2020-06

Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze

Ward, E

10.1016/j.cognition.2020.104241
Cognition
Elsevier BV

Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze.

Eleanor Ward*, Giorgio Ganis, Katrina L. McDonough, Patric Bach

School of Psychology, University of Plymouth, Drake Circus, Devon, UK, PL4 8AA

*Corresponding author/lead contact: Eleanor.ward@plymouth.ac.uk

Word count: 7996

Abstract

Other peoples' (imagined) visual perspectives are represented perceptually in a similar way to our own, and can drive bottom-up processes in the same way as own perceptual input (Ward, Ganis & Bach, 2019). Here we test directly whether visual perspective taking is driven by where another person is looking, or whether these perceptual simulations represent their position in space more generally. Across two experiments, we asked participants to identify whether alphanumeric characters, presented at one of eight possible orientations away from upright, were presented normally, or in their mirror-inverted form (e.g. "R" vs. "Я"). In some scenes, a person would appear sitting to the left or the right of the participant. We manipulated either between-trials (Experiment 1) or between-subjects (Experiment 2), the gaze-direction of the inserted person, such that they either (1) looked towards the to-be-judged item, (2) averted their gaze away from the participant, or (3) gazed out towards the participant (Exp. 2 only). In the absence of another person, we replicated the well-established mental rotation effect, where recognition of items becomes slower the more items are oriented away from upright (e.g. Shepard and Meltzer, 1971). Crucially, in both experiments and in all conditions, this response pattern changed when another person was inserted into the scene. People spontaneously took the perspective of the other person and made faster judgements about the presented items in their presence if the characters were oriented towards upright to them. The gaze direction of this other person did not influence these effects. We propose that visual perspective taking is therefore a general spatial-navigational ability, allowing us to calculate more easily how a scene would (in principle) look from another position in space, and that such calculations reflect the spatial location of another person, but not their gaze.

Keywords: *visual perspective taking; perceptual simulation; gaze cuing; navigation; mental rotation; mental imagery*

Perspective taking as virtual navigation: perceptual simulation of what others see reflects their location in space but not their gaze.

Visual perspective taking (VPT) allows people not only to derive *what* others can see, but also *how* they see it (Flavell, Everett, Croft, & Flavell, 1981). Doing so helps us navigate our social environment (Kozhevnikov, Motes, Rasch & Blajenkova, 2006), engage in joint action (Freundlieb, Kovács, & Sebanz, 2016) and work out what others will do next (Bach, Fenton-Adams, & Tipper, 2014; Creem-Regehr, Gagnon, Geuss, & Stefanucci, 2013; Kovacs, Teglas, & Endress, 2010). Deriving how the world appears to other people may also form a foundation for more sophisticated abilities to reason about others' mental states, such as their beliefs, emotions, and goals (Batson, Early, & Salvarani, 1997; Erle & Topolinski, 2015; Tomasello, Carpenter, Call, Behne, & Moll, 2005; Mattan, Rotshtein & Quinn, 2016; for meta-analytic neuroimaging evidence, see Schurz, Aichhorn, Martin & Perner, 2013). Yet, the mechanisms underlying this ability, and the stimulus features that trigger shifts to another's perspective, are largely unclear.

Recent research has started to conceptualize perspective taking as a form of perceptual simulation, which simulates relevant aspects of another's view and "paints" or "inserts" them onto one's perceptual processes, as if they were one's own perceptual input (Kampis, Parise, Csibra, & Kovács, 2015; Surtees, Apperly, & Samson, 2013; Ward, Ganis, & Bach, 2019; see Cole & Millet, 2019, for a critical view). If another's perspective were represented in such a (quasi-)perceptual manner, it could drive action and decision making processes just like own input, explaining the link between VPT and higher level mentalizing abilities (e.g., Hamilton, Brindley, & Frith, 2009; Schurz et al., 2013) and between VPT and joint action (Freundlieb, Kovács & Sebanz, 2016; Tversky & Hard, 2009).

Direct evidence for such a proposal had been lacking however. Consistent with a similar representational format of one's own and others' perspectives, it has been shown that one's own judgments of a stimulus are harder if another person around us would make the same judgments differently. People find it harder, for example, to judge how many objects they can see from their own perspective when another actor sees a different number (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010), to identify a character as a "6" if it looks like a "9" to another person (Surtees, Samson, & Apperly, 2016), or to judge whether an object is to the left or the right to oneself when this object is in another location for another person (Tversky and Hard, 2009; Zwickel, 2009; Zwickel and Müller, 2010). Yet, these studies leave open whether the interference happens on the perceptual, conceptual, response or even metacognitive level, perhaps simply indexing the uncertainty when one becomes aware that others would judge – or respond to – the same stimulus differently, without representing how specifically they see it.

A recent series of studies from our lab (Ward et al., 2019) provided the first direct evidence that another's visual perspective is represented perceptually. We reasoned that if people have (quasi-)perceptual access to the content of another's' perspective, then this perspective should not only interfere with one's perceptual judgements (Sampson et al., 2010; Surtees et al., 2016), but could also *facilitate* them, precisely if this other perspective would offer a better view on an object than one has oneself. To test this, we let participants take part in a variant of the classic mental rotation task (Shepard & Metzler, 1971) in which they simply judged whether alphanumeric characters presented at varying orientations were 'normal', or mirror-inverted (e.g. "R" v. "Я"). The classic finding is that these judgements become progressively slower with increasing angular disparity away from upright (Cooper, 1975; Shepard & Metzler, 1971) because people must mentally rotate each item back into its canonical orientation before judging it. We found that this classic pattern changed dramatically when another person, who would see the items from a different perspective, was inserted into the scenes. Items oriented away from participants were identified more quickly if they appeared upright to the other person, as if people could simply identify it from this person's alternative perspective. Conversely, they judged it more slowly if the item was rotated even further away from the other person. Regression analyses showed that recognition times across letter orientations can be predicted from the angular disparity of the item not only to the participant as it is the case in classical studies on mental rotation (e.g., Shepard & Metzler, 1971), but also to the other person, suggesting that participants mentally rotated and identified the items from both their own and the other's perspective.

These data showed that people can represent the *content* of another's (imagined) viewing perspective in a form that can drive perceptual judgements in the same bottom-up manner as our own perception (e.g. Roelfsema & de Lange, 2016), allowing them to make judgments from others' perspectives that would be more difficult from our own. Importantly, these shifts to the others' perspective occurred spontaneously, even when the persons in the scene were completely task-irrelevant. Yet, they were enhanced when participants were explicitly instructed to take the others' perspective but eliminated when the other person was replaced by an inanimate object (e.g., a lamp, Ward et al., 2019), suggesting that they capture another's ability to see (or at least mentally represent) the world.

This paradigm provides us with a unique window into the processes that underlie visual perspective taking. Here, we use this task to test which cues drive the spontaneous shifts into another's visual perspective. One important candidate is another person's gaze. Gaze has long been associated with perspective taking, exemplified perhaps by the folk-psychological notion that perspective taking allows one to "see through another's eyes", and several findings that demonstrate a close link between mentalizing processes and gaze. People are highly sensitive to others' gaze, to the extent that one's own spatial attention is involuntarily directed to the object another person is looking at (for a review, see Frischen, Bayliss and Tipper, 2007). Importantly, these effects are at least partially driven by mental state attributions. They disappear when the person looks at an item but cannot see it, for example, when wearing opaque glasses (Morgan, Freeth & Smith, 2019). Other findings show that people assimilate another's emotional response towards a gazed-at object into their own evaluations (e.g., Bayliss, Paul, Cannon & Tipper, 2006; for review, Becchio, Bertone & Castiello, 2008). Moreover, a person's gaze signals how they intend to act upon an object (Ambrosini, Costantini & Sinigaglia, 2011), it promotes later (automatic) imitation of these actions (Wang & Hamilton, 2012), and increases activity in brain regions related to Theory of Mind and mentalizing (e.g., Mosconi, Mack, McCarthy & Pelphrey, 2005; Williams, Waiter, Perra, Perret & Whiten, 2005). Together, these finding suggest that seeing someone gaze at an object might not only guide our own attention to this object but may also serve as a trigger to represent how the gazed-at object looks from their perspective.

Strikingly, there is very little evidence about whether another's eye gaze is an important cue for representing another's visual perspective of an object. Most studies have demonstrated only that following another's gaze triggers more limited judgments of *what* this person can or cannot see (Level 1 perspective taking), not *how* they see it (Level 2 perspective taking). Indeed, this limited form of Level 1 perspective taking is often described as a simple process of "mentally scanning" (Kosslyn, Ball, & Reiser, 1978) the line of sight between another's eyes and the objects they can potentially see, which can induce interference when they see different objects to oneself (e.g., Michelon & Zacks, 2006; Surtees et al., 2013). Indeed, disrupting an observed person's line of sight (e.g. by giving them opaque glasses) or replacing the other person with a non-seeing object (e.g. an arrow) reduces the interference effects between one's own and the other perspective (e.g., Furlanetto, Becchio, Samson & Apperly, 2016; Nielsen, Slade, Levy, & Holmes, 2015).

It is an open question whether others' eye gaze also contributes to Level 2 perspective taking, people's ability to work out *how* objects look from another person's perspective. To the best of our knowledge, only two studies have investigated this question, with both providing non-conclusive results. Mazzarella, Hamilton, Troiano, Mastromauro and Conson (2012) found that seeing another person reach for an object induces spatial interference effects from the others' perspective, so that participants were less likely to judge object location from their own perspective. However, merely seeing the other person shift their eyes towards the object (and not reaching for it) did not induce a detectable shift to their perspective. Similarly, Furlanetto, Cavallo, Manera, Tversky and Becchio (2013) showed that, while viewing someone gaze at an object numerically increased shifts to their spatial perspective, this was not statistically robust, and even removing gaze cues altogether (by blurring the eye region) did not affect whether people judged objects from another's or their own perspective.

