Where was I? Taking alternative visual perspectives can make us (briefly) misplace our own.

Manuscript accepted for publication in the Quarterly Journal of Experimental Psychology

Samuel, Steven[1,3]

Legg, Edward. W.[1]

Manchester, Callum[1]

Lurz, Robert[2]

Clayton, Nicola S.[1]

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

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

[3] Department of Psychology, University of Essex, U.K.

Word count: 6692 (excluding title page, abstract, tables, refs, acknowledgments and captions)

Key words: Theory of Mind, Spatial cognition, Perspective-Taking, Embodied cognition

Address for correspondence: Steven Samuel. University of Essex, Department of Psychology, Wivenhoe Park, C04 3SQ. Tel: 01206 874918. Email: ssamuea@essex.ac.uk

Abstract

How do we imagine what the world looks like from another visual perspective? The two most common proposals—embodiment and array rotation—imply that we must briefly imagine either movement of the self (embodiment) or movement of the scene (array rotation). What is not clear is what this process might mean for our real, egocentric perspective of the world. We present a novel task in which participants had to locate a target from an alternative perspective but make a manual response consistent with their own. We found that when errors occurred they were usually manual responses that would have been correct from the computed alternative perspective. This was the case both when participants were instructed to find the target from another perspective and when they were asked to imagine the scene itself rotated. We interpret this as direct evidence that perspective-taking leads to the brief adoption of computed perspective—a new imagined relationship between ourselves and the scene—to the detriment of our own, egocentric point of view.

Whether understanding that something on our right is on another's left, or that an upside-down 6 looks like a 9, we need to understand that visual perceptions of the world vary according to one's physical location. These two examples both concern what has historically been defined as Level 2 perspective-taking (Level 2 VPT), which concerns judging *how* something appears (Flavell, Everett, Croft, & Flavell, 1981; Masangkay et al., 1974; Sodian, Thoermer, & Metz, 2007). This is different from Level 1 visual perspective-taking (Level 1 VPT), which involves judging a binary state of *whether* or not an object is seen. Level 2 VPT is typically considered to tap into a richer understanding of another perspective. For example, one might simply draw imaginary lines in space to determine that two people sitting either side of a table in a restaurant can both see the table number from where they are (a Level 1 VPT task). Such a process is essentially a geometric one, and does not require an actual perspective to be taken (Michelon & Zacks, 2006). Level 2 VPT on the other hand is not solvable by line-of-sight processes alone, and is more likely to lead to the generation of a representation of another agent's perspective (Lurz, 2009). Imagining others' visual perspectives is considered a key component of our ability to understand others' mental states (Ferguson, Apperly, & Cane, 2017), often called our 'theory of mind' (Premack & Woodruff, 1978). It is for this reason that we choose here to investigate Level 2 VPT.

To date, there is a growing body of evidence that we engage our motor representations to imaginatively 'embody' a new physical location in Level 2 VPT (Kessler & Rutherford, 2010; Kessler & Thomson, 2010; Michelon & Zacks, 2006; Surtees, Apperly, & Samson, 2013a, 2013b). Much of this evidence comes from tasks that have found that visual perspective-taking is more difficult when the participant's body is manipulated to be incongruent with the movement required to 'reach' the imagined location. For instance, the time needed to judge what is on another person's right or left increases with the angle of disparity between our location and the to-be-embodied one (Erle & Topolinski, 2017; Surtees et al., 2013a, 2013b; Yu & Zacks, 2017), and physically rotating the body or fixing its posture in a manner incongruent with the shortest path to an imagined location also impairs performance (Deroualle, Borel, Devèze, & Lopez, 2015; Kessler & Thomson, 2010; Surtees et al., 2013b; Yu & Zacks, 2017) as does imagining a body part rotated in a manner inconsistent with plausible body articulation (Parsons, 1987).

A second strategy that we can adopt concerns the imagined movement not of ourselves but of the scene. Given that the output of array rotation is the same as embodied perspective-taking, a common means of investigating differences between the two strategies has been to word the instructions given to participants such that one or other method is privileged. Such experiments have typically found that adults are usually faster and more accurate at making Level 2 VPT judgments when they are asked to imagine themselves in a different location than when imagining the scene itself rotated (Amorim & Stucchi, 1997; May, 2004; Presson, 1982; Wraga, Creem, & Proffitt, 2000; Zacks & Tversky, 2005), suggesting embodied visual perspective-taking might be the preferred or default mode of Level 2 VPT.

Both the embodiment and array-rotation strategies imply the calculation of a new relationship between the self and the scene, either brought about through imagined self-movement (embodiment) or imagined scene movement (array rotation). An interesting question that arises from both of these possibilities is whether we might erroneously act according to this imagined relationship as a result of perspective-taking, even when we are required to make a manual response consistent with our actual, egocentric frame of reference. This would suggest that in order to see the world differently we briefly adopt that perspective to the detriment of our own.

We tested this possibility using a novel Level 2 VPT task. In Experiment 1, we asked participants to locate a target from an avatar's perspective but make a manual response consistent with the target's location as the participant herself saw it. This form of instruction favours an embodiment process rather than array rotation, as it promotes the alignment of one's self with the agent's point of view. Participants were shown an avatar at one of the four edges of a 2 x 2 grid, seen from above (see Figure 1). Each grid contained two digits with the target digit always upright from the avatar's perspective. On each trial, participants needed to find the digit the avatar named (a '4' or a '6') from the avatar's perspective, but to press a key on a 2 x 2 keypad to indicate the target's location from their *own* perspective. Given that the participant had four response options, this design allowed us to distinguish between different types of errors. One such error would involve the participant pressing the button that corresponded to the location of the distractor digit from their own perspective (Incorrect – Distractor). Another error could be the selection of one of the two empty squares from their own perspective.

