 130.42), ($F(2,18) = 5.86$, $p = .006$, $\eta_p 2 = 0.236$). This is consistent with the operation of a spatial cueing mechanism (See Figure 4).

Planned contrast analyses revealed a significant difference between the 2-avatar condition and the 0-avatar condition, with performance being significantly better in the former than in the latter (t (19) = 3.10, $p = .006$, $d = 0.148$), and also between the 2-avatar condition and the 1-avatar condition, with performance being significantly better in the former than in the latter (t (19) = 2.49, $p = .022$, $d = 0.127$). Both of those results are consistent with the hypothesis of a spatial cueing mechanism and not with the hypothesis that a perspective-taking mechanism underlies the effects of avatars on performance in this task. There was no significant difference between the 1-avatar and the 0-avatar conditions (t (19) = 0.588, $p = .56$, d= 0.024).

Accuracy Analysis

We performed a three-way ANOVA for hit rates (HRs), which revealed no significant differences among the in the 2-avatar condition (*M* = 96.11%, *SD* = 3.44%), the 0-avatar condition (*M*= 96.28, *SD*= 2.33), and the 1-avatar condition (95.37% *SD*= 4.28%), ($F(2,18) = 1.33$, $p = .275$, $\eta_p 2 = 0.236$), (See Figure 5).

Planned contrast analyses revealed no significant differences among the conditions: neither between the 2-avatar condition and the 0-avatar condition, ($t(19) = .377$, $p = .71$, $d = 0.058$), nor between the 2-avatar condition and the 1-avatar condition ($t(19) = 1.3$, $p = .211$, $d = 0.19$), nor between the 1-avatar and the 0-avatar condition ($t(19) = .126$, $p = .224$, d= 0.263).

Figure 3: Inverse Efficiency Scores (IES). Error bars represent the within-subject confidence intervals (following the method proposed by Cousineau, 2005; cf. Loftus & Masson, 1994). Symbols indicate significance level (*** p<0.001; ** p<0.01; * p<0.05; ns= non-significant).

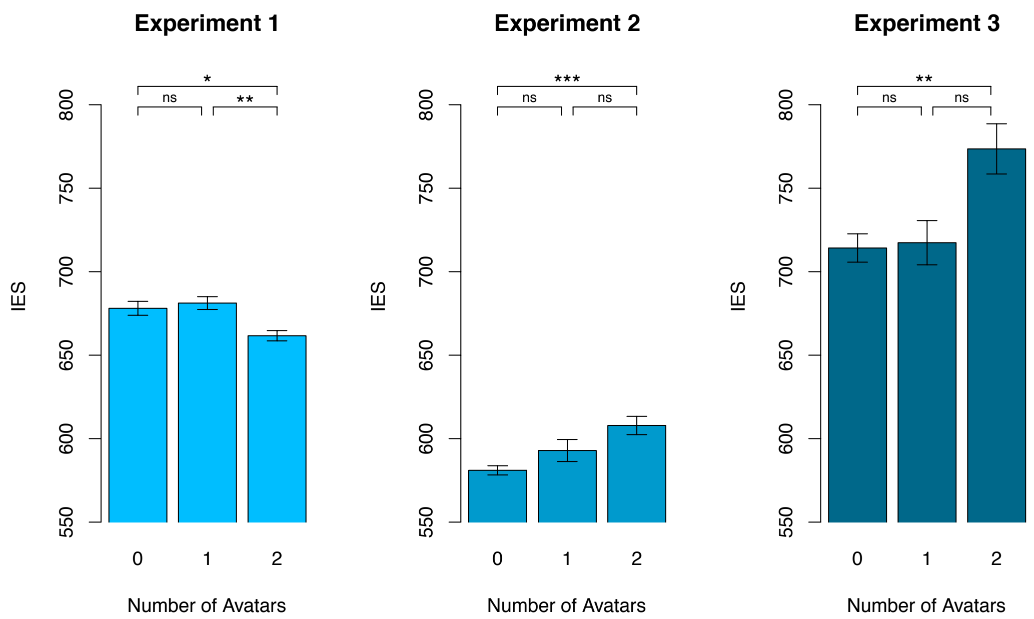


Figure 4: Reaction Times. Error bars represent the within-subject confidence intervals (following the method proposed by Cousineau, 2005; cf. Loftus & Masson, 1994). Symbols indicate significance level (*** p<0.001; ** p<0.01; * p<0.05; ns= non-significant).

