Title
The dot perspective task revisited: Evidence fordirectional effects

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
O'Grady, Cathleen
Scott-Phillips, Thom
Lavelle, Suilin
et al.

Publication Date
2017

Peer reviewed

The dot perspective task revisited: Evidence for directional effects

Cathleen O'Grady (C.J.O'Grady@sms.ed.ac.uk)[a], Thom Scott-Phillips[b,c], Suilin Lavelle[a], Kenny Smith[a]

[a] School of Philosophy, Psychology & Language Sciences, University of Edinburgh
[b] Dept. of Cognitive Science, Central European University, Budapest
[c] Dept. of Anthropology, Durham University, UK

Abstract

Humans are highly social creatures. Evidence from the dot perspective task suggests that humans automatically track the perspective of other individuals – a disposition that, if true, may help to facilitate social interaction. However, variants of the original dot perspective task suggest the alternative interpretation that the effect in the task is not due to perspective taking. Here, we present a new variant, using improved stimuli to address these issues. Our results replicate previous findings, across both animate and inanimate stimuli, and suggest that the effect is due to directional cueing rather than automatic perspective taking.

Keywords: perspective taking; dot perspective task; automaticity; theory of mind; mindreading

Introduction

The ability to reason about other individuals' mental states ("mindreading") is thought to be a central component of social cognition in humans (Corballis, 2011; Graziano, 2013; Tomasello, 2008, 2014). In order to explain the social abilities that are best accounted for by mindreading, it seems necessary that certain forms of mindreading are highly efficient (Apperly, 2011; Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). Evidence for efficient mindreading comes from various experimental paradigms (Freundlieb, Kovács, & Sebanz, 2016; Schneider, Slaughter, & Dux, 2017; Scott & Baillargeon, 2017), including the dot perspective task (DPT) (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010), which suggests that participants rapidly and automatically calculate the perspective of other agents.

However, the interpretation of these results is disputed. Different variants of the DPT (e.g. Cole, Atkinson, Le, & Smith, 2016; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014) have produced results that may be explained by a simple directional effect, in which attention is directed not exclusively by the gaze of an agent, but rather by any directional stimuli. If the task results are indeed attributable to directional cueing, it would undermine the use of this task as evidence for fast and automatic mindreading. We describe the different variants in the next section, before describing a new variant, using Lego figures, that may be used to address these issues, and the experimental results obtained using it.

Variants of the Dot Perspective Task

In the dot perspective task, participants observe scenes and answer a simple yes/no question based on the number of dots in the scene. The scenes that participants view feature an on-screen human avatar standing in a room. Arranged on walls around the room are various dots. In some scenes, the dots all appear in front of the avatar, making the avatar's perspective of the dots *consistent* with the participant's: e.g. if there are two dots on the front wall, the avatar and the participant both see two dots. In other scenes, some of the dots are behind the avatar, making the avatar's perspective *inconsistent* with the participant's: the avatar might see only one dot, while the participant can see two.

Participants are shown a digit (e.g. "2"), followed by one of these scenes, and asked to confirm whether the number of dots matches the pre-scene digit by answering "Yes" or "No." In three different experiments, Samson et al. (2010) found longer reaction times for inconsistent scenes compared to consistent scenes, which they interpreted as evidence for "altercentric interference": the participant had to suppress the avatar's perspective in order to answer the question of whether the digit matched their own perspective, resulting in a delayed response. This suggests that perspective taking, even for an on-screen avatar, is rapid and automatic.

In the first two of these three experiments, participants were asked to judge their own perspective on certain scenes (cued by the word "YOU" appearing before the digit), and the avatar's perspective on others (cued by the word "HE" or "SHE"). Because this may have caused participants to take the avatar's perspective in all scenes, Experiment 3 instructed participants to ignore the stimuli in the middle of the room and judge only their own perspective; the consistency effect persisted.