Another possibility is therefore that spontaneous visual perspective taking is separate from these gaze-based processes and primarily driven by the presence – and spatial location – of another person in the scene. In other words, people might be putting themselves generally into the shoes of others, without truly looking through their eyes. The relevance of another's location in space to perspective taking is well supported. Several studies have shown, for example, that it takes longer to explicitly judge where an object is from another's perspective the more this person is oriented away from us, as we mentally rotate our own bodies into the other's perspective, with this transformation taking longer the more we are oriented away from them (Kessler & Thompson, 2010; Kessler & Rutherford, 2010; Kozhevnikov et al., 2006; Surtees et al., 2013). Consistent with such location-based accounts, other studies link perspective taking not to Theory of Mind and mentalizing, but to more fundamental navigational abilities (Allen, Kirasic, Dobson, Long & Beck, 1996; Hegarty & Waller, 2004; Kozhevnikov et al., 2006), such as the ability to judge from which other locations one could interact with the environment (Gunalp, Moossaian & Hegarty, 2019; Kessler & Thompson, 2010).

Our paradigm provides an ideal means to disentangle whether viewing another's gaze at an object triggers shifts into their visual perspective of it. By measuring how much faster items are identified the more they are oriented towards other people in the scenes (and slower the more they are oriented away from them), our paradigm provides a quantitative measure of participants' tendency to spontaneously take another's visual perspective. If another's gaze is critical to trigger these VPT processes, then these recognition biases should only be seen for an actor who gazes at the items but be eliminated if the actor is not looking directly at them, but gazes elsewhere. If, on the other hand, VPT represents a mental transformation from one's own position in space into that of another, then the gaze direction of this other person will not influence recognition times, and *all* conditions will show shifts towards the incidentally presented people in the scenes, even if they gaze away from the relevant items. Here, we test these predictions in two experiments.

2. **Experiment 1 – manipulating gaze between trials**

Experiment 1 tests how another's gaze affects perspective taking, using a within-subject design. Participants were shown a series of scenes depicting a table in front of them. In each trial, an alphanumeric character appeared on the table presented either normally, or in its mirror-inverted form, at varying rotations away from upright. Participants had to simply judge, with a button press, whether the items were mirrored or normal ("R" vs. "Я"). In some trials, another person was seen sitting to the left or the right of the table. Our previous studies have shown that people spontaneously make use of another person's perspective when making these judgements. When letters are facing another person, it was easier to judge whether they are mirrored or normal than it would be if this other person was not there or sitting in the opposite direction, indicating that people judged the letters through this person's eyes. Here we manipulated – between trials – the direction in which the person was looking, in order to test whether people's tendency to spontaneously represent how an object looks to another person is affected by whether this person is actually looking at the object. In different trials, the person on the screen could either (1) look at the item on the table, as in our prior work, or (2) look away from it to the wall in the room. If perspective taking takes the others' gaze into account, then the ability to recognize an object from another's perspective should be enhanced – or only found – when the person is actually looking at the to-be-judged item, but less so if looking at the participant or the room's back wall. If, however, perspective taking emerges from a general rotation of one's own body into the spatial location of another then it should happen irrespective of where the actor is looking.

2.1. Method

2.1.1. Participants

Thirty-seven (31 females) naive participants were initially recruited via the University of Plymouth participation pool. All participants were adults (age range 18-35) and gave written informed consent according to the declaration of Helsinki. Approval was obtained from the University of Plymouth Ethics Committee. Participants received course credit as compensation. As in Ward et al. (2019), participants with more than 20% error rates across all conditions were not considered for analysis (*n*=5). The remaining thirty-two participants

(27 females; mean age: 20.1 years, range: 18-35) provide 80% power to detect effects in the range of d =.45. Prior work on this paradigm (Ward et al., 2019) has revealed that effect sizes are substantially larger (.747 < d < 1.08 for the main perspective taking effect).

2.1.2. Apparatus, stimuli and procedure

The study was conducted in behavioural testing lab space of the Action Prediction lab of the University of Plymouth. It was administered using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Stimuli were presented on a 19” LED computer monitor (Resolution: 1900x1200; Refresh rate: 60Hz). Responses were made on a standard computer keyboard with UP, DOWN, and SPACE keys as active response keys. Red and green stickers were positioned on the DOWN and UP keys, respectively.

Participants sat upright facing the screen at a distance of approximately 60cm and were given written and verbal instructions. Before the experiment proper commenced, they were given examples of the rotated items that would appear on the screen and completed eight training trials that were identical to the main experiment. The experiment proper consisted of 520 trials. Each trial (Figure 1) started with a fixation cross displayed for 400ms, followed by 300ms blank screen. The subsequent stimulus sequence included three frames, each measuring 33.4 by 23.5 degrees of visual angle. The first frame was presented for 800 ms. In one third of the trials, it showed a view onto a corner of a square table in a grey room without a person being present (No-Person trials). The remaining trials showed a person (either male or female) sitting at the same square table, gazing outwards at the participant. The person sat either on the left side of the table (Person-left trials) or on its right side (Person-right trials), so that they were positioned at roughly 270 or 90 degree angle to the position at which the rotated items would appear, relative to the participant.

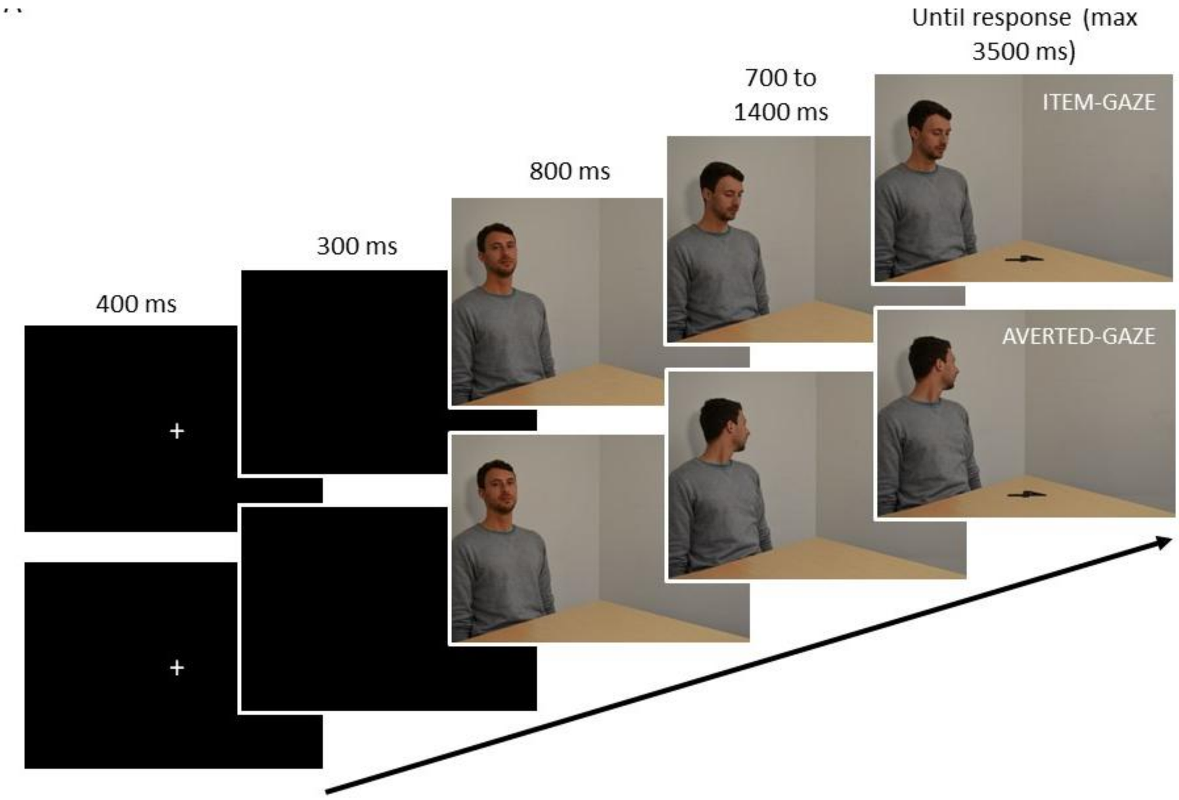

Figure 1. **Example stimuli used in Experiment 1.** Looking direction (Item-gaze or Averted-gaze) varied between trials.

The second frame in the sequence was identical to the first for the No-Person trials. In the other trials, the person looked down towards the middle of the table in half the trials, where the item would appear, and in the other half they looked away towards the corner of the room. Because this second frame was presented without inter-stimulus interval, it created the impression of apparent motion (Wertheimer, 1912), so that the person seemed to shift their gaze from the participant towards the table or towards the back of the room. Then, after a random interval between 700ms and 1400ms, the third frame was presented. This was identical to frame 2, but now one of 48 possible items appeared on the table, at the location on the table the on-screen person was gazing at. This item was one of three alphanumeric characters (4, P, or R), presented either in the canonical version or mirror-inverted about their vertical axis, in one of eight orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°, with 0° denoting the upright canonical orientation and angles increasing in a counter-clockwise fashion) relative to the participant. The characters always appeared in the same position on the table, half-way between the outward corner of the table and its centre, such that the persons to the left and right would gaze at the table from roughly 90° and 270°, respectively (perpendicular to the viewpoint of the participant). Character rotation occurred around the character's centre point.

Participants were asked to judge whether each character was presented in its canonical or mirror-inverted form (e.g. "R" vs "R"). Participants responded using their right hand by pressing the green key to indicate a canonical item and the red key to indicate a mirrored item. Response times were measured relative to item onset. The third frame remained on the screen until a response was made to a maximum duration of 3500ms.

2.2. Analysis

Data (pre-)processing and analysis was identical to Ward et al., (2019) and conducted in Microsoft Excel (2010) and JASP (2018). Violin plots were created using Raincloud Plots (Version 1; Allen, Podiaggi, Whitaker et al., 2019). Power analyses were conducted in G*Power (Version 3.1; Erdfelder, Faul, & Buchner, 1996).

Dependent measures were the recognition times (measured from item onset) for each character orientation (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315), depending on Person location (No-Person, Person-left, Person-right) and Gaze direction (Item-gaze, Averted-gaze).

As in our original study (Ward et al., 2019), the recognition times for each angle and condition were mapped onto two orthogonal and statistically independent summary measures, so that they can be compared across conditions, without accruing alpha inflation due to multiple testing, which would result if each of the eight angles were compared separately. These two summary measures are derived by treating each participant's recognition time for this character orientation as a vector in a polar coordinate system, with the recognition time providing the distance from the origin and the rotation angle the polar angle.