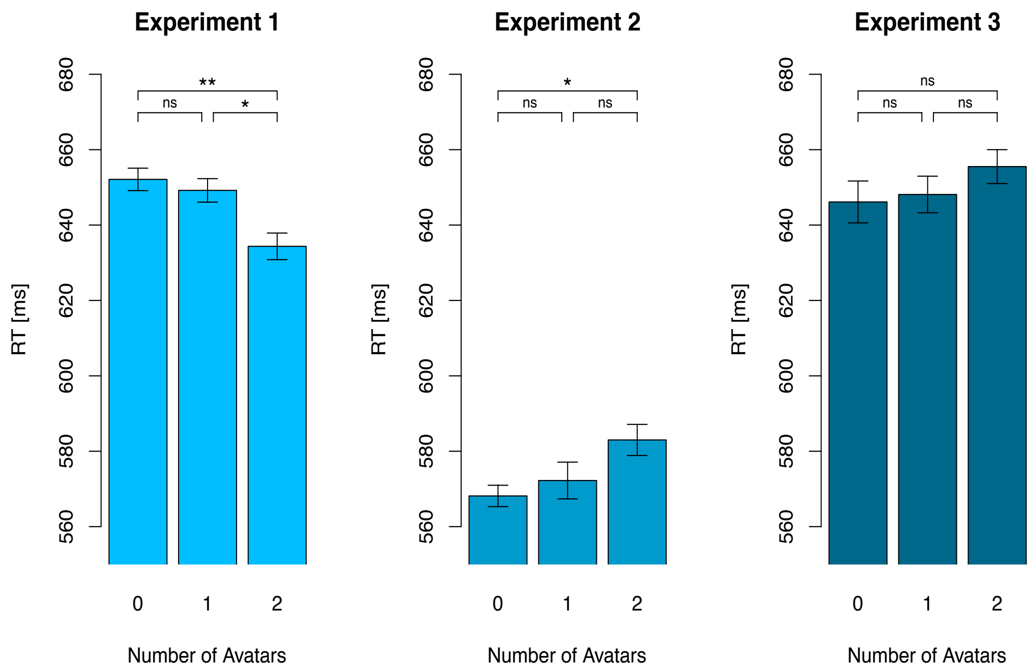
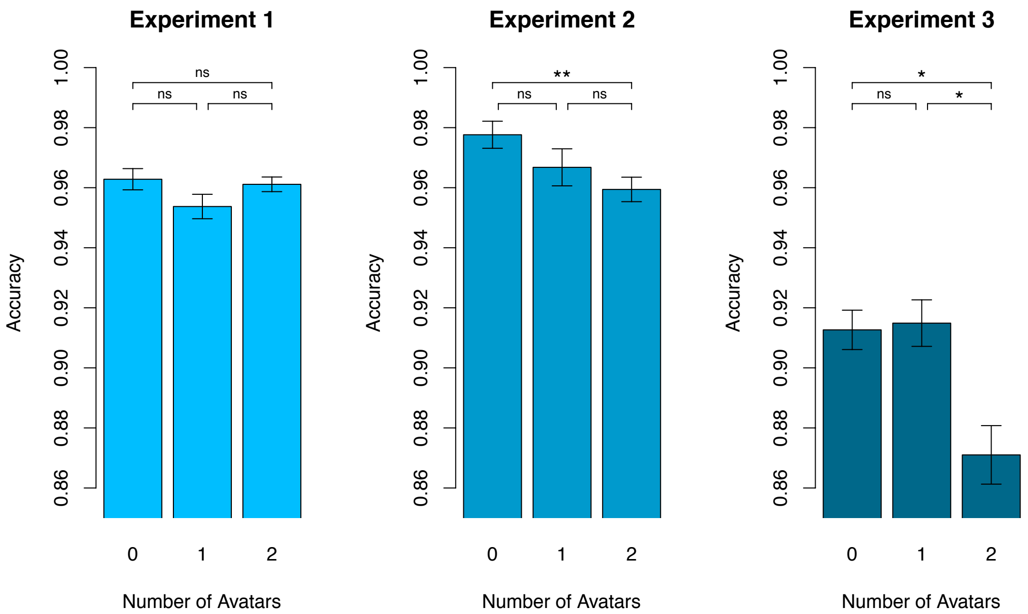


Figure 5: Hit Rates. Error bars represent the within-subject confidence intervals (following the method proposed by Cousineau, 2005; cf. Loftus & Masson, 1994). Symbols indicate significance level (*** p<0.001; ** p<0.01; * p<0.05; ns= non-significant).