What we were most interested in, however, was a specific type of empty-square selection that might occur when the avatar was located at a 90-degree angle to the participant. For example, if the avatar was on the right of the grid, a target in the *top* left corner of the grid from the avatar's point of view would require a *bottom* left motor response from the participant. This conflict between the avatar's perspective and the participant's created an opportunity for 'altercentric' errors, whereby participants fail to disengage from the avatar's perspective prior to making their response and erroneously press the top left key—the response that would be correct from the avatar's perspective. Such errors, which we termed Incorrect – Altercentric responses, would provide evidence that participants had erroneously made a manual response consistent with the computed location of the target from the avatar's perspective rather than their own. As a baseline measure, we could compare Incorrect – Altercentric errors to the number of Incorrect – Control responses, which correspond to the *other* empty square from the participant's perspective (the distractor digit from the avatar's). Importantly, both Incorrect - Altercentric and Incorrect – Control responses correspond to empty squares from the participant's perspective but only one is correct from the avatar's perspective. We chose left- and right-perspective trials for this particular analysis because on shared perspective trials we cannot know which perspective participants used (their own or the avatar's), and on opposite-perspective trials the distractor would resemble the target, making their selection more likely. Additionally, we chose to have the two digits in diagonally opposite squares so that we could discriminate between Incorrect – Distractor and Incorrect – Altercentric responses, which would otherwise culminate in the same button press on left- and right-perspective trials. We hypothesised that if participants compute the avatar's perspective to the (brief) detriment of their own, then this should be evident through a higher frequency of Incorrect - Altercentric responses than Incorrect – Control responses on left- and right-perspective trials.

On *Related* condition trials (50%) the second, distractor digit was the target digit inverted (e.g. a ‘6’ coupled with a ‘9’ or a ‘4’ with an upside-down 4) and on *Unrelated* condition trials it was a different number (one of a ‘6’ and a ‘4’). This distinction allowed us to investigate whether altercentric errors were more likely to occur when the disambiguation of targets was more demanding and the need to take a perspective more crucial; 6 and 9 are physically identical and require the correct perspective to be adopted to be disambiguated, whereas for 6/4 combinations the target could potentially be distinguished from the distractor by a consideration of their forms alone.

Experiment 1

Method

Participants

An a priori power analysis using G*Power found that 28 participants were required for an 80% chance of detecting a medium effect size (*dz* = .5), based on a one-tailed Wilcoxon matched-pairs test of the alternative hypothesis (H_1) that the median of the difference between the frequency of Incorrect - Altercentric errors and Incorrect - Control trial errors would exceed zero (owing to more of the former). We set a medium effect size because previous research has found either approximately medium or large effects of angular disparity or body movement on perspective-taking (e.g. Deroualle et al., 2015; Erle & Topolinski, 2017). Given the expectation that some data would be lost, a total of 32 participants participated in Experiment 1. One participant’s data were removed for evidence of their having misunderstood the instructions (always taking one perspective). Final sample size was thus 31 (9 male, 22 female, 30 right-handed, Mean age = 24, range = 19-40). All participants provided informed consent and were compensated financially for their time. The study was approved by the Cambridge Psychology Research Ethics Committee (PRE.2015.085).

Materials and procedure

Participants sat at one edge of a computer screen laid flat on a table. An overlay on the screen created a square white frame around a 720 x 720 pixel area on which the scenes were displayed. Each trial began with an instruction (1000 ms) ostensibly provided by the avatar and heard via headphones, which was always either ‘six’ or ‘four’ spoken in a female voice (the avatar was described as female). The screen was blank during the instruction and for 250 ms afterwards. An empty grid then appeared (100 ms duration), followed immediately by the two digits and the avatar. Participants were instructed to respond as quickly as possible. The avatar and the two digits remained on screen until the trial terminated. The trial terminated when the participant responded or after 3500 ms had elapsed (a timeout). There was then 2000 ms of blank screen prior to the next trial.

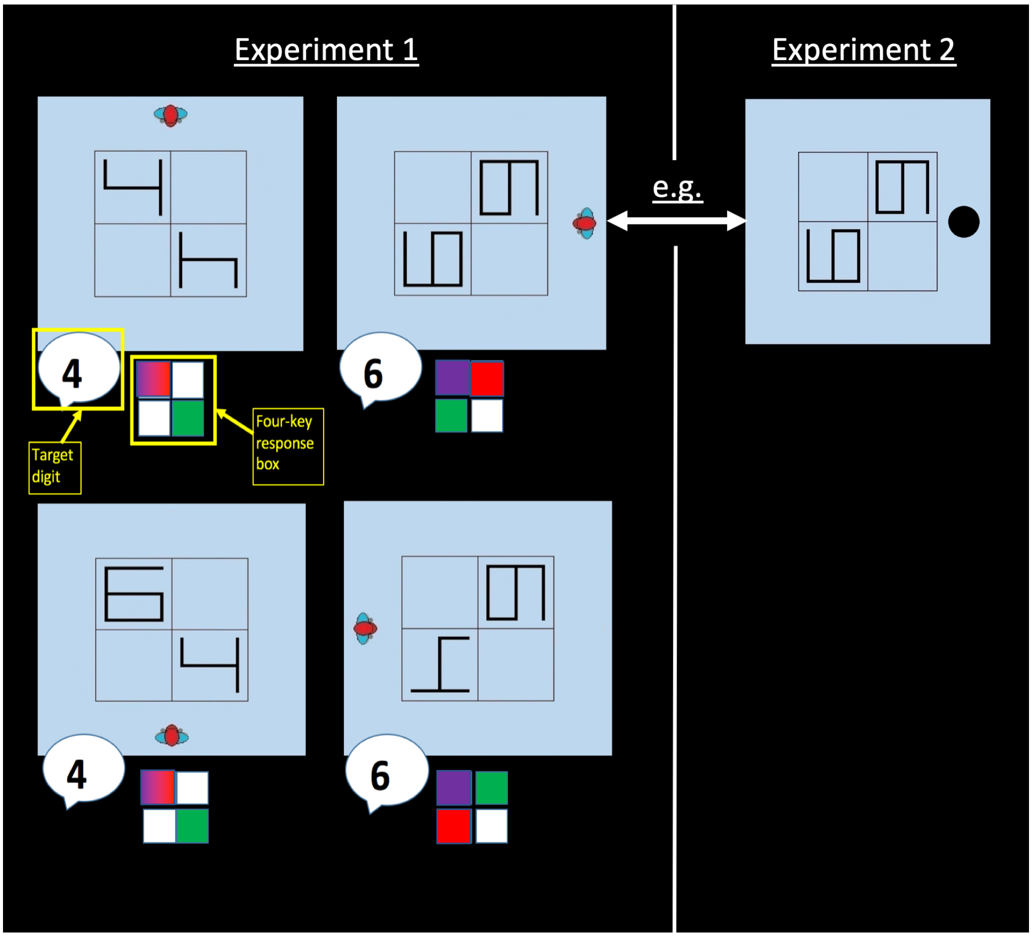

Fig 1. Examples from Experiment 1 (left), showing (clockwise) one opposite-, right-, left-, and shared-perspective trial. The top two examples come from the Related condition, the bottom two from the Unrelated condition. In the task, participants were instructed to find a target digit from the avatar's perspective but to press the key that corresponded to the target's location from their own *perspective. Participants used a physical response box and heard instructions via headphones. For illustration, correct responses are highlighted in green, Incorrect – Distractor responses in red, Incorrect – Altercentric responses in purple, and Incorrect – Control in white. Note that Incorrect – Distractor responses cannot be distinguished from Incorrect – Altercentric responses on opposite-perspective trial (hence the dual colour of the key). In Experiment 2 (example in right panel), participants saw a black dot instead of an avatar.*