The first summary measure (Toward/Away-bias) quantifies the mental rotation effect (Shepard & Meltzer, 1971) and indexes to what extent characters are recognized faster the more they are facing towards the participants (0°) rather than away from them (180°), separately for each participant and condition (No-Person, Person-left Item-gaze, Person-left Averted-gaze, Person-right Item-gaze, and Person-right Averted-gaze). A character orientation's contribution to the Toward/Away-bias is derived simply from the recognition times multiplied with the negative of the cosine of the orientation angle. As a result, characters contribute negatively the more they face the participants (315°, 0°, 45°) and positively they more they are oriented away from them (225°, 180°, 135°). Positive values on this score therefore indicate faster recognition of letters oriented towards rather than away from the participant.

The second summary measure (Left/Right-bias) indexes how much faster characters were recognized when oriented towards the left (270°) rather than right (90°), and therefore allows us to quantify whether people present in these location induce shifts in how quickly these letters can be identified. It was derived analogously to the Towards/Away-bias. The contribution of a character's orientation to the Left/Right-bias was calculated as the recognition time multiplied with the sine of the orientation angle. Character orientations contribute negatively the more they face to the right (45°, 90°, 135°) and positively the more they face to the left (225°, 270°, 315°). A zero value on the Left/Right-bias therefore indicates that items oriented left and right are identified equally quickly, while positive numbers indicate faster identification for left- compared to right-oriented letters, and vice versa for right-oriented letters.

By averaging these values, separately for each summary measure, participant and condition (No-Person, Person-left-item, Person-left-averted, Person-right-item, Person-right-averted), we can calculate, first, whether items are recognized faster the more they appear in the canonical orientation to the participant compared to when they are oriented away (positive values on the Toward/Away-bias), reflecting the expected mental rotation effect. Similarly, they allow us to calculate to what extent characters are recognized faster the more they are oriented leftwards and would appear in their canonical orientation to a person sitting to the left (positive values on the Left/Right-bias) rather than rightwards, appearing in their canonical orientation to a person sitting on the right (negative values). We were then able to determine if this difference in left/right biases changed depending on whether the other person was looking at the object or not. Note that the direct comparison of the Person-left and Person-right conditions is statistically identical to the comparison of how much the presence of person shifts mental rotation performance in the Person-left and Person-right conditions relative to the No-Person baseline (i.e. how much person presence shifts recognition times away from 0 towards either 90^0 or 270^0), as this would involve subtracting the same baseline value from each of the two conditions for each participant, and would therefore not affect the absolute difference between them.

2.2.1. Across-participant regression analyses

In prior work, the mental rotation effect is typically additionally characterised in terms of a linear regressions of an items' recognition time to its angular disparity relative to the participant, for each participant separately (Shepard & Meltzer, 1971). The results reveal linear increases with increasing angular disparity for the large majority of participants. Here, we used this analysis model to test whether an item's recognition times can be described, on a single participant basis, as a linear increase of the character's angular disparity *both* to the participant and to the other person. To this end, we entered each participant's item mean recognition times for each character orientation in each condition (No-Person, Person-item, Person-averted) as dependent variable in a multiple regression analysis, with the item's angular disparity to the participant and to the other person as two statistically independent predictors. This analysis provides regression coefficients for both predictors – angular disparity to participant and other person – for each participant and condition separately. We report mean across-participant regression coefficients for each of these two predictors and compare them with t-tests against zero, and against each other.

3. Results

Erroneous responses (7% on average) were excluded from the analysis of recognition times (RTs), as well as trials with RTs longer than 2000ms, or shorter than 150ms.

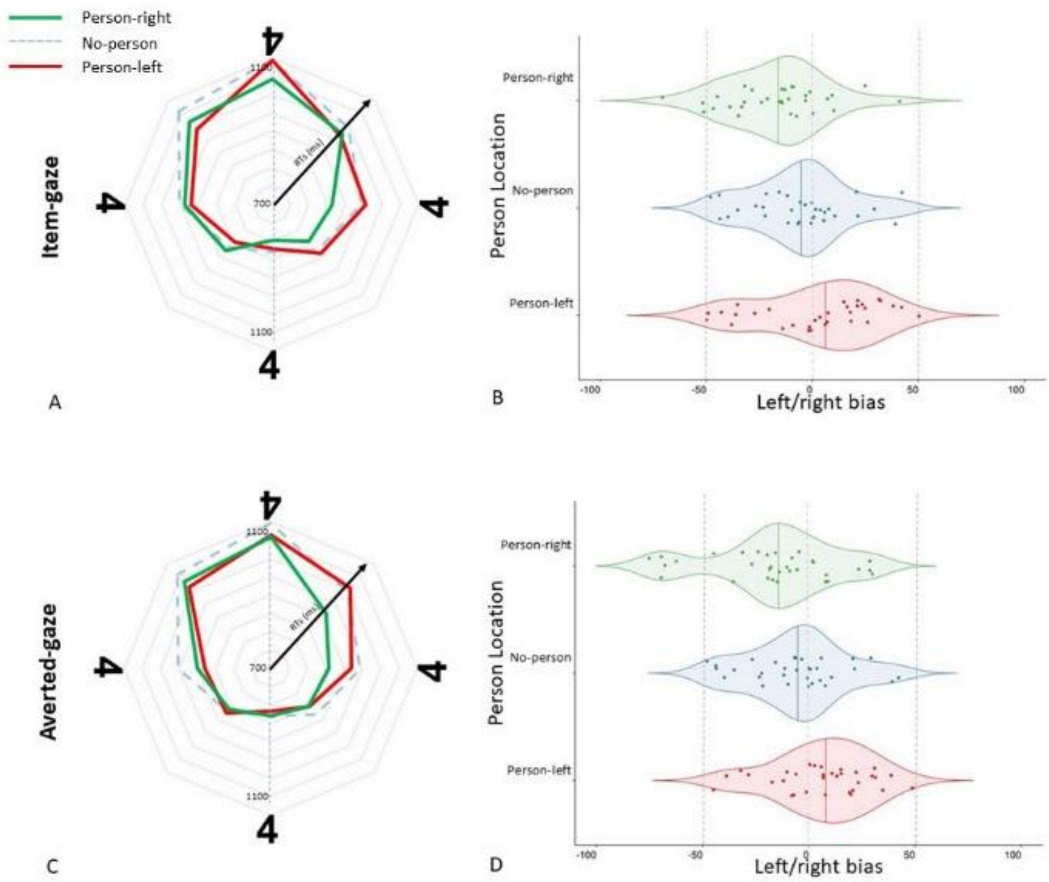

Figure 2. **(A)** Mean recognition times (RTs) for mirror-inverted/canonical judgements for Item-gaze trials when the actor is inserted on the right and the left showing faster recognition times for items oriented upright to the actor. (B) Left/right bias when the inserted person is looking at the table showing faster recognition times for items appearing upright to the actor. (C) Mean recognition times (RTs) for mirror-inverted/canonical judgements for Averted-gaze trials (D) Left/right bias when then inserted person appears to look away from the table (towards to wall).

3.1.Mental rotation

We first determined whether our data replicated the classical mental rotation effect (Shepard & Metzler, 1971). To this end, we first compared the overall (across conditions) towards/away bias, indexing, in milliseconds, how much more slowly items are identified the more they are oriented away from the participant compared to towards them, with a simple t-test against zero. The towards/away bias across all conditions was positive, M=56.34;

SD=28.05, *t*(31)=11.36, *p*<.001, *d*=2.01, BF_{10}=5.797e+09, showing – unsurprisingly – that items are identified more quickly the more they are oriented towards rather than away from participants. This mental rotation effect was also confirmed by regressing each item's recognition time to the expected linear increase with angular disparity, as in prior research, (Shepard & Metzler, 1971; Ward et al., 2019), revealing positive slopes in all participants, mean β =1.4; *t*(31)=14.60, *p*<.001, *d*=2.6, BF_{10}=3.181e+12.

We then verified that this overall mental rotation effect was not affected by person presence. As the actors would sit at 90 and 270 degree angle to the participant, and their location was therefore orthogonal to the towards/away axis, we did not expect that the presence/location of the other person would affect the overall mental rotation effect. Indeed, a 2x3 ANOVA on the towards/away-biases across conditions with the factors Gaze (item, averted) and Location (person left, person right, no person) did not reveal any main effects or interactions, *F*<2.68, *p*>.08, for all. When the six conditions were analysed separately, slower recognition of turned away items was present in all conditions, *t*(31) > 8.29, *p* <.001, *d* >1.5 for all.

3.2.Perspective taking

The main question was whether people would spontaneously take the perspective of the other person in the scenes, such that items were recognized faster when oriented towards compared to away from them, and whether this effect, in turn, was determined by where this person was looking. We therefore calculated, for each participant and condition separately, the left/right-bias (for details, see analysis-section), indexing, in ms., how much faster items were identified the more they were oriented to the left than to the right. The resulting left/right-bias summary scores were then entered into a 2 x 2 repeated measures ANOVA with within-subjects factors of Location (Person-left, Person-right) and Gaze (Averted-gaze, Item-gaze). This analysis revealed decisive evidence for a main effect of Location, *F*(1,31)=27.921, *p*

<.001, ηp^2=.474, BF_{10}=4379.04. As in our prior work (Ward et al., 2019), left/right-biases were more negative (indexing faster recognition of rightwards- than leftwards oriented letters) when someone was sitting on the right, and more positive (indexing faster recognition of leftwards- than rightwards oriented letters) when someone was sitting on the left. Strikingly, this analysis provided moderate evidence against a main effect of Gaze, $F(1,31)$=.545, p=.466, ηp^2=.017, BF_{10}=.203, and against an interaction of Location and Gaze, $F(1,31)$=.291, p=.593, ηp^2=.009, BF_{10}=.216, showing that the location of the actor, and not where they were looking, was the critical cue to induce perspective taking. Indeed, when the difference in left/right-bias between Person-left and Person-right conditions was compared separately for each gaze condition, the predicted differences were apparent for both Item-gaze trials $t(31)$=-3.287, p=.003, d=.58, BF_{10}=14.55, and Averted-gaze trials $t(31)$=5.044, p<.001, d=1.01, BF_{10}=1145.37, with effect sizes, if anything, being larger in the away trials.

Table 1. Means (M) and Standard Deviations (SD) for the Left/Right- and Towards/Away-biases in Error rates in Experiment 1 and Experiment 2. Relates to Figure 2 and Figure 4. Forward/Away and Left/Right-biases were calculated analogously as for the recognition times. *p<.05. **p<01, ***p<.005.