Experiment 2

The results of Experiment 1 indicate that the stimuli and the number verification procedure used in the dot perspective task can be used to trigger a spatial cueing mechanism, and that attention can be facilitated at two locations with avatars oriented in opposite directions. This is consistent with the operation of a spatial cueing mechanism. However, it would not be justified to conclude that the standard dot perspective task taps a spatial cueing mechanism and not a perspective-taking mechanism. This is because Experiment 1, like gaze cueing paradigms but unlike the standard dot perspective task, included an SOA. We do not know what effect an SOA may have on how perspective taking processes influence judgments. In order to support the view that a spatial cueing mechanism underlines performance on the standard dot perspective task, it would be necessary to observe the same pattern of findings with no SOA. The aim of Experiment 2 was to do just this.

Method

Participants

Twenty participants (12 females; age range: 20-31, *M* = 25.67, *SD* = 3.16.) were recruited from student organizations in the Budapest area, and received gift vouchers for their participation. A statistical power analysis for a one-way repeated measures ANOVA with three levels using G*Power 3.1 (Faul et al., 2009) confirmed that, based upon the effect size ($\eta_p 2$ = 0.218) observed in experiment 1 (and for 80% statistical power and an alpha level of .05), twenty was the appropriate sample size for experiment 2. For the analyses, we excluded the data from two participants: one of these had an IES more than 3 *SD* greater than the mean for the group, and the other failed to complete the task. All were naïve to the purpose of the study, reported normal or corrected to normal vision, and signed informed consent prior to the experiment. The experiment was conducted in accordance with the Declaration of Helsinki and was approved by the (EPKEB) United Ethical Review Board for Research in Psychology.

Procedure

The procedure was the same as in Experiment 1 with one exception: when the image of the room with 0,1 or 2 avatars appeared, the dots on the walls appeared simultaneously, i.e. there was no SOA. The image remained until a response was given (or until 2000 ms passed), as in Experiment 1.

Results and Discussion

As for Experiment 1, in calculating mean reaction times (RTs), response omissions due to the timeout procedure (0.17 % of the data) and erroneous responses (3.86% of the data) were eliminated from the data set. We also removed trials with responses that were more than 2.5 *SD*s greater or less than the mean for each participant for each condition (2.78% of the data).

Inverse Efficiency Score Analysis

We then performed a three-way ANOVA for IES, which revealed a significant main effect of number of avatars, with performance being worse in the 2-avatar condition (M = 607.87, SD = 120.80) than in the 1-avatar condition (M= 592.87, SD = 122.85) or the 0-avatar condition (M= 581.02, SD = 111.49), ($F(2,16) = 4.47$, $p = .019$, $\eta_p^2 = 0.208$) (See Figure 3). A planned contrast analysis revealed a significant difference between the 2-avatar condition and the 0-avatar condition, with performance being worse in the 2-avatar condition than in the 0-avatar condition (t (17) = 4.79, $p < .001$, $d = 0.23$). These results are consistent with the hypothesis of a perspective-taking mechanism and difficult to account for by appealing to the operation of a spatial cueing mechanism. There was no significant difference between the 2-avatar condition and the 1-avatar condition (t (17) = 1.27, $p = .221$, $d = 0.123$), nor between the 1-avatar and the 0-avatar conditions (t (17) = 1.639, $p = .118$, $d = 0.1$) (See Figure 3).

Reaction Time Analysis

We performed a three-way ANOVA for RT, which revealed no significant difference among the 2-avatar condition (M=582.00, SD=116.66), the 1-avatar condition (572.23 SD=114.45) and the 0-avatar condition (M=568.14, SD=113.15) ($F(2,16) = 2.41$, $p = .104$, $\eta_p^2 = 0.124$) (See Figure 4).