Santiesteban et al. (2014) argue that the effect of the avatar on reaction times was driven not by perspective taking of the avatar but rather by a directional effect: because the avatar faced one or the other side of the room, the participant's attention might be directed towards stimuli on that side. They repeated the experiment using avatar-sized arrows (rather than columns) as controls, finding a consistency effect for both avatars and arrows, both when perspective switched between trials (Experiment 1), and

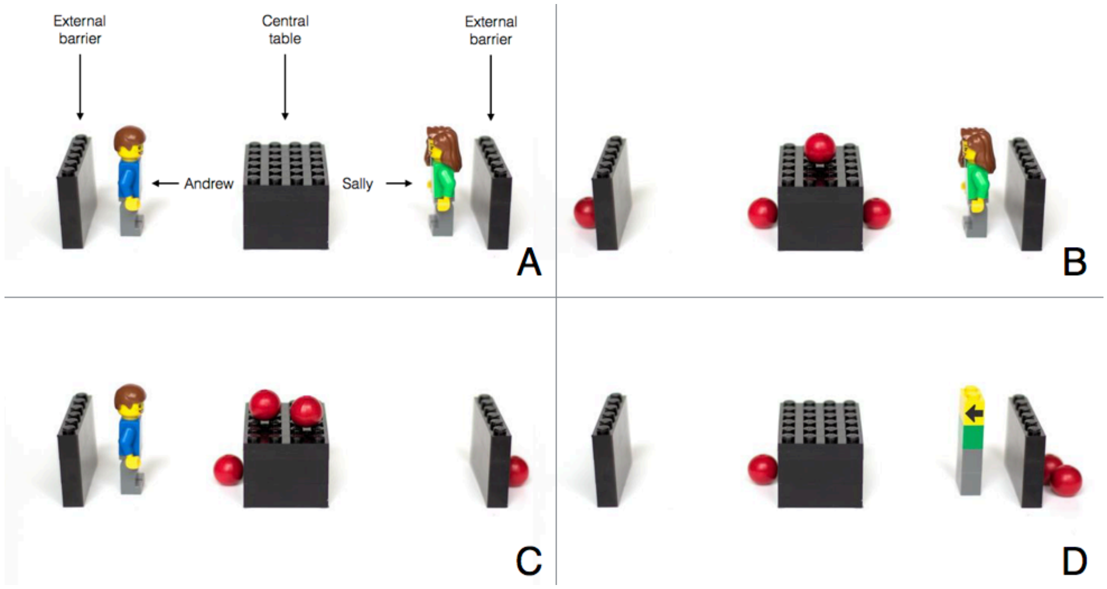


Figure 1: Example scenes.
A: The main components of each scene. B: Example scene with Sally.
C: Example scene with Andrew. D: Example scene with arrow.

when participants were instructed to ignore the stimuli in the centre of the room and judge only their own perspective (Experiment 2). However, because both kinds of stimulus were presented to all participants (i.e. the avatar vs arrow manipulation was within-subjects), it is possible that participants were transferring the "perspective taking" of the avatar over to the arrow.

Cole et al. (2016) note a further problem with this experiment: although arrows and avatars produce a similar effect on reaction times, these effects may in fact be driven by different processes—perspective taking in the case of the avatar, and directional cueing in the case of the arrows. Indeed, Marotta, Lupiáñez, Martella, & Casagrande (2012) find that, while eye gaze cues participants to a specific location, an arrow provides a more general cue. This suggests that different processes are involved in following the directional cue of an arrow and an avatar.

As an alternative control, Cole et al. (2016) use a set of stimuli that includes a barrier in front of the avatar, as is used in mentalising experiments in non-human animals (Hare, Call, & Tomasello, 2001). When the barrier "window" is open, allowing the avatar to "see" the dots, they find the expected consistency effect; but they also find the effect when the barrier window is closed, suggesting that the effect is driven simply by the directional effect of the avatar, rather than by mental state attribution.

However, the stimuli used in this experiment do not make it perfectly clear whether or not the barrier is transparent, and the depth and angle of the barrier placement within the room could be ambiguous. Further, the temporary nature of the barriers may create a problem: given that the participant likely assumes that the avatar is a single agent, it is possible that participants infer the agent's knowledge of what is on the other side of the barrier on the basis that they can sometimes see what is there, and may have done so before the barrier window closed.