		Toward/away bias			Left/right bias		
Exp.	Looking direction	Person-left M (SD)	Person-right M (SD)	No-Person M (SD)	Person-left M (SD)	Person-right M (SD)	No-Person M (SD)
1.	Item	.021 (.027)***	.022 (.027)***	.021 (.024) ***	-.0001 (.024)	-.005 (.02)	.0004 (.02)
	Averted	.024 (.028) ***	.025 (.026)***	.021 (.024) ***	.003 (.021)	-.006 (.022)	.0004 (.02)
2.	Item	.028 (.026) ***	.024 (.026) ***	.022 (.03) ***	.004 (.011)	-.001 (.017)	.004 (.015)
	Averted	.028 (.027) ***	.031 (.032) ***	.022 (.025) ***	-.0019 (.014)	-.0023 (.015)	-.003 (.016)
	Participant	.034 (.029) ***	.025 (.024) ***	.033 (.022) ***	-.007 (.02)	-.01 (.021)	-.005 (.019)

3.3. Regression analyses

As in our previous studies, we tested whether recognition times could be described as independent linear increases depending on an item's angular disparity to the participant as well as the other person, by using both disparities as orthogonal predictors in a simple regression model, for each participant and condition separately (and then comparing them against zero). Overall, these revealed very strong evidence for independent contributions of both the angular disparity to the participant, mean β =1.39, $t(31)=11.09$, $p<.001$, $d=1.96$, $BF_{10}=3.278e+09$, and to the other person, mean β =.27, $t(31)=4.95$, $p<.001$, $d=.87$, $BF_{10}=919.8$, showing that recognition times can be described by independent mental rotation functions form one's own and the other person's perspective.

To test how these relationships were affected by the other person's gaze, regression coefficients for each participant and condition were entered into a 2 x 2 repeated measures ANOVA with Gaze (Item, Averted) and Perspective (Self, Other) as within-subject factors. This analysis only revealed a main effect of Perspective, $F(1,31)=68.96$, $p<.001$, $\eta p^2=.69$, $BF_{10}=2.118e+18$ showing that, as expected, angular disparity towards the participants determined recognition times to a stronger extent that angular disparity to the other person. Replicating the main analysis, the analysis provided moderate *against* a contribution of gaze, neither revealing a main effect of Gaze, $F(1,31)=.267$, $p=.609$, $\eta p^2=.009$, $BF_{10}=.24$, nor an interaction of Gaze and Perspective, $F(1,31)=.044$, $p=.835$, $\eta p^2=.001$, $BF_{10}=.193$. In Item-gaze trials, both the angular disparity away from upright to the participant, mean $\beta=1.39$, $t(31)=9.55$, $p<.001$, $d=1.7$, $BF_{10}=1.044e+08$, and to the actor, mean $\beta=.25$, $t(31)=3.273$, $p=.003$, $d=.58$, $BF_{10}=13.98$ contributed to recognition times. This pattern was the same in the Averted-gaze trials, with the stronger contribution again coming from the angular disparity to the participants, mean $\beta=1.42$, $t(31)=10.81$, $p<.001$, $d=1.9$, $BF_{10}=1.799e+09$, and the smaller contribution coming for the angular disparity to the other person in the scene, mean $\beta=.30$, $t(31)=5.25$, $p<.001$, $d=.91$, $BF_{10}=2002.84$.

3.4. Exploratory analyses

One might speculate that variations in an observed individual's gaze might make a relatively late contribution, when the mechanism that analyses *where* they are looking has time to interact with the perceptual representations of *how* the task relevant item looks to them. To address this question, in a post-hoc analyses prompted by a reviewer, we replicated the main 2 x 2 repeated measures ANOVA (see 3.2 Perspective taking) with the within-subjects factors of Location (Person-left, Person-right) and Gaze (Averted-gaze, Item-gaze), but added the additional factor SOA (Short, Long), indicating whether participants saw the person look at or away from the item for more or less than 1050ms before the rotated item appeared. However, this analysis did not reveal the crucial interactions that would indicate such an influence, $F(1,31) < 1$, with Bayesian analysis providing evidence against such an interaction, $BF_{10} = 0.197$. Numerically, however, shifts into the other's perspective indeed decreased more strongly for longer SOAs in the away-gaze (short, 31ms; long 22ms) than the item-gaze trials (short, 27ms; long, 23ms).

4. Experiment 2 – manipulating gaze between participants

Experiment 1 confirmed that the mere presence of a person is sufficient to induce a shift into their perspective. As in our prior research (Ward et al., 2019), we found that rotated items were recognized more rapidly if they appeared closer to their canonical orientation to another person in the scenes. Moreover, recognition times were well described by mental rotation functions centred on both one's own and the other person's perspective. Strikingly, however, these shifts into the others' perspective were independent of the other person's gaze and occurred irrespective of where they are looking at the to-be-recognized item or away from it.

This suggests that taking another's visual perspective is based on their location in space, but not their gaze at (or away from) the relevant items.

One reason for this lack of a difference might be that within-participant experiments such as ours suffer from influential companion (e.g., Poulton, 1992) and carry-over effects (MacFie, Bratchell, & Greenhoff, 1989), for example when a participant's memory of the actor's gaze (e.g. at the item) in a previous trail influenced their processing in the current trial (e.g. at the wall). In Experiment 2, we therefore varied gaze *between* participants to rule out such an explanation. In addition, we added a third condition, in which the on-screen actors looked outwards at the participant. Prior work suggests that such gaze at the participant plays a special role in social cognition. It serves as an alerting signal, captures attention, and eliminates gaze cuing altogether (e.g. Mazzarella et al. 2012; for a review, see Hamilton, 2016). It may therefore provide the strongest manipulation of gaze. If visual perspective taking takes into account what another actually sees through their eyes – rather than what they *could* see from their position in space – then it should now be disrupted.

4.1. Method

4.1.1. Participants

Ninety-one (61 females) naive participants were initially recruited via the University of Plymouth participation pool. All participants were adults (age range 18-35) and gave written informed consent according to the declaration of Helsinki. Approval was obtained from the University of Plymouth Ethics Committee. Participants received course credit as compensation. After exclusion (>20% error), 80 participants (51 females; mean age: 19.9

years, range: 18-30) were considered for analysis, with at least 26 participants for each condition. This provides enough power to detect effects in the range of $d = .5$ for our critical comparison of left/right-bias across person locations. Results of the previous study (Ward et al., 2019) and as well as Experiment 1 show that effects are likely larger.

4.1.2.. Apparatus, stimuli and procedure

Participants completed 584 trials following the same apparatus and procedure as in Experiment 1. Between participants, we varied whether the actor always looked out towards the participant (participant-gaze), looked directly at the item on the table (item-gaze), or looked away from the item towards the corner of the room (averted-gaze). In each group, an equal number (1/3rd) of trials showed a person on the left, a person on the right, or no person. The starting frame in each trial showed the table, with either the person present or absent. Depending on the participant's group allocation, the person in the scenes either looked directly at the future item location, at the corner of the room or outwards towards participant. This frame was displayed for a randomized time between 1500 and 2220ms. It was directly followed by the appearance of the item on the centre of the table for 3500ms (while the person remained present), or until a response was made (Figure 3).

Note that, in contrast to Experiment 1, the stimulus sequences here did not include apparent motion in which the person in the scene would first look outwards at the participant and then switch their gaze towards the object or elsewhere. This initial frame was removed so that the outward-gaze frames could be used as experimental stimulus. It also ensures that participants view the specific gaze – and represent that view it affords of the scene – for the whole action sequence, and not just during the last frame. Removing apparent motion was possible because our prior study (Ward et al., 2019) had shown that person presence induced shifts into the others' perspective of equal size irrespective of whether apparent motion drew attention to the persons/items in the scenes or not.

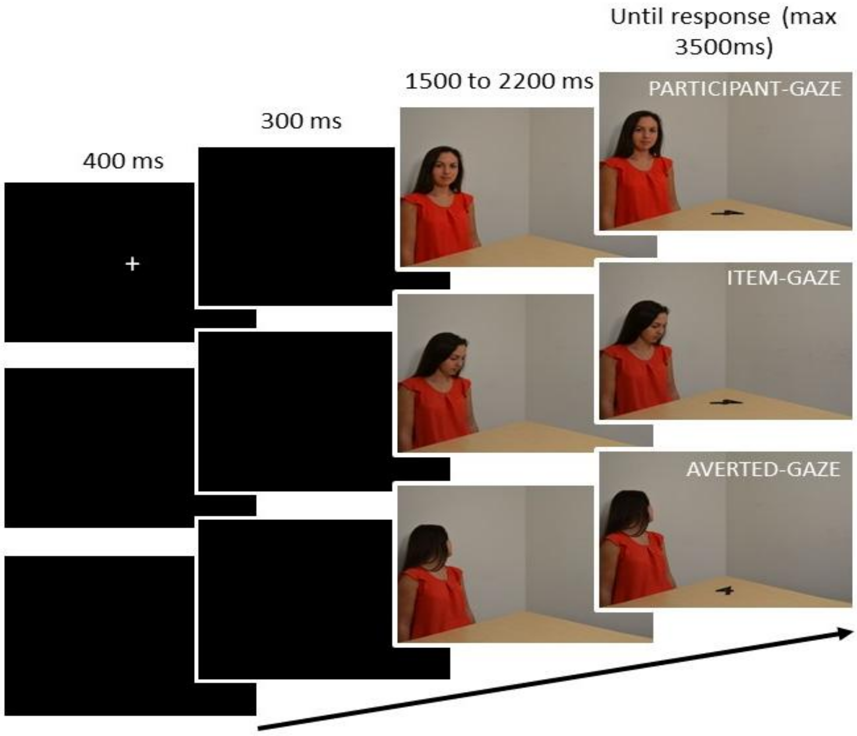

Figure 3. **Example stimuli used in Experiment 2.** Participants viewed only one of the three possible looking directions in a between-subjects design.

5. Results

As in Experiment 1, erroneous responses (10% on average) were excluded from the analysis of recognition times (RTs), as well as trials with RTs longer than 2000ms, or shorter than 150ms.

5.1. Mental rotation

We again first verified that our data would replicate the mental rotation effect (Shepard & Metzler, 1971). As in Experiment 1, a simple t-test comparing toward/away bias across all conditions against zero confirmed that recognition times were generally slower the more items were rotated away from participants, $t(79)=20.69$, $p<.001$, $d=2.31$, $BF_{10}=2.223e+30$. This mental rotation effect was also confirmed by regressing each item's recognition time to the expected linear increase with angular disparity, revealing positive slopes in all bar one participant, mean β =.85; t(79)=54.96, $p<.001$, $d=6.15$, $BF_{10}=9.342e+60$.