Planned contrast analyses revealed a significant difference between the 2-avatar condition and the 0-avatar condition, with performance being worse in the 2-avatar condition than in the 0-avatar condition (*t*(17) = 2.90, *p* = .01, *d* = 0.23). This result is consistent with the hypothesis of a perspective taking mechanism and not with the hypothesis that a spatial cueing mechanism underlies the effects of avatars on performance in this task. The difference between the 2-avatar condition and the 1-avatar condition did not reach significance (*t* (17) = 1.26, *p* = .23, *d* = 0.093); nor did the difference between the 1-avatar and the 0-avatar conditions (*t*(17) = 0.60, *p* = .56, *d*= 0.04).

Accuracy Analysis

We performed a three-way ANOVA for hit rates (HRs), which revealed no significant differences among the 2-avatar condition (*M* = 95.94%, *SD* = 2.65%), the 0-avatar condition (*M*= 97.76%, *SD*= 2.84%), and the 1-avatar condition (96.68% *SD*= 2.95%), (*F*(2,16) = 2.24, *p* = .112, $\eta_p 2$ = 0.236) (See Figure 5).

Planned contrast analyses revealed a significant difference between the 2-avatar condition and the 0-avatar condition, with performance being worse in the 2-avatar condition than in the 0-avatar condition (*t* (17) = 3.04, *p* = .007, *d* = 0.663). This result is consistent with the hypothesis of a perspective taking mechanism and not with the hypothesis that a spatial cueing mechanism underlies the effects of avatars on performance in this task. The difference between the 2-avatar condition and the 1-avatar condition did not reach significance (*t*(17) = 0.78, *p* = .446, *d* = 0.262); nor did the difference between the 1-avatar and the 0-avatar conditions (*t*(17) = 1.08, *p* = .29, *d*= 0.375).

Experiment 3

In Experiment 2, the pattern of results we found in Experiment 1 is reversed. The reversal indicates that a perspective-taking mechanism, rather than a spatial cueing mechanism, may underpin performance on standard dot perspective tasks. However, to be at all confident in this interpretation we must further investigate the effects of the SOA. Taken at face value, the results of Experiments 1 and 2 suggest that there may be two separate mechanisms at work, with the perspective-taking mechanism predominating at earlier time points and the spatial cueing mechanism predominating at later time points. If so, it may be possible to selectively intervene on the spatial cueing mechanism and prolong the effects of the perspective-taking mechanism.

To test whether this is indeed possible, we re-introduced the 800ms SOA in Experiment 3, but also instructed participants to concurrently perform a secondary task. We reasoned that the secondary task may interfere with the spatial cueing mechanism, since there is evidence that attention shifts in response to central cues can be inhibited through the use of concurrent secondary tasks to increase processing demands (Jonides, 1981; Müller & Rabbitt, 1989; Frischen et al., 2007). In light of Qureshi, Apperly & Samson (2010)'s finding that the perspective-taking mechanism was not inhibited by the concurrent performance of a secondary task, we predicted that it would predominate in this version of the task, and that we would therefore again observe better performance with 1 avatar than with 2. For the secondary task, we chose an auditory-tone-monitoring task, and recorded verbal responses in order to rule out any visuospatial or motor interference with the dot perspective task.

Method

Participants

A statistical power analysis for a one-way repeated measures ANOVA with three levels using G*Power 3.1 (Faul et al., 2009) confirmed that, based upon the effect size ($\eta_p 2$ = 0.218) observed in experiment 1 (and for 80% statistical power and an alpha level of .05), twenty was the appropriate sample size for experiment 2. Twenty participants (10 females; age range: 21-30, *M*= 24.77, *SD* = 2.63.) were therefore recruited from student organizations in the Budapest area, and received gift vouchers for their participation.

All were naïve to the purpose of the study, reported normal or corrected to normal vision, and signed informed consent prior to the experiment. The experiment was conducted in accordance with the Declaration of Helsinki and was approved by the (EPKEB) United Ethical Review Board for Research in Psychology.