Cole et al. (2016) do attempt to deal with these problems. The open or closed barriers were shown in different blocks of trials, and at the beginning of each block, participants were explicitly told whether or not the avatar could see the wall that was blocked by the barrier. However, given the visual ambiguity of the stimuli, it is possible that this kind of explicit knowledge is not taken into account in fast processing, when at a glance the image might be interpretable in different ways.

Using different stimuli and a modified experiment design, we conducted a conceptual replication of Experiment 3 in Samson et al. (2010) and Experiment 2 in Santiesteban et al. (2014). Although our experiment was designed to address details of Yes vs. No responses and arrow vs. avatar stimuli, the design also allowed us to explore the effect of barriers as in Cole et al. (2016), while addressing the problems of ambiguity. Unlike Samson et al. (2010) but following Santiesteban et al. (2014) we used arrows as a directional control for avatars; unlike Santiesteban et al. (2014), we manipulated avatars vs arrows in a between-participants design, rather than within-participants. Our stimuli did not have the same temporal and physical ambiguity as the images used by Cole et al. (2016) (see Figure 1). We used photographs of Lego figures in scenes with unambiguous depth in the third dimension, and solid black barriers were used, preventing any ambiguity in whether or not Lego figures were able to see through them.

A variety of hiding places allowed balls (our equivalent of dots/discs) to be hidden from view of the Lego figures, even when placed in front of them. This allowed us to test the claim that the altercentric effect could be explained by the general directionality of the avatars, rather than perspective taking.

In addition, the use of arrows as control stimuli should indicate whether, as in Marotta et al. (2012), the arrows have a more general directional effect than the avatars. If this were the case, one would expect arrows to cause a reaction time delay only when there are balls placed in the opposite direction to that indicated by the arrow; and the more specific perspective attributed to avatars to cause a reaction time delay in all cases where there are balls not in its field of view (regardless of whether they are hidden behind a barrier in front of, or behind, the avatar).

Method

Participants

Sixty participants were recruited through the University of Edinburgh Student and Graduate Employment Service. They were compensated £4 for their participation, which lasted approximately 20 minutes. Thirty participants viewed stimuli with the Lego figures, and thirty viewed control stimuli showing columns with arrows on them. One further participant was excluded from analysis because a post-experiment questionnaire indicated that they had successfully guessed the purpose of the experiment.

Materials

Participants observed scenes consisting of photographs (Figure 1) of Lego figures (dubbed "Sally" and "Andrew" for ease of reference), a series of barriers created by Lego bricks, and red beads that, at Lego scale, had the appearance of red balls. Control stimuli consisted of Lego columns with the same colours and proportions as Sally and Andrew, with a black arrow on the yellow block, pointing in the same direction as a figure's direction of facing. Each scene featured either Sally or Andrew (each figure could appear on either side of the screen), and between 0 and 4 balls (with a maximum of two balls in any given location).

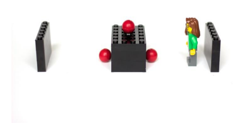
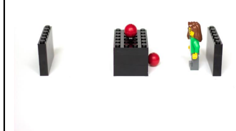
Procedure

On each trial, participants were presented with a fixation cross for 750 ms, followed by a digit between 0 and 4 (displayed for 750 ms), followed by a Lego scene, with the words "Yes" and "No" in the bottom corners of the screen (Yes-side was counterbalanced across participants but remained consistent across trials for a given participant). Participants were instructed to judge whether the picture had the same number of balls as the digit they had been shown – with no other comment given about the other elements of the scene – using a two-button button box, pressing the Yes-side button for yes and the No-side button for no. Scenes timed out within 2000 ms if no response was given, and moved on to the following trial.

After completing 12 practice trials with correct/incorrect feedback on responses, participants completed 324 trials (36 filler trials with zero balls, and 288 test trials), in random order, divided into four blocks, with a self-paced break between blocks. These 288 trials balanced three different variables: the number of balls in a scene, the consistency between the Lego avatar's and participant's perspective, and the match between the digit shown and the number of balls in the scene.