To confirm that this overall mental rotation effect was not affected by person presence, we ran a 3x3 ANOVA on the towards/away-biases across conditions with the factors Gaze (item, averted, participant) and Location (person left, person right, no person) as in Experiment 1. As before, this did not reveal any main effects or interactions, $F(2/4,154/77)<2.47$, p>.088, BF<.404 for all. The mental rotation effect was present in all conditions, $t(25/26) > 9.17$, p <.001, $d > 1.8$, for all.

5.2.Perspective taking

We then tested whether the task would replicate the shift to the other person's perspective and whether this shift was modulated by their gaze. Left/right-biases from all participants were entered into a 2 x 3 repeat measures ANOVA with within-subjects factors of Location (Person-left, Person-right) and between-subject factors of Gaze (Averted-gaze, Item-gaze, Participant-gaze). It again revealed a main effect of Location, $F(1,77)=29.472$, $p<.001$, $\eta p^2=.277$, $BF_{10}=15474.14$. As can be seen in Figure 3, left/right-biases were more negative when someone was sitting on the right and more positive when someone was sitting on the left. Thus, letters were identified more quickly when pointing leftwards when someone was also sitting on the right, but identified more quickly when oriented rightwards when someone was sitting on the left. However, as in Experiment 1, there was neither a main effect of Gaze,

F(2,77)=1.189, p=.310, ηp^2=.030, BF_{10} =.230 , nor an interaction of Location and Gaze, F(2,77)=2.11, p=.128, $\eta p2$=.052. BF_{10} = .554. Pairwise comparisons of the critical Left/Right actor locations for all conditions confirmed that regardless of where the actor was looking, recognition times were faster when items that would usually be difficult to recognise were easier when appearing upright from this person's position, t(26)=2.347, p=.027, d=.5, $BF_{10}=2.055$, t(26)=2.730, p=.011, d=.54, $BF_{10}=4.239$, and t(25)=4.173, p<.001, d=.89, BF_{10}=94.93 for Item, Participant, and Averted-gaze, respectively.

Note that in the participant (outward) gaze condition, responses seemed generally faster for items oriented towards the right compared to the left, as indicated in the radar plots (figure 4, panel b) and the perceptual shift measures for the three conditions (person-left, person-right, no person). Such a bias is sometimes (but not always) observed in studies on mental rotation (e.g. Koriat & Norman, 1985; Liesefeld & Zimmer, 2011; Robertson & Palmer, 1983). It simply reflects that many people make clockwise mental rotations (required for rightwards oriented items) more quickly than counter-clockwise ones (required for leftwards oriented items). However, this general shift is statistically independent from our measure of interest, which reflects the difference *within each group* between the perceptual shift between the person-left and person-right conditions. Moreover, as there was no main effect of Gaze in the main analysis, behaviour in the participant-gaze group did not differ statistically from other groups (item-gaze, averted-gaze), and the general rightward bias in this group was not robust when evaluated separately against the (Bonferroni-corrected) threshold for incidental findings in one of three groups (p=.054).

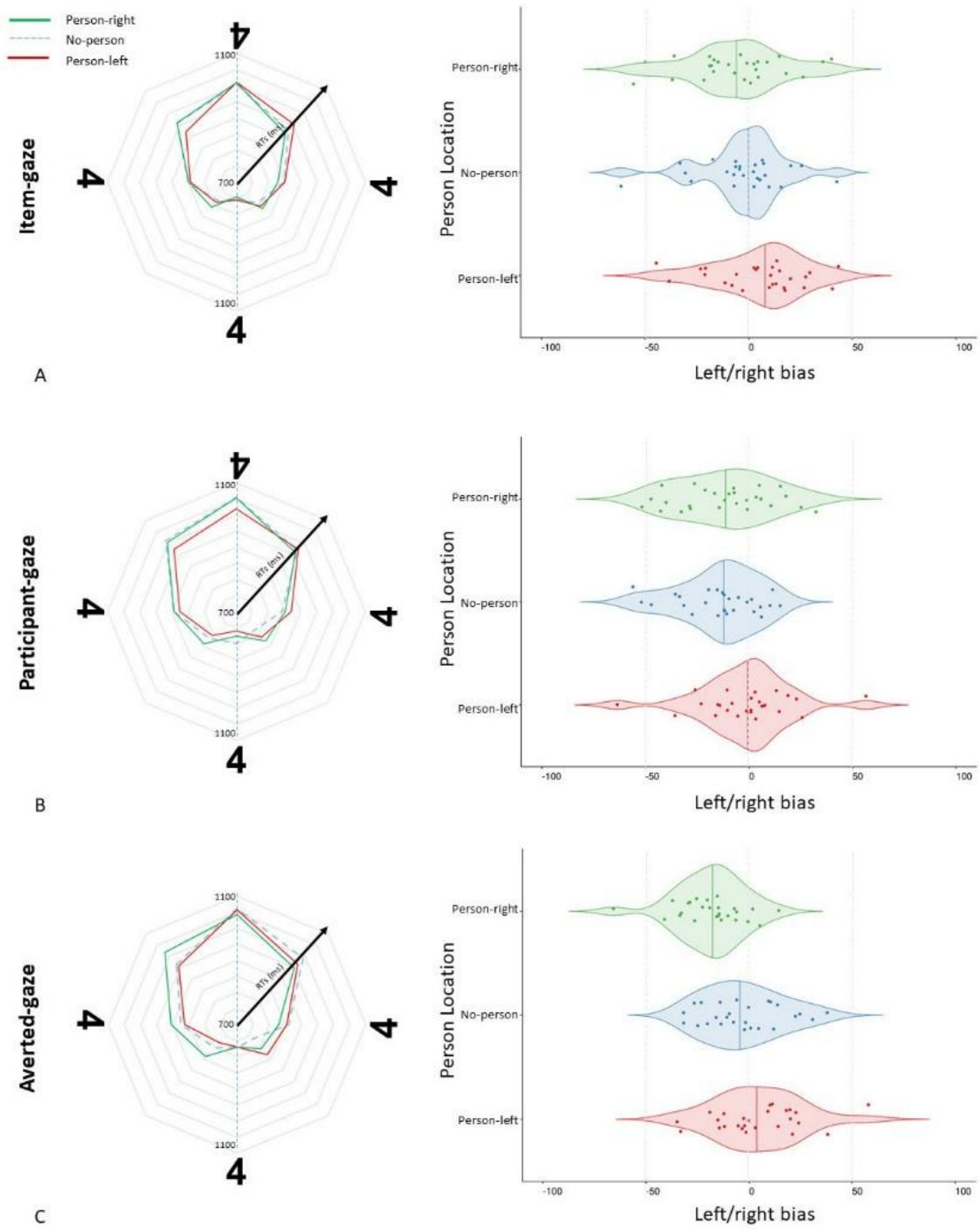

Figure 4. (Left) Mean recognition times (RTs) for mirror-inverted/canonical judgements and (Right) Left/right bias when the inserted person is looking at the table showing faster recognition times for items appearing upright to the actor for (A) Item-gaze, (B) Participant-gaze, and (C) Averted-gaze.

5.3.Regression analyses

As in Experiment 1, we tested whether recognition times can be described as independent linear increases depending on an items angular disparity to the participant as well as the other person. Mean beta coefficients for all gaze conditions were entered into a simple regression model, with the angular disparity of the character relative the actor and the participant as independent predictors of recognition times. To test for differences between gaze conditions, mean Beta for all participants were entered into a 3x2 ANOVA with the between-subjects factor, Gaze (item, participant, averted), and within-subjects factor, Perspective (own, other). As in Experiment 1, a 2 x 3 ANOVA revealed only a main effect of Perspective, $F(1,77)=285.99$, $p<.001$, $\eta p2=.788$, $BF_{10}=1.587e+34$, showing that the angular disparity of an item to one's own perspective determined recognition times to a stronger extent than to the other's perspective. There was no main effect of Gaze, $F(2,79)=.231$, $p=.794$, $BF_{10}=.099$, and no interaction of Perspective and Gaze, $F(2,77)=.23$, $p=.795$, $\eta p2=.006$, $BF_{10}=.130$. Step-down analyses confirmed that, in each condition, recognition times could be described by a linear combination of an item's angular disparity to the participant (Item-gaze, $\beta=1.45$, $t(26)=10.207$, $p<.001$, $d=1.95$, $BF_{10}=7.276e+7$; Averted-gaze, $\beta=1.57$, $t(26)=13.697$, $p<.001$, $d=2.64$, $BF_{10}=6.344e+7$; Participant-gaze, mean $\beta=1.63$, $t(25)=10.352$, $p<.001$, $d=2.03$, $BF_{10}=3.399e+10$) and to a lesser extent to the other person (Item-gaze, mean $\beta=.13$, $t(26)=2.234$, $p=.034$, $d=.429$, $BF_{10}=1.684$; Averted-gaze, $\beta=.13$, $t(26)=2.699$, $p=.012$, $d=.52$, $BF_{10}=68.782$; participant-gaze, $\beta=.28$, $t(25)=4.032$, $p<.001$, $d=.79$, $BF_{10}=3.982$).

5.4.Exploratory analyses

As in Experiment 1, we ran a post-hoc analyses whether the individuals' gaze affect the shifts into the others' perspective, particularly when the individual's gaze was seen for longer and has time to interact with the perceptual representations of their perspective. We again replicated the main ANOVA with within-subjects factor Location (Person-left, Person-right) and between-subject factors Gaze (Averted-gaze, Item-gaze, Participant-gaze), but added the additional factor SOA (Short, Long), indicating whether participants saw the person look at or away from the item for more or less than 1850ms before the rotated item appeared. One additional participant had to be excluded, because not all cells were filled for both the long and short SOA conditions. In contrast to Experiment 1, the predicted interaction of Gaze, Location and SOA was observed, $F(2,76)=5.694$, $p=.005$, $BF_{10}=8.482$, and followed the expected pattern. While the shift into the other's perspective increased with longer SOAs when seeing the other persons gaze at the relevant items ($t(26)=2.508$, $p<.019$, $d=.483$, $BF_{10}=2.761$), it did not change when seeing them looking away ($t(24)=.517$, $p=610$, $d=.10$, $BF_{10}=.238$), and decreased when looking outwards at the participants ($t(26)=2.122$, $p=.044$, $d=.408$, $BF_{10}=1.390$). While this result reveals the patterns expected of gaze-following affecting Level 2 perspective taking relatively late, please note that it reflects an incidental finding from a not pre-specified analysis, and in a paradigm that was not optimized for this contrast (e.g. SOAs varied randomly and were not fully counterbalanced across conditions). It therefore needs to be interpreted with caution before it is replicated.