Prior to the experimental trials described, participants first performed a brief (12-trial) practice session in which the aim was merely to press the button that corresponded to the location of a single stimulus (a '+' sign, no distractor) in one of the squares in the 2 x 2 grid. Instructions were in written form, and an experimenter was available to answer any questions regarding the procedure. This practice session gave participants experience pressing a button according to their own perspective of the grid (e.g., they pressed the top-right button when the '+' appeared in the top right square of the grid). The experimenter presided over the practice session and reminded participants that they should use only the forefinger of their dominant hand to respond, and that that finger should rest in the middle of the four buttons when not responding. No avatar was present in the practice phase, and no instructions were given to participants to take any perspective other than their own. Following the practice phase, participants received a second set of written instructions about the main task. All participants were required to show the experimenter which button on the response box they would press for three example scenes printed in the instructions, two horizontal perspectives (left and right), and one from the participant's own perspective (i.e. the scene depicted the avatar on the same side of the grid as the participant). The experimenter checked that participants responded according to the instructions. If the participant did not display the correct choices, further examples and clarification were provided until the participant did.

A total of 64 experimental trials were presented in a random order. To ensure participants realized the importance of the avatar's perspective and to increase the difficulty of the task more generally, on 16 trials the avatar appeared directly opposite the participant and hence the distractor represented a perfect match for the target from the participant's perspective; and on 16 trials the avatar shared the participant's perspective and hence the answer from the participant's own perspective was in fact the correct one. The crucial test of our hypothesis, however, concerned those trials in which the avatar was at a 90-degree angle to the participant, namely on left-perspective trials (16) and right-perspective trials (16 trials), rather than those trials from the opposite perspective. This is because it is not possible on opposite-perspective trials to distinguish between an Incorrect - Altercentric response and an Incorrect – Distractor response, as both result in the same response. The

64 experimental trials were equally divided between *Related*/*Unrelated* trials, 'four'/'six' targets, target and competitor location, and avatar position. This meant that of the 16 trials from the left perspective (for example), four trials had a target 6 with a distractor 6 (Related condition), once with the target in the top left, once in the top right, once in the bottom left, and once in the bottom right. Target location was not a factor of interest in the analysis but was included only to ensure a balanced number of correct responses across each of the four response buttons.

Results

Types of errors. Accuracy across the task as a whole was very high (95%, SD = 6%). Figure 2A displays the number and type of errors made. The crucial analyses concerned those trials where the avatar appeared on either the left or right (N = 992). On these trials, there were a total of 34 Incorrect - Altercentric errors, made by a total of 13 participants[1] (42%). Critically, the other empty square (i.e. Incorrect – Control) was *never* selected on such trials. A Wilcoxon matched-pair signed-rank test found that the median of differences (Incorrect - Altercentric minus Incorrect - Control) was greater than zero, W(31) = 0, Z = 3.210, $p = .001$, $r = .408$ (note all tests that do not form a priori power analyses are always two-tailed). Of the Incorrect - Altercentric responses, 15 (44%) occurred in the Related condition and 19 in the Unrelated condition (56%), a difference that was not significant, W(32) = 31, Z = 0.639, $p = .523$, $r = .080$. An additional 11 errors were timeouts, and 17 involved Incorrect – Distractor responses. Thus there were more Incorrect - Altercentric responses, though only numerically, W(31) = 65.5, Z = 0.528, $p = .598$, $r = .067$, than the combined total of all other error types on left- and right-perspective trials.

[1] Specifically, four participants made one Incorrect-Altercentric response, five made two each, one made three, two made five and one made seven.