There were 72 trials for each number of balls; that is, 72 scenes with one ball, 72 with two balls, and so on. Half of the trials were consistent in perspective: that is, the figure/arrow could "see" (i.e. had unobstructed line of sight to) the same number of balls that the participant could see. The other half were inconsistent, with balls hidden from the figure or arrow by either the central, table-like barrier or the external wall-like barriers, introducing an inconsistency between the participant's perspective and that of the avatar/arrow. The match between the digit shown and the on-screen perspective was balanced (Table 1); the results of analysis of this variable will be reported in a future paper.

Table 1: Match between digit and perspective

	Inconsistent			**Consistent**	
	 Avatar sees 2; participant sees 3			 Avatar sees 2; participant sees 2	
Digit shown	3	2	4	2	3
Correct answer	Yes	No	No	Yes	No
Condition	Yes	No-Other	No-None	Yes	No-None

Post-experiment questionnaires were used to assess whether participants' intuitions about the figures' lines of sight matched those of the experimenters. Pictures showing a variety of scenes with balls in different positions were displayed, and participants were asked to note how many balls the Lego figure could see (regardless of whether they had just completed the avatar or arrow condition of the experiment). All responses to these questionnaires indicated that participants did not expect the Lego figures to be able to see balls hidden by either the central or external barriers, but did expect them to see balls either on the table or at their feet.

The experiment was implemented using PsychoPy (Peirce, 2010).

Results

We used lme4 (Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2016) to perform a series of linear mixed effects analyses on reaction time (RT); RT was our only dependent variable given the lack of effect on error rate found in our own data and in previous studies. We removed training trials, trials with zero balls on screen, timed-out trials (0.69%, n = 119), and trials where participants made an incorrect response (3.12%, n = 533). As per Whelan (2008), trials in which the response RT was lower than 100 ms were also removed, on the assumption that these trials could not be genuine responses to the stimuli (0.01%, n = 2). No trimming was conducted on higher reaction times, given the imposed cut-off of 2000 ms on all trials. Visual inspection of the reaction time data revealed an obvious deviation from the normal distribution, necessitating a log transform of the data (Baayen & Milin, 2010).

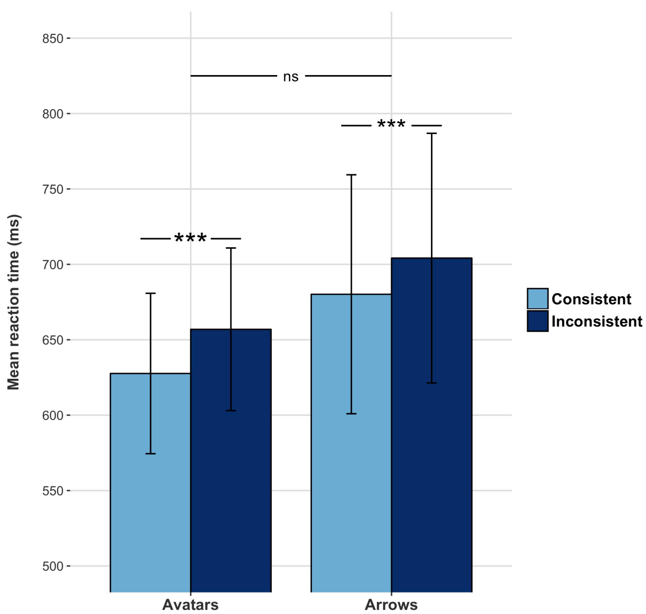


Figure 2: Mean RTs showing a significant effect of Consistency (error bars show 95% CI) and no Stimulus x Consistency interaction. The effect of Stimulus is not significant (note that Stimulus, unlike Consistency, is manipulated between-subjects). Y-axis limited for easier comparison with earlier experiments.