6. General Discussion

Using our recently developed task (Ward et al., 2019), we tested whether spontaneous shifts into another's visual perspective depend on the other person looking at the to-be-judged item. We replicated the finding that people spontaneously represent – perceptually simulate – the content of another's viewing perspective and use it to recognize items that would be more difficult to recognize from their own perspective. Thus, participants identified leftwards oriented items oriented more quickly when another person was sitting to the left than on the right, and vice versa for rightwards oriented items. Moreover, as in our original study (Ward et al., 2019), recognition times were well described by mental rotation functions based on the item's orientation away from the participant *and* the other person. Our results therefore confirm that participants spontaneously shift their perspective into the location of another person and represent how an item would look from this position. Once represented in such a quasi-perceptual manner, the content of another's viewing perspective can feed into perceptual decision making processes in a similar manner as one's own perceptual input.

Here we asked whether these spontaneous acts of visual perspective taking only reflect the other person's location in space or whether they are also driven by the person's gaze at the task-relevant objects (or away from them). We therefore manipulated, in two experiments, whether the other person was directly looking at the item, towards the wall, or towards the participant (Experiment 2 only). Neither experiment revealed an influence of gaze on visual perspective taking. The signature perspective taking effect – that items are identified more quickly when appearing upright to another person – was robustly present irrespective of whether this person was looking at the item, towards the wall, or directly at the participant, and irrespective of whether looking direction was manipulated between-trials in Experiment 1 and between participants in Experiment 2, preventing any influence of carry over or influential companion effects (MacFie et al., 1989; Poulton, 1982).

These data show that spontaneous visual perspective taking – deriving how an object looks from another viewpoint – is primarily driven by the other person's location in space but does not take into account where they are looking. This insensitivity to another's gaze may be surprising under the (folk-psychological) assumptions that perspective taking is based on a sophisticated mechanism that allows us to "see" through another's eyes (e.g., Furlanetto et al., 2016). Consistent with such views, studies have established that following another's gaze towards an object induces shared/joint attention (for a review, see Frischen et al., 2007), and a bias to judge this object similarly as the other person (e.g., Becchio et al., 2008; Samson et al., 2010; Surtees et al., 2016). In contrast, disrupting another's gaze (e.g. by showing them with opaque glasses) reduces this crosstalk between own and others' judgments (e.g. Furlanetto et al., 2016). Strikingly, of course, all these studies measured relatively simple Level 1 perspective taking processes – whether the other person sees *the same objects* as oneself –, for which gaze is of obvious relevance. In contrast, our task taps into more sophisticated Level 2 processes for representing *how* specifically relevant objects will appear to another person. Our data – and previous failed attempts to demonstrate gaze-based effects in Level 2 perspective taking (Furlanetto et al., 2013; Mozzarella et al., 2012) – suggest that these abilities to embody another's view of the world may draw upon different, non-gaze-based mechanisms.

The prior literature is entirely consistent with such a view. While Level 1 and Level 2 perspective taking follow a developmental sequence, the mechanisms underlying both appear qualitatively different. Michelon and Zacks (2006) demonstrated that the time to judge *what* another person can or cannot see (Level 1 VPT) increases with the distance between the person and their potential targets. The location of the other person, relative to the participant, did not influence response times. This implies a straightforward tracing of the line between someone's eyes and an object and testing for potential obstructions. In contrast, Level-2 judgments of *how* an object looks to another person *does* crucially depend on this person's location, increasing the more their location and orientation differs from ours (e.g. Kessler & Rutherford, 2010; Kessler & Thompson, 2010; Kozhevnikov, Motes, Rasch & Blajenkova, 2006). In contrast to Level 1 judgments, this implies an imaginistic "embodied" process that mentally rotates one's own body into the position of the other person so that one can derive how an item would, in principle, look from this alternative vantage point (e.g., Kessler & Rutherford, 2010; Surtees et al., 2013), irrespective of what they are currently looking at.

This sensitivity to person location but not their gaze aligns this form of visual perspective taking with fundamental navigational abilities, particularly the ability to navigate an imagined environment, rather than with more sophisticated Theory of Mind and mentalizing abilities. Several authors have made such a case, reporting for example a correlational link between perspective taking and navigational skills, or that people are just as ready to judge items from the perspective of another person or from the perspective of a location they could, potentially, occupy (Gunalp et al., 2019; Hegarty & Waller, 2004; Kessler & Thompson, 2010; Kozhevnikov et al., 2006). Similarly, it is well-established that a key node of the theory of mind network, the Temporoparietal junction (TPJ) is also involved in visual perspective taking (e.g., Zacks, Rypma, Gabrieli, Tversky and Glover, 1999; for meta-analysis, see Schurz et al., 2013) and self-other distinction (David et al., 2006; Lamm, Bukowski, & Silani, 2016), perhaps on the basis of the spatial difference vector between self and other (e.g., Jeannerod, 2007). However, the TPJ, and posterior parietal regions in general, are also implicated in imagined navigation (Committeri, Piccardi, Galati & Guariglia, 2015), specifically the translation of allocentric hippocampal place representations into representation of a location into egocentric space (Boccia, Sulpizio, Palermo, Piccardi, Guariglia, & Galati, 2017). The perhaps most dramatic example for this relationship is the link between the TPJ and out-of-body-experiences, in which patients subjectively “see” the world (and themselves) from outside their own body. The TPJ is the one region typically impaired in patients that report these experiences and stimulation of the TPJ in healthy individuals, via transcranial magnetic stimulation (TMS), can induce them (Blanke et al., 2005). We therefore speculate that the ability to visually perspective-take might emerge from an evolutionary ancient mechanism that allows us to calculate what we can see – and what/how we can interact with – from different positions in space, therefore serving both epistemic and action functions. More sophisticated abilities for mentalizing and Theory of Mind can then capitalise on the emerging insights that other perspectives might provide one with very different knowledge of the same scenes (Frith & Frith, 2007; Kessler & Thompson, 2010; Mundy & Newell, 2007), giving rise to a common neuronal network for Theory of Mind, visual perspective taking, self-other distinction and (imagined) navigation.

Further studies need to establish how complete the distinction between Level 1 Perspective taking (*what* another person can see) and the Level 2 (*how* they see it) really is. Our study shows that spontaneous shifts into another's visual perspective are not driven by gaze, but merely their position in space. But what about other manipulations that vary whether an actor has visual access to the task-relevant item? For example, one might argue that because the task-relevant items were always in the actor's peripersonal space, the insensitivity to gaze might happen because the *potential* to look at the object is always there. What would happen, however, if the actor's view onto the item from their position was further away or blocked in principle, for example, by a placing a barrier between them (see Cole & Millet, 2019) or by giving the actor opaque glasses through which they would be unable to see (see Furlanetto et al., 2016). Such more drastic changes to what the person can or cannot see might have larger impact on how their visual perspective is represented.

Similarly, previous research has established a distinction between the ability to *spontaneously represent* somebody else's perspective and the ability to *intentionally select* it when it becomes task relevant (Schwarzkopf, Schilbach, Vogeley & Timmermans, 2014; Qureshi, Apperly & Sampson, 2010). Thus, while people might spontaneously integrate how an item looks to someone else into their own perceptual decision making, the ability to intentionally weigh this perspective over one's own might be impaired when it is clear that the other person does not look at – or does not have visual access to – the item in question. Note in this respect that, in Experiment 2, shifts into the other's perspective increased with longer exposure times when persons looked at the items, but decreased when they were looking away or at the participant. While we caution against over-interpreting this incidental finding (which was also not replicated in Experiment 1), this late changes are entirely consistent with such an secondary re-weighting process, with which people explicitly up- or down-regulate spontaneous shifts into another's perspective depending on whether they can currently see the object. Because our task provides a robust measure of these perspective shifts, and the resulting perceptual changes, it may provide a powerful window to resolve these interrelationships.

A final question for future research is how complete this simulation of other's perspectives is and what aspects of the other person are taken into account. In our experiments, the other person's posture remained consistent across conditions, and their head was averted to the corner of the room, or outwards towards the participant. What, however, would happen if not only the actor's head, but their whole body was turned to face away from the table in our task? Such studies would allow one to disentangle if visual perspective taking is primarily based on locations in space, or whether it is embodied, in the sense that it is derived from a mental model of oneself in a target posture (Amorim, Isableu, & Jarraya, 2006). If so, this would predict that one's own ability to embody another's posture through our own behaviour might influence these effects. Aligning one's posture with that of another person has been shown to make judgements from their perspective easier (Kessler & Thompson, 2010; Kessler & Rutherford, 2010). In our own task, could participants' inadvertent imitation of their virtual partner's posture then explain why mental rotation can be performed with such ease from their perspective, even without any instruction to do so?

6.1.Conclusion

This study provides the first evidence that perceptual shifts into another person's perspective are unaffected by where this person is looking. It therefore confirms that people perceptually represent the content of another's perspective in a (quasi)perceptual manner, but that these perceptual simulations are based on their location in space, but do not take into account whether this person can actually 'see' target items. These data link perspective taking to more fundamental navigational skills and the ability what one could, in principle, see from another's vantage point, irrespective of what this person is currently looking at.

References

Allen, G. L., Kirasic, K. C., Dobson, S. H., Long, R. G., & Beck, S. (1996). Predicting environmental learning from spatial abilities: An indirect route. *Intelligence, 22*(3), 327-355.

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of neurophysiology*, *106*(3), 1437-1442.

Amorim, M. A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations:" Body analogy" for the mental rotation of objects. *Journal of Experimental Psychology: General*, *135*(3), 327.

Bach, P., Fenton-Adams, W., and Tipper, S.P. (2014). Can't touch this: the first-person perspective provides privileged access to predictions of sensory action outcomes. *J Exp Psychol Hum Percept Perform*. 40(2), 457.

Batson, C. D., Early, S., and Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and Social Psychology Bulletin*. 23(7), 751-758.

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic bulletin & review*, 13(6), 1061-1066.