Procedure

The procedure for the dot perspective task was the same as in Experiment 1. In addition, however, participants performed an additional practice block of trials for the auditory-tone-monitoring task, and in all non-practice trials they concurrently performed the auditory-tone-monitoring task. For the auditory task, two tones were presented over a pair of headphones during the 800 ms. The first tone was presented 100ms after the image of the room appeared, and the second tone was presented 150 ms later, i.e. 250 ms after the appearance of the image of the room with the avatar(s). On half of the trials, the tones were presented at the same pitch. On the other half of the trials, one of the tones was presented at a high pitch and the other at a low pitch. Participants were instructed to give a verbal response (by saying 'Same!' into a microphone) if the two tones were the same, and otherwise to give no response. They were instructed to make their responses (if at all) as quickly as possible, and in any case before the discs on the wall appeared. Participants first performed one practice block of the dot perspective task, as in Experiments 1 and 2. Then they performed one practice block of the auditory-tone-monitoring task alone, without the dot perspective task. Next, they performed a practice block of the dot perspective task in conjunction with the auditory-tone-monitoring task. After these three practice blocks, they proceeded to the test blocks.

Results and Discussion

As for Experiments 1 and 2, in calculating mean reaction times (RTs), response omissions due to the timeout procedure (1.12 % of the data) and erroneous responses (7.05% of the data) were eliminated from the data set. We also removed trials with responses that were more than 2.5 *SD*s greater or less than the mean for each participant for each condition (3.18 % of the data).

Inverse Efficiency Score Analysis

We then performed a three-way ANOVA for IES, which revealed a significant main effect of number of avatars, with performance being worse in the 2-avatar condition (M = 773.56, SD = 261.83) than in the 1-avatar condition (717.37 SD= 188.80) or the 0-avatar condition (M= 714.20, SD = 190.60), ($F(2,18) = 4.71$, $p = .015$, $\eta_p 2 = 0.199$) (See Figure 3). A planned contrast analysis of performance in the 2-avatar condition and the 1-avatar condition revealed a marginally significant difference, with performance in the 2-avatar condition being worse than in the 1-avatar condition ($t(19) = 2.08$, $p = .051$, $d = 0.246$). A planned contrast of performance in the 2-avatar condition and the 0-avatar condition revealed a significant difference between the 2-avatar condition and the 0-avatar condition, with performance being worse in the 2-avatar condition than in the 0-avatar condition ($t(19) = 2.89$, $p = .009$, d = 0.259). There was no significant difference between the 1-avatar and the 0-avatar conditions ($t(19) = 0.194$, $p = .849$, $d = 0.017$) (See Figure 3).

On the auditory-tone-monitoring task, the overall hit rate for the group was 78.49%. Mean reaction time, (from the initiation of the second tone) was 354.83 ms (SD=210.91). We decided not to exclude any participants on the basis of their performance on the secondary task, since it is not possible to infer how distracting a participant found this task from performance (some will have found the task more difficult than others), and the purpose of this task was to distract participants from the dot perspective task.

Reaction Time Analysis

We performed a three-way ANOVA for reaction times (RTs), which revealed no significant difference among the 2-avatar condition (M = 655.51, SD =160.30), the 0-avatar condition (M= 646.13, SD= 153.39), and the 1-avatar condition (M= 648.11 SD= 144.46), ($F(2,18) = 0.66$, $p = .525$, $\eta_p^2 = 0.033$) (See Figure 4).

Planned contrast analyses revealed no significant difference between the 2-avatar condition and the 0-avatar condition (t (19) = 3.10, p = .006, d = 0.148), nor between the 2-avatar condition and the 1-avatar condition (t (19) = 2.49, p = .022, d = 0.127), nor between the 1-avatar and the 0-avatar conditions (t (19) = 0.588, p = .56, d= 0.024).

Accuracy Analysis

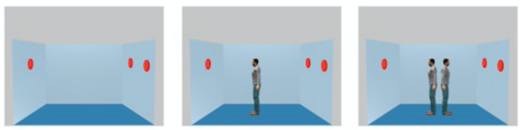
We performed a three-way ANOVA for hit rates (HRs), which revealed a significant main effect of number of avatar, with performance being worse in the 2-avatar condition (M = 87.10%, SD = 10.27%) than in the 0-avatar condition (M= 91.27%, SD= 6.49%), and the 1-avatar condition (91.49% SD= 8.02%), ($F(2,18) = 6.17$, $p = .005$, $\eta_p^2 = 0.250$), (See Figure 5).