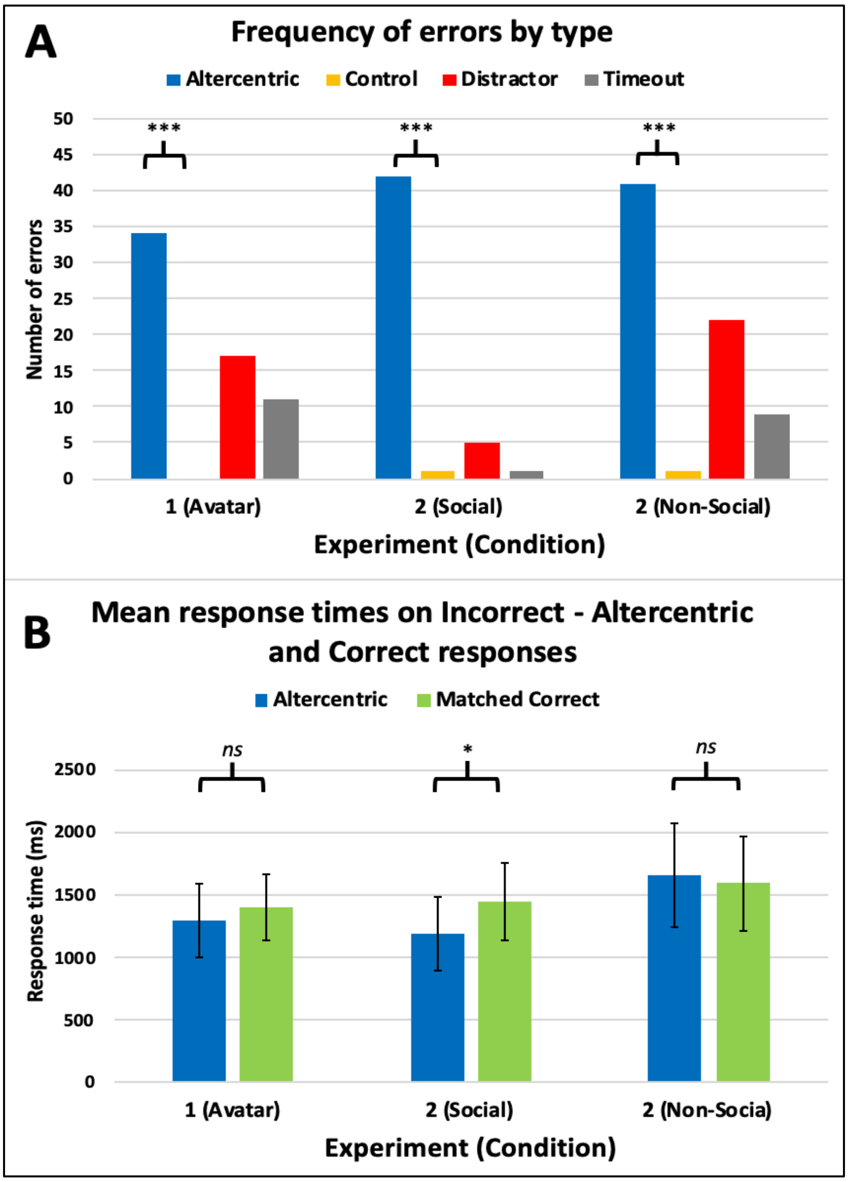

Figure 2. *(A) Number and types of errors by Experiment on left- and right-perspective trials. Additionally, in Experiment 1 there were six Incorrect – Distractor responses from the Shared perspective and 25 from the Opposite perspective, one empty square selection from the Shared and five from the Opposite, and two timeouts from the Shared and one from the Opposite. In Experiment 2 these figures were 7, 19, 3, 13, 0 and 6 (Social) and 2, 19, 2, 5, 1 and 15 (Non-Social) respectively; (B) RTs on Incorrect – Altercentric and Correct responses (matched to the former for Perspective, Target, and Relatedness), with 95% confidence intervals.*

Other results. We analysed RTs (initially non-normally distributed, so log transformed) by means of a 4 x 2 x 2 fully within-subjects ANOVA (Perspective: Shared vs. Left vs. Right vs. Opposite; Target: 4 vs. 6; Distractor: Related vs. Unrelated). The assumption of sphericity was not met in relation to the factor Perspective according to Mauchly's test, $\chi^2 = 13.370$, $p = .002$, ε (Greenhouse-Geisser) = .762, and Greenhouse-Geisser corrections are reported as a result.

Condition means can be found in Figure 3 (top panel). This revealed main effects of Perspective, F (2.287, 68.613) = 46.652, $p < .001$, $\eta_p^2 = .609$, $BF_{10} = 3.662e+24$ (all BF values based on a null model with all other factors and interactions included). Participants were fastest on shared-perspective trials (M = 937ms) followed by left- (M = 1097ms), right- (M = 1177ms), and then opposite-perspective trials (M = 1212ms). Pairwise comparisons using the Bonferroni correction found all contrasts significant (adjusted ps < .02) except right vs. opposite $p = .695$. There was a main effect of Distractor, F (1, 30) = 156.147, $p < .001$, $\eta_p^2 = .839$, $BF_{10} = 2.735e+75$. Participants were on average 436ms slower when the distractor was related to the target than when it was a different figure. Significant two-way interactions between Perspective and Target, F(3, 90) = 4.592, $p = .005$, $\eta_p^2 = .133$, $BF_{10} = 0.581$, and Perspective and Distractor, F (3, 90) = 6.088, $p = .001$, $\eta_p^2 = .169$, $BF_{10} = 8.128$, were subsumed within a significant three-way interaction, F (3, 90) = 4.928, $p = .003$, $\eta_p^2 = .141$, $BF_{10} = 1.454$. This pattern appeared to be driven by performance on opposite-perspective trials, where related trials and trials where participants had to select a 6 which looked like a 9 were particularly hard relative to unrelated trials and selecting a 4 that was upside-down. This latter effect was likely due to a difficulty in selecting a target that was clearly identifiable as another number. Finally, there was no effect of Target, F (1, 30) = 2.577, $p = .119$, $\eta_p^2 = .079$, $BF_{10} = 0.605$, and no Distractor x Target interaction, F (1, 30) = 0.183, $p = .672$, $\eta_p^2 = .141$, $BF_{10} = 0.146$.