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, *12*(7), 254-258.

Blanke, O., Mohr, C., Michel, C. M., Pascual-Leone, A., Brugger, P., Seeck, M., ... & Thut, G. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *Journal of Neuroscience*, *25*(3), 550-557.

Boccia, M., Sulpizio, V., Palermo, L., Piccardi, L., Guariglia, C., & Galati, G. (2017). I can see where you would be: patterns of fMRI activity reveal imagined landmarks. *NeuroImage*, *144*, 174-182.

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

Committeri, G., Piccardi, L., Galati, G., & Guariglia, C. (2015). Where did you “left” Piazza del Popolo? At your “right” temporo-parietal junction. Cortex, 73, 106-111.

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, *7*(1), 20-43.

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., and Stefanucci, J.K. (2013). Relating spatial perspective taking to the perception of other's affordances: Providing a foundation for predicting the future behavior of others. *Frontiers in Human Neuroscience*. 7, 596.

David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., ... & Vogeley, K. (2006). Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience*, *18*(6), 898-910.

Driver IV, 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.

Erle, T. M., & Topolinski, S. (2015). Spatial and empathic perspective-taking correlate on a dispositional level. *Social Cognition*. 33(3), 187-210.

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

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.

Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: Knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *Journal of experimental psychology: human perception and performance*, *43*(6), 1065.

Frischen, Alexandra, Andrew P. Bayliss, and Steven P. Tipper. (2007)"Gaze cueing of attention: visual attention, social cognition, and individual differences." *Psychological bulletin* 133.4 (2007): 694.

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*(16), R724-R732.

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance, 42*(2), 158.

Furlanetto, T., Cavallo, A., Manera, V., Tversky, B., & Becchio, C. (2013). Through your eyes: incongruence of gaze and action increases spontaneous perspective taking. *Frontiers in Human Neuroscience*, *7*, 455.

Gunalp, P., Moossaian, T., & Hegarty, M. (2019). Spatial perspective taking: Effects of social, directional, and interactive cues. *Memory & Cognition*, 1-13.

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

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175-191.

JASP Team (2018). JASP (Version 0.10.0)[Computer software].

Jeannerod, M. (2007). Being oneself. *Journal of Physiology-Paris*, *101(4-6)*, 161-168.

Kampis, D., Parise, E., Csibra, G., and Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proc. R. Soc. B*. 282(1819), 20151683.

Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective-taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology*, *1*, 1–12.

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

Koriat, A., & Norman, J. (1985). Mental rotation and visual familiarity. *Perception & Psychophysics*, *37*(5), 429-439.

Kovács, Á. M., Téglás, E., and Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330(6012)*, 1830-1834.

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective- taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(3), 397-417.

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self–other representations in empathy: evidence from neurotypical function and socio-cognitive disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1686),* 20150083.

Liesefeld, H. R., & Zimmer, H. D. (2011). The advantage of mentally rotating clockwise. *Brain and cognition*, *75*(2), 101-110.

MacFie, H. J., Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first- order carry- over effects in hall tests. *Journal of Sensory Studies*, *4(2)*, 129-148.

Mazzarella, E., Hamilton, A., Trojano, L., Mastromauro, B., & Conson, M. (2012). Observation of another's action but not eye gaze triggers allocentric visual perspective. *The Quarterly Journal of Experimental* Psychology, 65(12), 2447-2460.

Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective-taking. *Percept. Psychophys.* 68, 327–337.

Morgan, E. J., Freeth, M., & Smith, D. T. (2018). Mental State Attributions Mediate the Gaze Cueing Effect. *Vision, 2(1)*, 11.

Mosconi, M. W., Mack, P. B., McCarthy, G., & Pelphrey, K. A. (2005). Taking an "intentional stance" on eye-gaze shifts: a functional neuroimaging study of social perception in children. *NeuroImage*, *27*(1), 247-252.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science, 16(5)*, 269-274.

Nielsen, M. K., Slade, L., Levy, J. P., & Holmes, A. (2015). Inclined to see it your way: Do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process?. *The Quarterly Journal of Experimental Psychology*, *68*(10), 1931-1951.

Poulton, E. C. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin*, *91*(3), 673.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230-236.

Robertson, L. C., & Palmer, S. E. (1983). Holistic processes in the perception and transformation of disoriented figures. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(2), 203.

Roelfsema PR, and de Lange FP (2016). Early visual cortex as a multi-scale cognitive blackboard. *Annual Reviews in Vision Science*. 2,131-151.

Samson, D., Apperly, I.A., Braithwaite, J.J., Andrews, B.J., and Bodley Scott, S.E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J Exp Psychol Hum Percept Perform*. 36(5), 1255.

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.

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

Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*. 171(3972), 701-703.

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

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43-52.

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97-105.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci*. 28, 675–691.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*(1), 124-129.

Wang, Y., & Hamilton, A. F. D. C. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Frontiers in human neuroscience*, 6, 153.

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*, *29*(5), 874-880.

Williams, J. H., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *Neuroimage*, *25*(1), 133-140.

Zacks, J., Rypma, B., Gabrieli, J. D. E., Tversky, B., & Glover, G. H. (1999). Imagined transformations of bodies: an fMRI investigation. *Neuropsychologia*, *37*(9), 1029-1040.

Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychonomic Bulletin & Review*, *16*(6), 1089-1093.

Zwickel, J., and Müller, H.J. (2010). Observing fearful faces leads to visuo-spatial perspective taking. *Cognition*. 117(1), 101-105.

Acknowledgements

We thank the members of the Action Prediction Lab, Plymouth University, (www.actionprediction.org) for discussion and comment on an earlier version of this article. Eleanor Ward was funded by a PhD student grant from the University of Plymouth.

Author contributions

EW and PB designed the experiment with GG and KM. EW programmed the study, prepared the stimuli and collected all data with KM. EW and PB and GG analysed the data. EW, PB, GG wrote the manuscript.

Declaration of interests

The authors declare no competing interests.

Data available at https://osf.io/2cxpr/

CRediT author statement

Eleanor Ward: Conceptualisation, Methodology, Software, Investigation, Formal analysis, Project Administration, Visualisation, Writing – Original Draft, Writing – Review & Editing. **Giorgio Ganis:** Conceptualisation, Resources, Writing – Original Draft, Supervision. **Katrina L McDonough:** Conceptualisation, Investigation, Writing – Original Draft, Project administration. **Patric Bach:** Conceptualisation, Methodology, Formal analysis, Writing – Original Draft, Writing – Review & Editing, Supervision.

Table(s)

Table 1. Means (M) and Standard Deviations (SD) for the Left/Right- and Towards/Away-biases in Error rates in Experiment 1 and Experiment 2. Relates to Figure 2 and Figure 4. Forward/Away and Left/Right-biases were calculated analogously as for the recognition times. **p*<.05. **p<01, ****p*<.005.

		Toward/away bias			Left/right bias		
Exp.	Looking direction	Person-left M (SD)	Person-right M (SD)	No-Person M (SD)	Person-left M (SD)	Person-right M (SD)	No-Person M (SD)
1.	Item	.021 (.027)***	.022 (.027)***	.021 (.024) ***	-.0001 (.024)	-.005 (.02)	.0004 (.02)
	Averted	.024 (.028) ***	.025 (.026)***	.021 (.024) ***	.003 (.021)	-.006 (.022)	.0004 (.02)
2.	Item	.028 (.026) ***	.024 (.026) ***	.022 (.03) ***	.004 (.011)	-.001 (.017)	.004 (.015)
	Averted	.028 (.027) ***	.031 (.032) ***	.022 (.025) ***	-.0019 (.014)	-.0023 (.015)	-.003 (.016)
	Participant	.034 (.029) ***	.025 (.024) ***	.033 (.022) ***	-.007 (.02)	-.01 (.021)	-.005 (.019)

Figure(s)

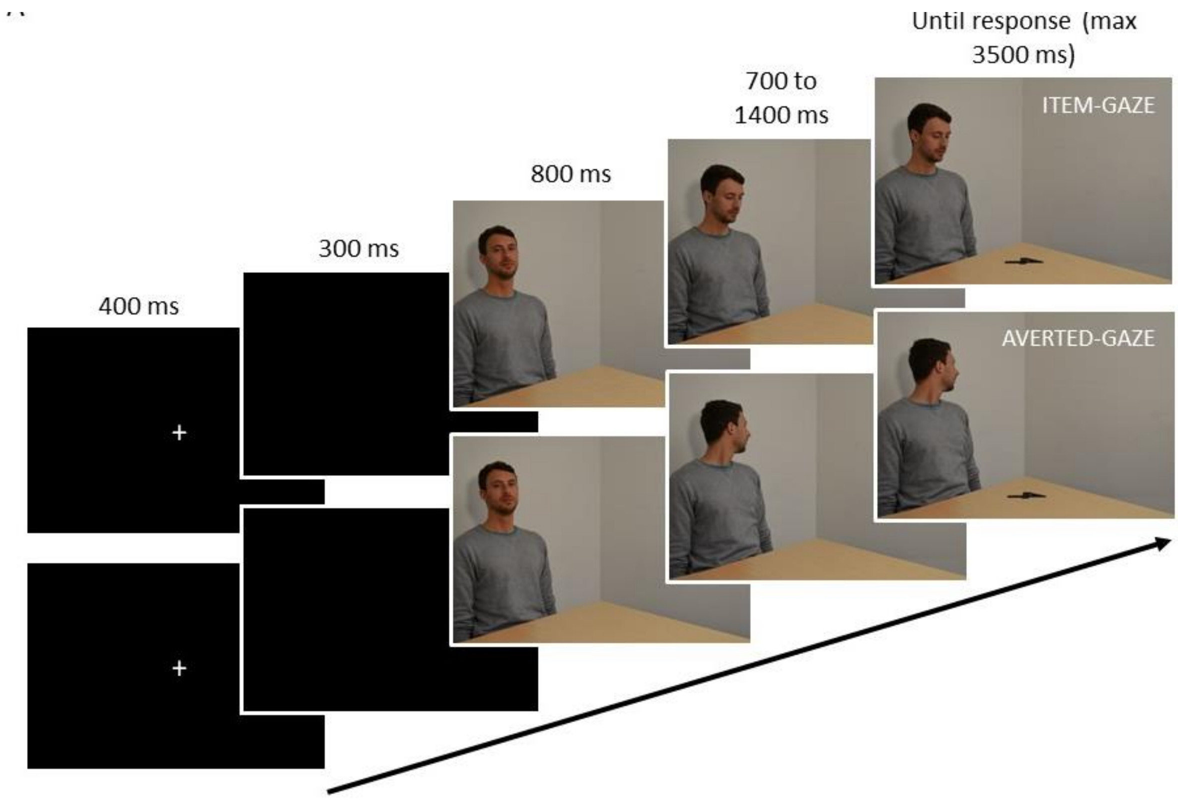

Figure 1. **Example stimuli used in Experiment 1.** Looking direction (Item-gaze or Averted-gaze) varied between trials.