Replication

We first conducted an analysis of the relationship between RT, Consistency and Stimulus. As fixed effects, we entered Consistency and Stimulus (with interaction term) into the model. As random effects, we included random intercepts for participants and images, as well as by-participant and by-image random slopes for the effects of Consistency and Stimulus (without interaction term, to facilitate model convergence).

Following Samson et al. (2010), the model showed a significant effect of Consistency (Figure 2), with consistent trials faster than inconsistent trials ($\beta = 0.047$[1], SE = 0.008, $p < .001$). Contra Samson et al. (2010) but consistent with Santiesteban et al. (2014), there was no effect of Stimulus ($\beta = -0.065$, SE = 0.049, $p = .187$) and no Stimulus x Consistency interaction ($\beta = 0.012$, SE = 0.01, $p = .220$). This suggests that an inconsistency in perspective resulted in slower responses, but that this was true for both avatars and arrows. Our between-subjects manipulation of avatars vs arrows ensures that, unlike for Santiesteban et al. (2014), this cannot be explained as a consequence of transfer from avatars to arrows: our participants seeing the arrow stimuli had not seen Lego figures in those positions.

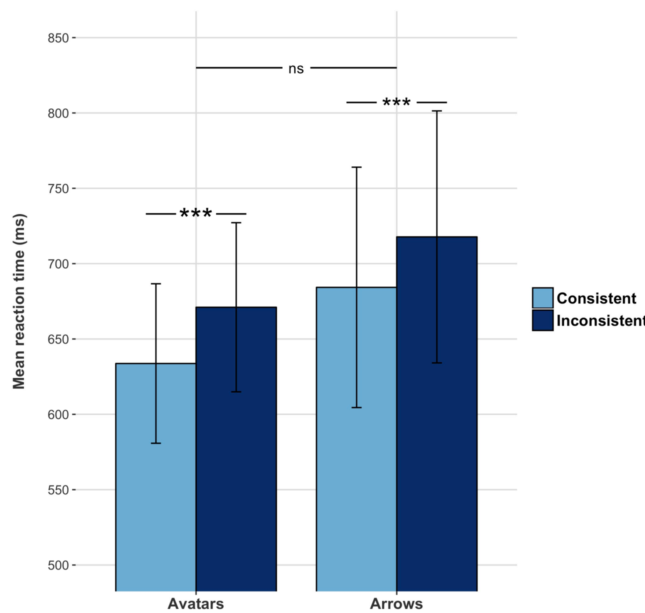


Figure 3: Mean RTs showing a significant effect of Directional Consistency.

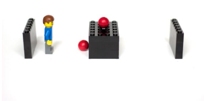
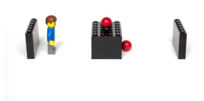
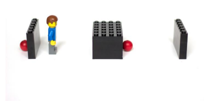
Directional Consistency

Our experimental setup also allowed us to test the hypothesis that the delay is caused not by processing of the altercentric perspective, but rather by preferential attention to objects in the direction of facing/arrow pointing. We predicted, based on Marotta et al. (2012), that the delay would appear only on those trials where balls in front of the avatar are within the avatar's actual field of view, and not on trials where there are balls in front of the avatar, but hidden by obstacles, consistent with the explanation of altercentric interference. We similarly predicted that when the stimulus was an arrow instead of an avatar, the delay would occur on all trials where there are balls within the arrow's field of reference, regardless of barriers between the balls and the arrow.

[1] Slope estimates represent log transformed RT data.

To test these predictions, we re-coded the data to classify all trials with balls in front of the avatar/arrow as direction-consistent, and only those trials where a ball appeared behind the avatar/arrow as direction-inconsistent. We then modelled the relationship between this Directional Consistency, Stimulus, and RT (Figure 3). Contrary to our predictions, the results showed that directional-inconsistent trials were slower than directional-consistent trials (β = 0.047, SE = 0.01, $p < .001$), with no significant effect of Stimulus (β = -0.058, SE = 0.049, p = .24) and no significant interaction (β = 0.004, SE = 0.011, p = .73). This suggests that the consistency effect may be driven by preferential attention to objects within a directional figure's direction of facing/pointing, regardless of the animacy of that figure.