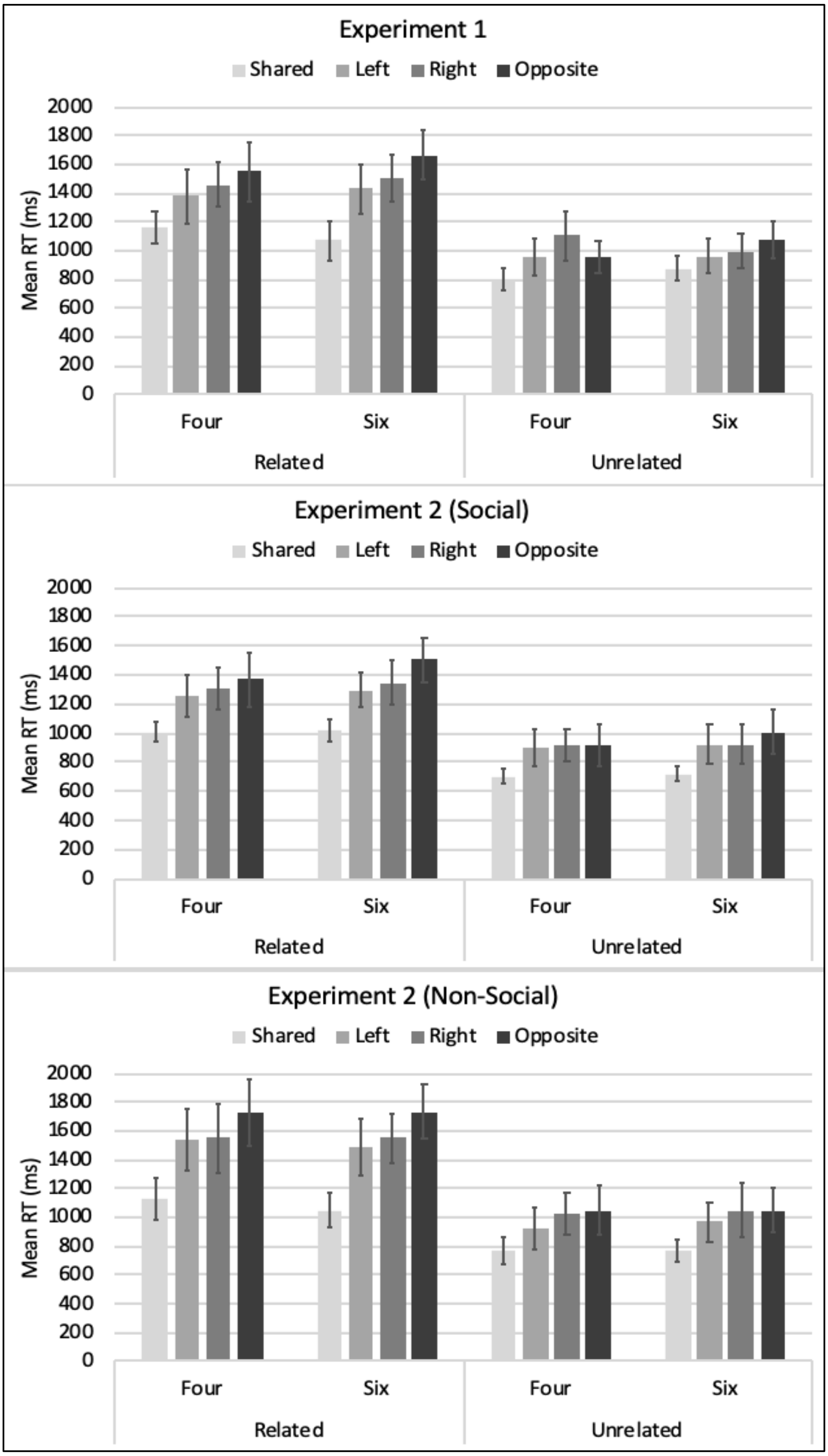

Figure 3. *Condition mean response times (ms) as a function of the relatedness of the distractor, the target digit, and the four possible perspectives (with 95% confidence intervals).*

Overall, the results from the analysis of RTs confirms our expectations that targets paired with a physically identical distractor (Related condition) would be harder than trials with a physically different distractor (Unrelated condition). The frequency of Incorrect – Altercentric responses, however, remained stable across both conditions.

We conducted a further, exploratory test concerning the time course of perspective-taking. Specifically, after computing another perspective it will be necessary to 'return' to our own prior to making a response. We can therefore predict that Incorrect – Altercentric responses could represent responses made *too early*, while the participant was still taking the other perspective. Response times (RTs) on Incorrect – Altercentric responses were compared to RTs for correct trials for the same participant on the same trial type, taking into account avatar position (left or right), distractor type (Related or Unrelated), and target (4 or 6). This analysis incorporated data from only 13 participants (those who made Incorrect-Altercentric responses), and thus can only be seen as offering an indication as to whether the hypothesis might be supported with a more suitably-powered design. The results are displayed in Figure 2B. Although altercentric errors were 104 ms faster (95% CI [-39,248]) than matched correct trials, this difference was not statistically significant, $t(12) = 1.586$, $p = .139$, $r = .350$, $BF_{10\ (\text{Correct RT} > \text{Altercentric RT})} = 1.391$. Finally, there was no evidence of a speed-accuracy trade-off, $r(29) = -.159$, $p = .393$.

Discussion

In Experiment 1, participants were given a task in which they were instructed to find a target from an avatar's perspective but to indicate its location according to their own. We found that the most common error on trials where the avatar was on either the right or left was the selection of the empty square that was correct from the computed perspective. Participants never indicated the empty square that was incorrect from either perspective. This demonstrates that participants were making a manual response upon a computed perspective rather than their actual one, contrary to instructions, and despite the obvious salience of one's own perspective.

In Experiment 2, we investigated whether this occurs only when we are instructed to take the perspective of social stimuli like avatars. It could be that we expect such agents to have a perspective, and therefore the need to briefly reduce the salience of our own perspective (to avoid competition) is more pressing. On the other hand, if we instruct participants to imagine the grid 'upright' rather than instruct them to take the perspective of an avatar, participants might adopt a different strategy, one that might not lead to such a reduction in egocentricity, and hence not produce as many Incorrect – Altercentric responses. In order to make this comparison, we replaced the avatar with a black dot of equivalent size and told one half of participants that the dot represented an avatar looking at the grid (Social condition), and the other half that the dot marked the bottom of the grid and that the target should be found assuming the display was upright (Non-Social). This distinction is very similar to the embodiment vs. array rotation difference described earlier, except we couched the distinction in social vs. non-social terms by describing the dot as an avatar or as a spatial marker. If computing perspectives rather than rotating arrays is what leads to altercentric responses, then Incorrect – Altercentric responses should be limited to or at least more frequent in the Social condition. In all other respects the task was identical to that in Experiment 1.

Experiment 2

Method

Participants

A total of 62 (new) participants took part in Experiment 2. An equal number of randomly selected participants were assigned to the Social and Non-Social conditions. Participant numbers were therefore the same as in Experiment 1 for each of the two new conditions, providing a sample size predicted to be powerful enough to detect medium effect sizes on a test comparing the number of Incorrect – Altercentric responses and Incorrect – Control responses. Given that we had actually found a medium-to-large effect size for this comparison in Experiment 1 (r = .408), and our hypothesis that this difference might be due to the instruction to take the perspective of a social stimulus, a further power analysis found that a between-group Mann-Whitney U test with an 80% chance of detecting a similar effect size (d = 0.66, one-tailed) requires 31 participants in each group. One participant's data were removed for evidence of their not having following instructions (always taking one perspective), leaving 61 (30 Non-Social: 11 male, 19 female, 27 right-handed, Mean age = 21, range = 18-27; 31 Social: 13 male, 18 female, 30 right-handed, Mean age = 21, range = 19-35). All participants provided informed consent and were compensated financially for their time.

Materials and procedure.

A black dot of equivalent size to the avatar replaced the avatar in Experiment 2 (see Fig 1, right panel). Participants were instructed to find a digit the avatar (now a black dot) named from the avatar's perspective (Social condition) or locate the named numeral assuming the black dot was the bottom of the grid (Non-Social condition), but always to respond with a key that corresponded to the target's location from the participant's own perspective. In all other respects the procedure was identical to that of Experiment 1.

Results

Types of errors. Accuracy across the task as a whole was very high (Social M = 94%, SD = 7%; Non-Social M = 95%, SD = 7%). The number and type of errors are displayed in Figure 2A. Overall, there were 41 Incorrect – Altercentric responses in the Non-Social condition (N = 960 possible errors, 15 participants[2] made at least one such error) and 42 in the Social condition (N = 992 possible errors, 12 participants[3] made at least one such error). That is, only one more Incorrect –

---

[2] Specifically, seven participants made one error each, four made two each, one made four, two made five and one made twelve.