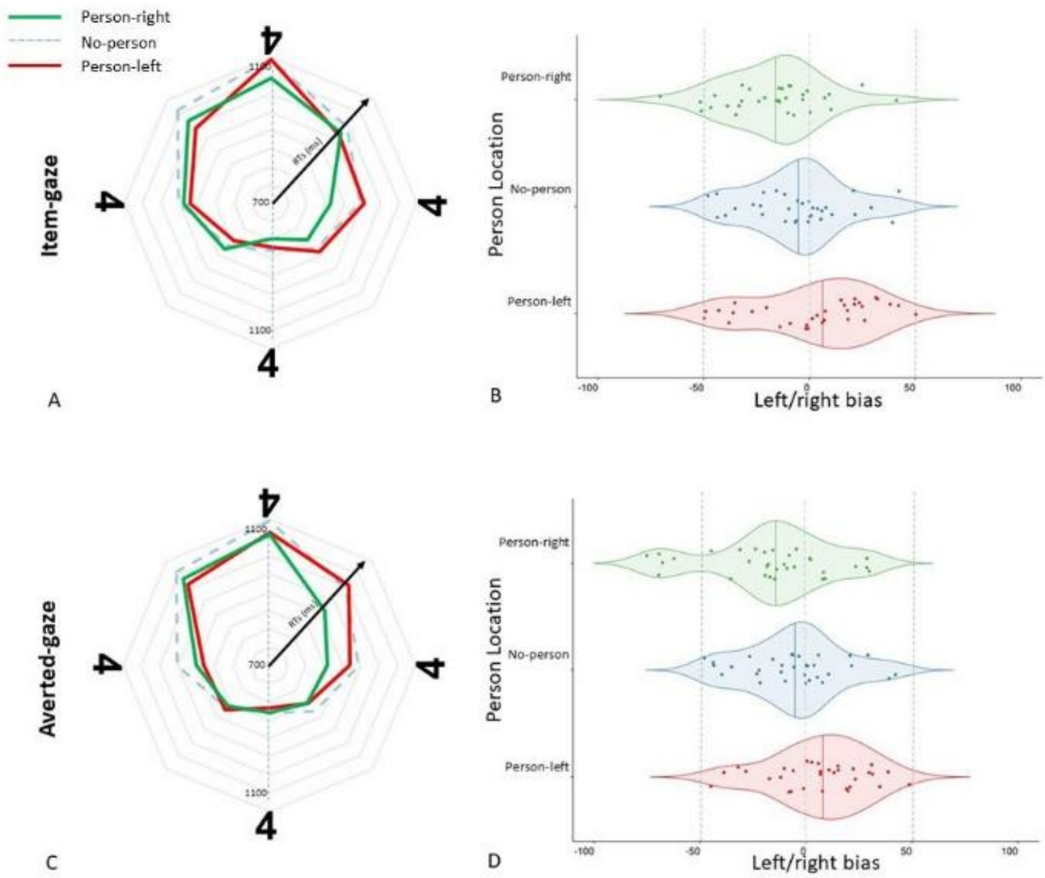

Figure 2. **(A)** Mean recognition times (RTs) for mirror-inverted/canonical judgements for Item-gaze trials when the actor is inserted on the right and the left showing faster recognition times for items oriented upright to the actor. (B) Left/right bias when the inserted person is looking at the table showing faster recognition times for items appearing upright to the actor. (C) Mean recognition times (RTs) for mirror-inverted/canonical judgements for Averted-gaze trials (D) Left/right bias when then inserted person appears to look away from the table (towards to wall).

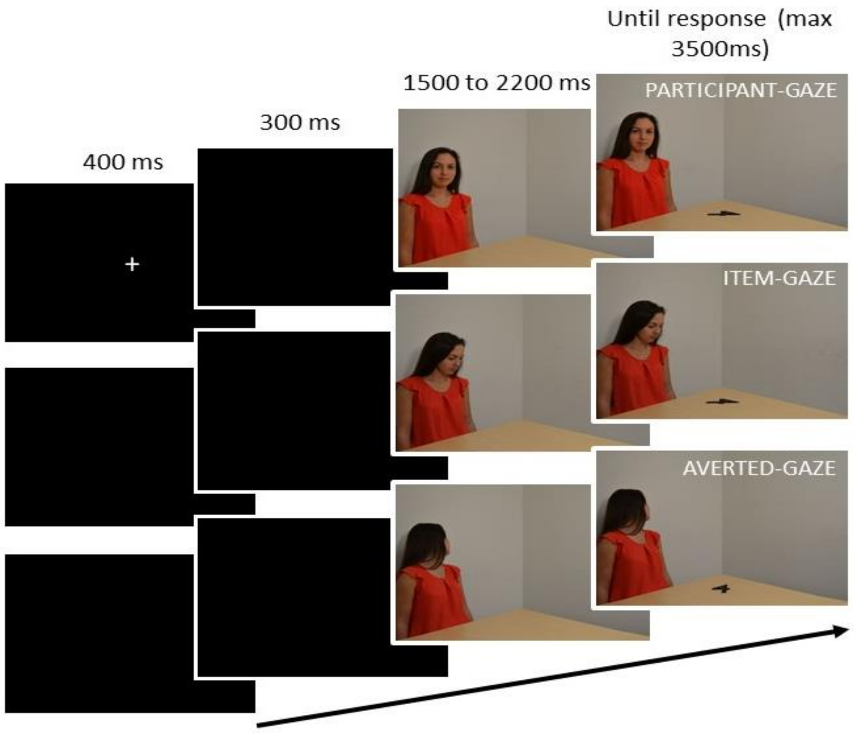

Figure 3. **Example stimuli used in Experiment 2.** Participants viewed only one of the three possible looking directions in a between-subjects design.

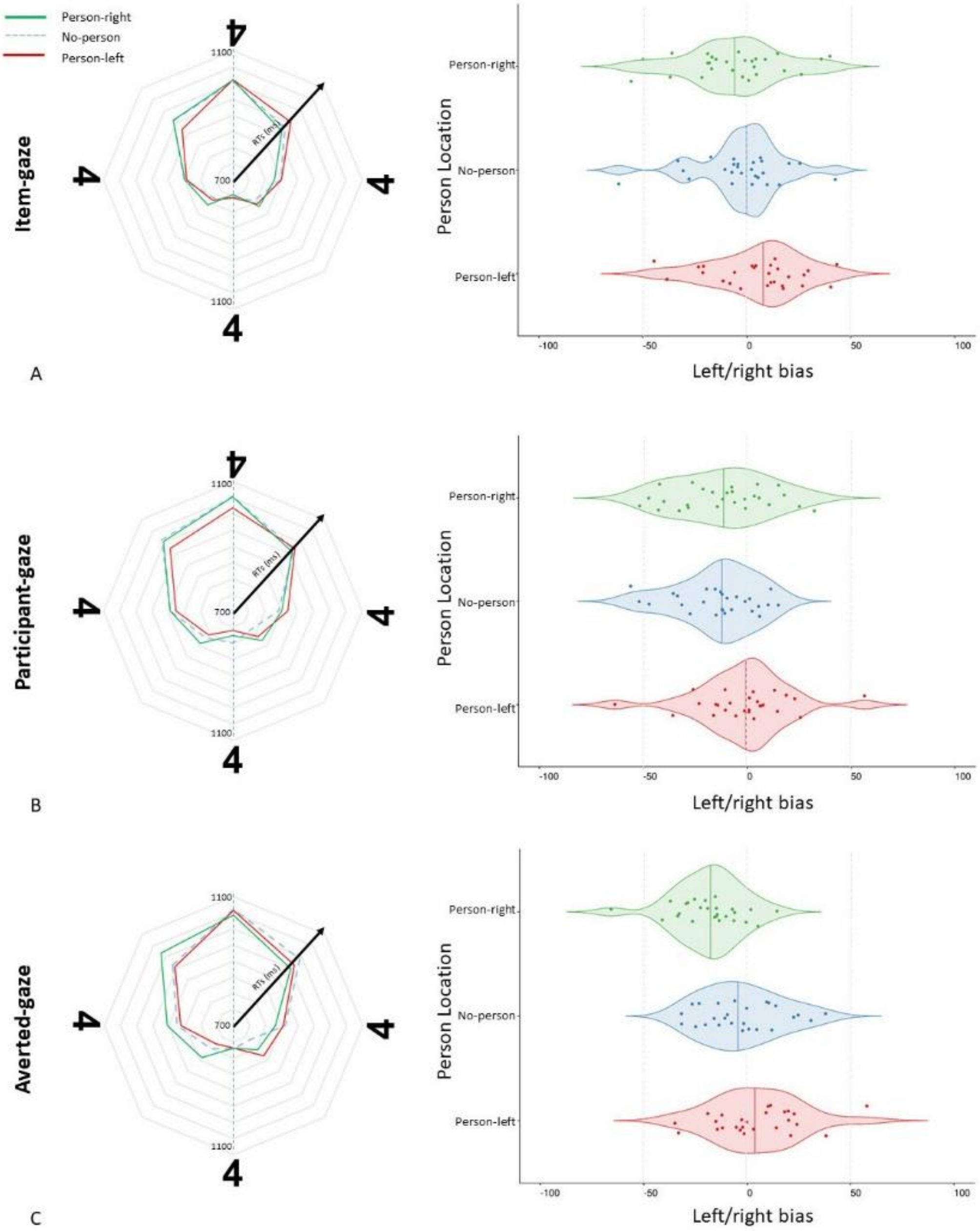

Figure 4. (Left) Mean recognition times (RTs) for mirror-inverted/canonical judgements and (Right) Left/right bias when the inserted person is looking at the table showing faster recognition times for items appearing upright to the actor for (A) Item-gaze, (B) Participant-gaze, and (C) Averted-gaze.