Planned contrast analyses revealed a significant difference between the 2-avatar condition and the 0-avatar condition, with performance being significantly worse in the former than in the latter (t(19) = 2.83, p = .011, d = 0.485), and also between the 2-avatar condition and the 1-avatar condition, with performance being significantly worse in the former than in the latter (t(19) = 2.69, p = .015, d = 0.476). These result are consistent with the hypothesis of a perspective taking mechanism and not with the hypothesis that a spatial cueing mechanism underlies the effects of avatars on performance in this task. There was no significant difference between the 1-avatar and the 0-avatar conditions ($t(19) = 0.213$, $p = .83$, $d = 0.031$).

Table 1a. The two competing hypotheses.

	What underlies performance on Inconsistent Self trials in Samson et al., (2010)?	How should the difference in performance on Consistent Self and Inconsistent Self trials in Samson et al., (2010) be interpreted?	Prediction: If there were two avatars rather than one or zero on Self Inconsistent trials with discs on both walls, would performance be better or worse?	Prediction: If you remove the avatar from Self Inconsistent trials with discs on both walls, would performance be better or worse?
The Perspective-Taking Hypothesis	A perspective taking mechanism	It is an altercentric interference effect.	Worse	Better
The Cueing Hypothesis	An attentional mechanism	It is a cueing effect.	Better	Equal or Worse

Table 1b. Summary of the predictions generated by the two competing hypotheses for Experiments 1-3.

Experiment	SOA	Secondary Task	Prediction: When there are two avatars rather than one or zero, will performance be better or worse? 	
			Perspective taking mechanism	Attentional mechanism
1	800ms	[none]	Worse	Better
2	0	[none]	Worse	Better
3	800ms	auditory-tone-monitoring task	Worse	[none]

GENERAL DISCUSSION

The results of Experiment 1 revealed a significant effect of the number of avatars, with participants performing better on trials with two avatars than on trials with one or zero avatars. This is consistent with the operation of a spatial cueing mechanism. In contrast, they are difficult to account for by appealing to the operation of a perspective-taking mechanism. This confirms that the stimuli and the number verification procedure used in the dot perspective task can be used to trigger a spatial cueing mechanism. However, it would be hasty to draw any conclusions from this about the mechanisms underpinning the findings from the original dot perspective task. This is because Experiment 1 differed from the original dot perspective task in that it included an SOA of 800 ms.

For this reason, we conducted Experiment 2, in which there was no SOA. Here we observed the reverse pattern: participants performed better on trials with one avatar than on trials with two avatars. This is consistent with the hypothesis that the dot perspective task taps a perspective-taking mechanism distinct from spatial cueing. On the perspective-taking hypothesis, the presence of any avatar with an inconsistent perspective should interfere with the task rather than facilitate it. Thus, given that all trials involve discs on both walls, performance should be worse on trials in which there are two avatars (each with a perspective that is inconsistent with the perspective of the participant) than on trials with one avatar (with just one perspective that is inconsistent with that of the participant) or on trials with no avatar (in which there is no inconsistent perspective).

The findings from Experiment 1 and Experiment 2, taken together, suggest that there may be two separate mechanisms at work, with the perspective-taking mechanism predominating at earlier time points and the spatial cueing mechanism predominating at later time points. If this is correct, it may be possible to selectively intervene on the spatial cueing mechanism and prolong the effects of the perspective-taking mechanism. Since the spatial cueing mechanism depends upon executive function (Jonides, 1981; Müller & Rabbitt, 1989; Frischen et al., 2007), it may therefore be suppressed through the imposition of a demanding secondary task. By contrast, there is evidence that the perspective-taking mechanism operates automatically (Qureshi, Apperly & Samson; 2010), and would thus be robust under dual-task conditions. We therefore conducted Experiment 3 in which there was an SOA of 800 ms, as in Experiment 1, but in which participants were also asked to concurrently perform a secondary task. We predicted that the secondary task would interfere with the spatial cueing mechanism, and that the perspective-taking mechanism would therefore predominate on this version of the task, which should result in better performance with 0 or 1 avatar than with 2. The results confirm this prediction, with performance being worse with 2 avatars than with 1 or 0 (See Tables 1a and 1b for a summary of the different patterns of predictions generated by the perspective-taking hypothesis and the spatial cueing hypothesis).