[3] Specifically, four participants made one such error each, two made two each, three made three each, one made four, one made seven and one made fourteen.

Altercentric response was made in the Social than Non-Social condition. In each group, there was only one Incorrect – Control response. A Wilcoxon matched-pair signed-rank test found that the median of differences (Incorrect - Altercentric responses minus Incorrect – Control responses) was higher than zero in both the Social condition, W(31) = 0, Z = 3.077, *p* = .002, *r* = .391, and the Non-Social condition, W(30) = 10, Z = 3.207, *p* = .001, *r* = .414. A Mann-Whitney *U* test found no evidence that the number of Incorrect - Altercentric responses differed between conditions, *U*(61) = 428, Z = 0.589, *p* = .556, *r* = .075.

In the Social condition, 17 (40%) of Incorrect – Altercentric responses occurred in the Unrelated condition and 25 (60%) in the Related condition, a difference that was not significant, W(31) = 34.5, Z = 1.469, *p* = .142, *r* = .186. From the same (left and right) perspectives, only one error was a timeout, and 5 were Incorrect – Distractor responses. As in Experiment 1, on left/right perspective trials there were more Incorrect - Altercentric responses than errors of other types combined, with the relevant comparison reaching significance, W(31) = 16, Z = 2.530, *p* = .011, *r* = .321.

Of the Incorrect – Altercentric responses in the Non-Social condition, 29 (71%) occurred in the Related condition, and 12 (29%) on Unrelated trials, a difference that was significant, W(30) = 91.5, Z = 2.507, *p* = .012, *r* = .324. From the same (left and right) perspectives, an additional 9 errors were timeouts, and 22 were Incorrect – Distractor responses. Again, on left/right perspective trials there were more Incorrect - Altercentric responses than errors of other types combined, though only numerically, W(30) = 112.5, Z = 0.105, *p* = .916, *r* = .014.

Response times on errors. We next compared the difference in RTs between Incorrect – Altercentric responses and matched correct trials (again, RTs for correct trials for the same participant on the same trial type, taking into account avatar position (left or right), distractor type (Related or Unrelated), and target (4 or 6)) for the Social and Non-Social conditions, as we did in Experiment 1 (see Figure 2B). Four Incorrect - Altercentric errors in the Social condition were not included in this analysis owing to the absence of any correct responses by the relevant participant on that trial type. It is important to note that only a small number of participants' data were included in this analysis, many of whom contributed only one data point and hence their mean scores represent a single trial. This, together with a significant Shapiro-Wilks test for one of the four cells suggesting non-normal distribution, leads us to suggest results are interpreted with caution. A mixed-design 2 x 2 ANOVA with Group: (Social vs. Non-Social) as a between -subjects factor and repeated measures over the factor Type (Correct vs. Altercentric Error) found an interaction, F(1, 25) = 5.492, MSE = 64364, *p* = .027, $\eta_p^2 = .18$., but no main effect of Type or Group (*p*s > .15). Planned comparisons confirmed that Incorrect – Altercentric responses occurred more quickly than matched correct trials in the Social condition (*M* = 1186 ms vs. *M* = 1442 ms respectively), *t*(11) = 2.346, *p* = .039, $BF_{10\ (Correct\ RT > Altercentric\ RT)} = 3.923$, but not the Non-Social condition (*M* = 1657 ms vs. *M* = 1587 ms), W(15) = 78, Z = 1.022, *p* = .307), *r* = .001, $BF_{10\ (Correct\ RT > Altercentric\ RT)} = 0.161$. As an additional test, we also analysed the RTs for both the Social condition and Experiment 1 combined, as both instructed participants to take the perspective of a social stimulus. This analysis therefore included data from 25 participants, providing greater power than either experiment in isolation. A paired-samples t-test found that Incorrect – Altercentric responses occurred more quickly (*M* = 1242 ms) than matched correct trials (*M* = 1419 ms), *t* (24) = 2.807, *p* = .01, $BF_{10\ (Correct\ RT > Altercentric\ RT)} = 9.686$. Finally, there was no evidence of a correlation between overall accuracy and overall RT in the Non-Social condition, *r(28)* = -.273, *p* = 144, but in the Social condition faster performance was associated with higher, not lower accuracy, *r(29)* = -.422, *p* = .018. Thus, the faster times for altercentric errors are highly unlikely to be the result of a speed/accuracy trade-off.

Other results. We analysed RTs (initially non-normally distributed, so logtransformed) by means of a 2: Group (Social vs. Non-Social) x 4: Perspective (Shared x Left x Right x Opposite) x 2: Target (4 vs. 6) x 2: Distractor (Related vs. Unrelated) mixed-design ANOVA with Group as the only between-subjects factor (condition means can be found in the bottom two panels of Figure 3). Levene's tests for homogeneity of variance revealed some concerns regarding trials from the shared and left perspectives. We therefore conducted parallel non-parametric tests (not reported here) and found that they conformed to the pattern of results described. The assumption of sphericity was not met in relation to the factor Perspective, $\chi^2 = 22.351$, $p < .001$, ε (Greenhouse-Geisser) = .779, and Greenhouse-Geisser corrected effects are reported as a result.

There was a significant main effects of Perspective, $F(2.337, 135.535) = 128.218$, $p < .001$, $\eta_p^2 = .689$, $BF_{10} = 6.538e+82$ (all BF values based on a null model with all unrelated factors and interactions included). Participants were fastest on shared-perspective trials (M = 895 ms), followed by left- (M = 1163 ms), right- (M = 1209 ms), and opposite-perspective (M = 1295 ms) trials, with all pairwise contrasts statistically significant at the p (adjusted) < .04 level. There was a main effect of Distractor, $F(1, 58) = 651.166$, $p < .001$, $\eta_p^2 = .918$, $BF_{10} = 2.168e+181$. Participants were 454 ms slower on Related than Unrelated distractor trials. There was a significant interaction between Perspective and Distractor, $F(3, 174) = 7.152$, $p < .001$, $\eta_p^2 = .011$, $BF_{10} = 30.323$, a result driven by particularly slow performance on Related trials compared to Unrelated trials when the opposite perspective was taken (and when the distractor therefore appeared at first glance to be a perfect match for the target). Mean response times in the Social (M = 1070 ms) and Non-Social conditions (M = 1211 ms) did not differ, $F(1, 58) = 1.722$, $p = .195$, $\eta_p^2 = .029$, $BF_{10} = 0.623$. Finally, there was no main effect of Target, $F(1, 58) = 3.726$, $p = .058$, $\eta_p^2 = .060$, $BF_{10} = 0.595$, and no further interactions[4].