Table 2: Congruence

		
Line of Sight consistent Directional consistent	Line of Sight inconsistent Directional consistent	Line of Sight inconsistent Directional inconsistent

However, a further model with both Consistency and Directional Consistency as fixed effects found a significant effect for both variables (β = 0.032, SE = 0.011, p = .005 and β = 0.04, SE = 0.008, $p < .001$ respectively). In order to explore this, the data was recoded to classify each scene as consistent and/or inconsistent for both definitions of consistency (Table 2). That is, each scene could be (a) line of sight consistent + directional consistent (balls within the avatar's direction of facing and actual field of view); (b) line of sight inconsistent + directional consistent (balls within the avatar's direction of facing, but hidden from the avatar's field of view); or (c) line of sight inconsistent + directional inconsistent (inconsistent based on both direction of facing and field of view).

A model with this variable (Congruence) and Stimulus as fixed effects (with interaction term) found that line of sight consistent + directional consistent trials were faster than both line of sight inconsistent + directional consistent (β = 0.036, SE = 0.012, p = .003) and line of sight inconsistent + directional inconsistent (β = 0.077, SE = 0.012, $p < .001$) trials (Figure 4); a re-levelled model showed that line of sight inconsistent + directional consistent was significantly faster than line of sight inconsistent + directional inconsistent (β = 0.041, SE = 0.01, $p < .001$). There was no effect of Stimulus or Stimulus x Congruence interaction.

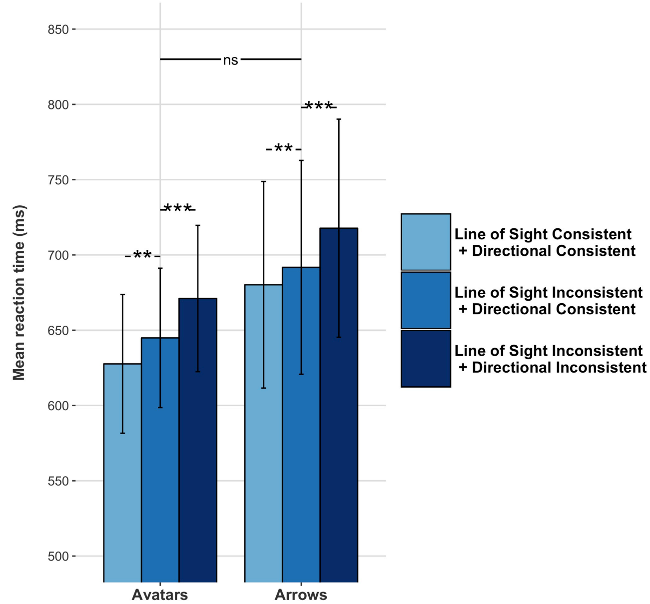


Figure 4: Mean RTs showing a significant effect of Congruence: scenes with consistent perspectives and unobstructed balls are faster than scenes with inconsistent perspectives created by barriers in front of the stimuli; which in turn are faster than scenes with balls hidden both in front of, and behind, the stimuli.

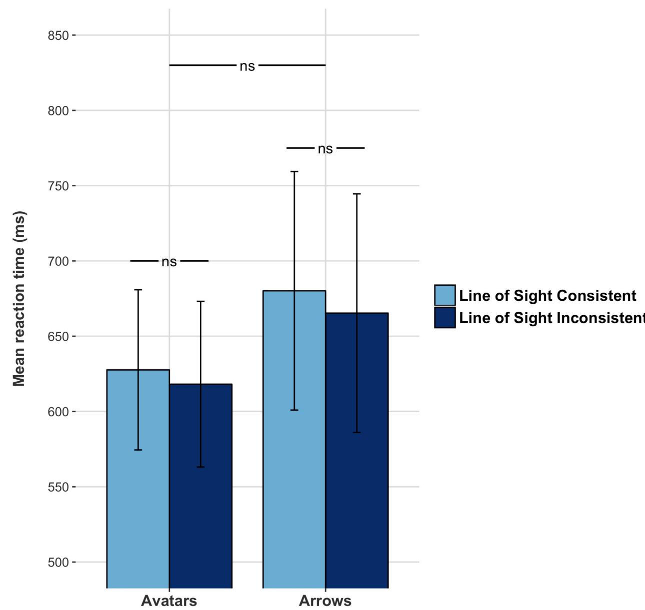


Figure 5: Without the confound of peripheral balls, there is no effect of Line of Sight consistency on RT.