Taken together, the results of Experiments 1, 2 and 3 suggest that a spatial cueing mechanism may be engaged in the dot perspective task if an SOA is involved, but that this is unlikely to explain the findings from standard versions of the paradigm that do not include an SOA. This is consistent with the findings of Marotta, Lupiánez, Martella, & Casagrande (2012), who systematically varied not only the locations of targets, but also the objects (i.e. rectangular figures) in which those targets appeared, and were thereby able to show that faces, unlike arrows, trigger a pure location-based cueing effect, whereas arrows, unlike faces, trigger a pure object-based cueing effect. They interpret this finding as evidence that faces and arrows engage qualitatively different (i.e. location-based versus object-based) orienting mechanisms.

In sum, our results strongly suggest that the perspective-taking mechanism engaged in the standard dot perspective task is distinct from spatial cueing. This finding vindicates the use of the dot perspective paradigm as a tool for investigating the nature and limitations of fast and efficient mindreading processes in neurotypical (Qureshi, Apperly, & Samson, 2010; Furlanetto et al., 2016) and autistic (Schwarzkopf et al., 2014) populations, and thus also as a means of testing theories of the cognitive architecture of mindreading (Butterfill & Apperly, 2013; Christensen & Michael, 2015, Westra, 2016). Further research may investigate how, if at all, this perspective-taking mechanism may be modulated by and/or integrated with attentional mechanisms.

Acknowledgements

John Michael was supported by a Marie Curie Intra European Fellowship (PIEF-GA-2012-331140). John Michael and Clément Letesson were supported by a Starting Grant from the European Research Council (n 679092, SENSE OF COMMITMENT). Jakob Hohwy was supported by The Australian Research Council grants FT100100322 and DP160102770 and by the Research School Bochum and the Center for Mind, Brain and Cognitive Evolution, Ruhr-University Bochum. We are grateful to Dana Samson for making the stimulus material available.

References

Bayliss, A. P., & Tipper, S. P. (2006). Predictive gaze cues and personality judgments should eye trust you? *Psychological Science*, 17(6), 514–520.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008a). Social attention and real world scenes: the roles of action, competition, and social content. *Quart. J. Exp. Psychol.*, *61*(7), 986–998.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008b). Gaze selection in complex social scenes. *Visual Cognition*, *16*(2/3), 341–355.

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, 5–13.

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*, 1-23.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in quantitative methods for psychology*, *1*(1), 42-45.

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.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*(1), 99.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*, 5, 490–495.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, *133*(4), 694.

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.* Vol 42(2), Feb 2016, 158-163.

Hamilton, A. F., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. Cognition, 113, 37–44.

Heyes, C. M. (2014). Submentalizing: I'm not really reading your mind. *Perspectives on Psychological Science*, 9, 131–143.

Jonides, J. Voluntary versus automatic control over the mind's eye's movement. In: Long, JB.; Baddeley, AD., editors. Attention and performance IX. Hillsdale, NJ: Erlbaum; 1981. p.187-203.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic bulletin & review*, *1*(4), 476-490.

Marotta, A., Lupiánez, J., Martella, D., & Casagrande, M. (2012). Eye gaze versus arrows as spatial cues: Two qualitatively different modes of attentional selection. Journal of Experimental Psychology: Human Perception and Performance, 38(2), 326–335.

Maylor, E. A. (1985). Facilitatory and inhibitory components of orienting in visual space. In: Eleventh International Symposium on Attention and Performance, Oregon, USA, 1-8 July 1984 pp. 189-204.

Müller HJ, Rabbitt PM. Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. Journal of Experimental Psychology: Human Perception and Performance. 1989; 15:315–330.

Peirce, JW (2007) PsychoPy - Psychophysics software in Python. *J Neurosci Methods*, 162 (1-2): 8-13.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25.

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, 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, 1255–1266.

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(3), 929–937.

Schneider, D., Slaughter, V.P., Bayliss, A.P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, 129, 410-417.

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.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. *Cognitive Theory*, 3, 200–239.