[4] Perspective x Group, $F(3, 58) = 2.198$, $p = .090$, $\eta_p^2 = .037$, $BF_{10} = 0.491$; Target x Group, $F(1, 58) = 2.197$, $p = .144$, $\eta_p^2 = .037$, $BF_{10} = 0.371$; Target x Distractor, $F(1, 58) = 0.054$, $p = .817$, $\eta_p^2 = .001$, $BF_{10} = 0.098$; Distractor x Group, $F(1, 58) = 3.127$, $p = .082$, $\eta_p^2 = .051$, $BF_{10} = 2.870$; Perspective x Target, $F(3, 174) = 1.971$, $p = .120$, $\eta_p^2 = .033$, $BF_{10} = 0.074$; Perspective x Target x Group, $F(3, 174) = 1.373$, $p = .253$, $\eta_p^2 = .023$, $BF_{10} = 0.144$; Perspective x Distractor x Group, $F(3, 174) = 2.399$, $p = .070$ $\eta_p^2 = .040$, $BF_{10} = 0.335$; Target x Distractor x Group, $F(1, 58) = 1.277$, $p = .263$, $\eta_p^2 = .022$, $BF_{10} = 0.292$; Target x Distractor x Perspective, $F(3, 174) = 1.275$, $p = .285$, η_p^2 =

Discussion

Overall, Experiment 2 replicated and extended the findings of Experiment 1. Participants continued to make errors that indicated they acted upon a computed perspective. Crucially, the number of Incorrect – Altercentric responses did not differ between the Social and Non-Social conditions, suggesting that such errors are not restricted to taking another's perspective, but also occur when we need to see things 'the right way up'.

General Discussion

The most common error on trials when the avatar or black dot was on either the left or right was an Incorrect – Altercentric response; namely the key that corresponded to the correct choice from the computed rather than the participant's physical perspective. Not only did the number of these Incorrect – Altercentric responses exceed the number of Incorrect – Control responses, but (at least numerically) also all other error types from these perspectives combined. This is particularly remarkable given that Incorrect – Altercentric responses consisted of a response to an *empty* square rather than a distractor. This finding strongly suggests that participants made their response according to a computed perspective even though they should have responded according to their own.

The number of Incorrect – Altercentric responses did not differ between the Social and Non-Social conditions in Experiment 2, suggesting that the reduction in salience of one's egocentric perspective is not restricted to taking the perspective of a social stimulus, but also occurs when we attempt to see things 'the right way up'. We found that these errors occurred more quickly than

---

.022, $BF_{10} = 0.151$; Perspective x Target x Distractor x Group, $F(3, 174) = 0.430$, $p = .732$, $\eta_p^2 = .007$, $BF_{10} = 0.061$.

correct responses in the Social condition, but altercentric errors and correct responses occurred at similar latencies in the Non-Social condition. It was only in the Non-Social condition that task difficulty showed any evidence of moderating performance, because there were more Incorrect – Altercentric responses on related than unrelated trials. These results come from analyses of only a small subset of the data and should therefore be interpreted with caution, but they suggest participants may have performed the Social and Non-Social conditions in subtly different ways. Precisely how they might differ is presently unclear. On the one hand, it seems that a serial process by which one first imagines oneself in the avatar's location prior to 'returning' to one's own can explain why altercentric errors occurred more quickly than correct trials in the Social condition. However, the same error occurs in the Non-Social condition, and with equivalent frequency, but not earlier than correct trials. This might suggest that a serial process was not involved when imagining in the Non-Social condition. However, we feel that until a larger-scale study which is able to elicit a larger number of errors can be conducted it is difficult to know whether this difference is reliable or not. What these results *do* show is that participants sometimes act upon computed level 2 perspectives regardless of whether that perspective is described as social or non-social, or promoted embodied perspective-taking or array rotation.

An important question that these results give rise to is how participants carried out the perspective-taking process. There are at least three possible accounts of our results. Firstly, participants might have integrated their motor representations with the imagined perspective, such that the top-left square from the desired perspective sometimes culminated in a top-left manual response. This account of altercentric errors would appear to be most consistent with the embodied perspective-taking account for Level 2 VPT favoured by many researchers (e.g., Erle & Topolinski, 2017; Kessler & Thomson, 2010). Secondly, participants may have engaged an array-rotation strategy, whereby they mentally rotated the grid in front of them. Thirdly, participants may have made altercentric errors because of a failure to resolve a conflict between location codes (e.g. top left vs. bottom left) ,without responses actually being prepared while in a non-egocentric reference frame.

Overall, we feel that our results are most consistent with the first, embodiment account, whereby our motor representations are temporarily integrated with a new perspective. This is because our task promoted such a strategy, which itself has been found to elicit embodiment, as well as because embodiment appears to be the easiest route for us to solve Level 2 VPT tasks (Wraga et al., 2000; Zacks & Tversky, 2005). There are other reasons why we feel embodiment may be the most appropriate account. Although the contrast did not reach statistical significance, performance in the social condition, which promoted embodiment, was 141 ms faster than performance in the non-social condition, which promoted array rotation. This pattern of data is in line both with an account of array rotation being the less efficient and dispreferred strategy. Finally, in our task participants needed to move their finger in one of four diagonal directions prior to pressing a response key. This movement, however short, represented an additional and explicitly *directional* motor movement which, in the case of Incorrect – Altercentric responses, would need to be consistent with the computed perspective. This movement could flag an altercentric error before it was made, leading to a correction, if the movement was inconsistent with our understanding of our physical position relative to the grid. The fact that altercentric errors occurred in spite of this movement suggests that participants thought they were acting consistently with their real-life environment.

However, there is an important caveat that our data place upon embodiment accounts, and this concerns the result that as many altercentric responses were made when participants were told the black dot represented the bottom of the screen as when they were told it was an avatar. Relative to the social conditions, this instruction should promote an array-rotation process. We believe that the finding of similar numbers of altercentric responses regardless of instruction can be reconciled if we posit a process whereby we update our motor representations to be consistent with the target to be viewed (i.e., the grid), but not with the real-world location required to view it. In other words, whether we imagine ourselves in a new location or imagine the array rotated, our motor representations update according to our new relationship with the array. This suggests that motor representations might not require imagined self-movement to occur, but at the same time array rotation becomes complementary to embodiment accounts, since it forms another way by which our motor representations are integrated with an imagined perspective. We do not wish to suggest that there is no difference between array rotation and imagined self-movement, for there is ample evidence that there is (Amorim & Stucchi, 1997; May, 2004; Presson, 1982; Schurz, Aichhorn, Martin, & Perner, 2013; Wraga et al., 2000; Zacks & Tversky, 2005). Our point is that they might both lead to an updating of motor representations vis-à-vis the target.