These results would suggest a role for both Consistency and Directional Consistency in affecting reaction times, but there is an important confound: within directionally consistent scenes, line of sight consistent scenes can only have balls in the centre of the screen, while line of sight inconsistent scenes may have balls on the periphery of the screen (the same confound does not apply across directional consistent vs. directional inconsistent scenes, which may both have peripheral balls). Once data is restricted to only those scenes with balls in the centre of the scene (all of which are directionally consistent), there is no longer an effect of line of sight consistency ($\beta = -0.008$, $SE = 0.013$, $p = .525$, Figure 5). This suggests that the consistency effect may be accounted for by the directional hypothesis.

Conclusion

These results replicate the headline result of Samson et al. (2010) by finding a robust effect of Consistency on reaction times. However, they also replicate the results of Santiesteban et al. (2014) by finding that the Consistency effect appears with inanimate but directional stimuli, even when those stimuli appear in a between-participants design. Additionally, the analysis of Directional Consistency suggests that the effect is driven by a directional cueing effect. These findings cast uncertainty on interpretation of DPT data as evidence for automatic mindreading.

Heyes (2014) argues that evidence for a directional explanation, such as the data we have presented here, is evidence against a mentalising explanation. This dichotomy may be too sharp: directionality and perspective taking are not unrelated. Taking another individual's perspective must entail first following the direction of their gaze; or, in other words, directional effects may be a necessary pre-condition of perspective taking. Our results (and other results too) suggest that directional effects – which are a relevant input into any possible fast and efficient perspective taking – are indeed automatic and efficient. They just do not seem to necessarily lead to perspective taking.

If this speculation is correct, it may be important to distinguish automatic cognitive processes (i.e. those that are mandatory upon the perception of relevant inputs) and spontaneous ones (i.e. those that occur quickly and efficiently as and when needs arise). Our results – and results from other experimental paradigms (e.g. Freundlieb et al., 2016; Schneider et al., 2017) – are consistent with the interpretation that perspective taking is spontaneous but not automatic. Future experimental research could test this possibility directly.

Acknowledgements

We would like to thank our colleagues Simon Kirby and Jennifer Culbertson, who provided insight and expertise that greatly assisted this research.

References

Apperly, I. A. (2011). *Mindreaders*. 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), 953–70.

Baayen, R. H., & Milin, P. (2010). Analyzing Reaction Times. *International Journal of Psychological Research*, *3*(2), 12–28.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Butterfill, S. A., & Apperly, I. A. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language*, *28*(5), 606–637.

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

Corballis, M. C. (2011). *The Recursive Mind: The Origins of Human Language, Thought, and Civilization*. Princeton, NJ: Princeton University Press.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective? *Journal of Experimental Psychology. Human Perception and Performance*, *42*(3), 401–12.

Graziano, M. S. A. (2013). *Consciousness and the Social Brain*. Oxford University Press.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, *61*(1), 139–151.

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-30.

Marotta, A., Lupiáñez, 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.

Peirce, J. (2010). PsychoPy - Psychology software for Python. *Integration The Vlsi Journal*.

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), 929–37.

Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition*, *162*, 27–31.

Scott, R. M., & Baillargeon, R. (2017). Early False-Belief Understanding. *Trends in Cognitive Sciences*, 1–13.

Tomasello, M. (2008). *Origins of Human Communication*. *Communication*. MIT Press.

Tomasello, M. (2014). A Natural History of Human Thinking. *2014*, *2*(1), 1–5.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*, 475–482.