An alternative, non-motoric account for altercentric errors is that they are the outcome of interference between the location codes, such that 'top-left' (avatar) might interfere with 'bottom-left' (participant), leading to the occasional erroneous response that has the appearance of an action upon a computed perspective but is instead the result of competing codes. Aside from the evidence from previous research that appears to privilege embodied perspective-taking accounts for this type of task, which we have already discussed, an important problem for this type of account concerns how the non-egocentric location code is retrieved in the first place if the computed perspective is not read. In this sense, whether participants make an altercentric response because of location code interference or because they made a manual response consistent with the computed perspective, both are manual responses based on the computed perspective. Overall, therefore, we feel that the most parsimonious explanation of our results is that participants generated a representation of a new, imagined relationship between their selves and the grids, and that this sometimes culminated in altercentric errors.

Our findings might also help inform research into how we interact physically with objects in complex scenes, and whether these manual actions—when they are based on memory of the target object's location rather than online perception—rely on egocentric or allocentric (relative to other external objects) frames (Klinghammer, Schütz, Blohm, & Fiehler, 2016; Schütz, Henriques, & Fiehler, 2013). Although our findings are based on button-presses rather than manual reaches for targets, they do suggest that we can make manual responses upon non-egocentric perspectives not only from memory but also online[5]. They are also consistent with the view that we treat some non-egocentric perspectives as if they were egocentric (Filimon, 2015).

Finally, our results suggest a number of potentially fruitful avenues for future research. We found that participants generally had no problem making responses from their own frame of reference as required, but an interesting question our study raises is why participants made altercentric errors when they did. The most likely explanation is a failure of performance monitoring; that is, participants sometimes failed to register that their motor representations were still integrated with the computed perspective at the moment of response. An interesting question for future work might therefore be to ask participants whether they were *aware* of their making altercentric errors. Additionally, given that visual perspective-taking is but one ability that is brought under the umbrella of theory of mind, another direction might be to design similar, error-based tasks to test whether we simulate others' *beliefs*. Although our task was about exclusively visuo-spatial perspectives, some have speculated that this ability may have been an evolutionary precursor of other, non-visuo-spatial forms of perspective-taking (Kessler & Thomson, 2010). If this is correct, then the principle that taking others' perspectives might lead to (brief) modulations of our own might extend beyond visuo-spatial perspectives alone. There already exists evidence that we are more likely to approximate others' responses to trivia questions (such as the year Einstein first visited the US) immediately after we have taken that person's visual perspective (Erle & Topolinski, 2017). An interesting extension of our research might therefore be to check whether being asked to process another's belief leads people occasionally to make judgments contrary to their own belief system.

[5] We are grateful to Dimitris Voudouris for this suggestion.

Acknowledgements

This work was supported by an ESRC grant (ES/M008460/1: PI = NSC) (SS, CM, RL and NSC) and by a Leverhulme grant (RPG-2014-353: PI = NSC) (EL and NSC). We are grateful to Greg Davis for fruitful discussions of our results.

Declarations of interest: None.

Open practices statement: The data for these experiments have been made available on a third-party archive (https://doi.org/10.17863/CAM.26026.) Original materials and accompanying E-prime scripts are available from the lead author.

References

Amorim, M.-A., & Stucchi, N. (1997). Viewer-and object-centered mental explorations of an imagined environment are not equivalent. *Cognitive Brain Research, 5*(3), 229-239.

Deroualle, D., Borel, L., Devèze, A., & Lopez, C. (2015). Changing perspective: The role of vestibular signals. *Neuropsychologia, 79*, 175-185.

Erle, T. M., & Topolinski, S. (2017). The grounded nature of psychological perspective-taking. *Journal of Personality and Social Psychology, 112*(5), 683-695.

Ferguson, H. J., Apperly, I., & Cane, J. E. (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *Quarterly Journal of Experimental Psychology, 70*(8), 1646-1660.

Filimon, F. (2015). Are all spatial reference frames egocentric? Reinterpreting evidence for allocentric, object-centered, or world-centered reference frames. *Frontiers in Human Neuroscience, 9*, 648.

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), 99-103.

Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology, 1*, 213.

Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition, 114*(1), 72-88.

Klinghammer, M., Schütz, I., Blohm, G., & Fiehler, K. (2016). Allocentric information is used for memory-guided reaching in depth: a virtual reality study. *Vision Research, 129*, 13-24.

Lurz, R. (2009). If chimpanzees are mindreaders, could behavioral science tell? Toward a solution of the logical problem. *Philosophical Psychology, 22*(3), 305-328.

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development*, 357-366.

May, M. (2004). Imaginal perspective switches in remembered environments: Transformation versus interference accounts. *Cognitive Psychology, 48*(2), 163-206.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & psychophysics, 68*(2), 327-337.

Parsons, L. M. (1987). Imagined spatial transformations of one's hands and feet. *Cognitive psychology, 19*(2), 178-241.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526.

Presson, C. C. (1982). Strategies in spatial reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(3), 243-251.

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in Human Neuroscience, 7*, 712.

Schütz, I., Henriques, D., & Fiehler, K. (2013). Gaze-centered spatial updating in delayed reaching even in the presence of landmarks. *Vision Research, 87*, 46-52.

Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science, 10*(2), 199-204.

Surtees, A., Apperly, I., & Samson, D. (2013a). Similarities and differences in visual and spatial perspective-taking processes. *Cognition, 129*(2), 426-438.

Surtees, A., Apperly, I., & Samson, D. (2013b). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in Human Neuroscience, 7*, 698.

Wraga, M., Creem, S. H., & Proffitt, D. R. (2000). Updating displays after imagined object and viewer rotations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 151-168.

Yu, A. B., & Zacks, J. M. (2017). Transformations and representations supporting spatial perspective taking. *Spatial Cognition & Computation, 17*(4), 304-337.

Zacks, J. M., & Tversky, B. (2005). Multiple systems for spatial imagery: Transformations of objects and bodies. *Spatial Cognition and Computation, 5*(4), 271-306.