2021

Spontaneous visual perspective-taking: level 2 representations of another's perspective are not related to what they actually see

Aldrich, L.

Spontaneous visual perspective-taking: level 2 representations of another's perspective are not related to what they actually see.

Louis Aldrich

Project Advisor: Prof. Patric Bach and Dr. Giorgio Ganis, School of Psychology, Faculty of Health and Human Sciences, University of Plymouth, Drake Circus, Plymouth, PL4 8AA

Abstract

The ability to understand what another person can see is a fundamental feature of human interaction, commonly referred to as 'visual perspective-taking' (VPT). Traditionally, this ability has been conceptualised along two 'levels', in line with developmental observations (Flavell, Everett, Croft & Flavell, 1981). 'Level-1 VPT' is a simple, automatic, and rapid process enabling understanding of what another individual can see. In contrast, 'Level-2 VPT' is a more complex process requiring conscious control, which enables understanding of how features of the environment look from another's location. This has led to a widely accepted hypothesis that Level-1 VPT is necessary in the formation of Level-2 understanding. Recent research however is beginning to challenge these previous notions. Using a new paradigm which incorporates mental rotation (Shepard and Metzler, 1971) and perceptual interference (Riecke, Cunningham & Bülthoff, 2007) phenomena, Ward, Ganis and Bach (2019) demonstrated that Level-2 representations emerge rapidly, spontaneously and can facilitate (as well as inhibit) one's own perceptual judgements. Building upon these findings, the current experiment implemented barriers into Ward et al's paradigm. By blocking the other person's line of sight to the target stimuli, this manipulation was able to assess how Level-1 judgements affect Level-2 understanding. Surprisingly, findings indicated that Level-2 representations of how the other person saw the target stimuli persisted, even when they could not see the stimuli in reality, and in contrast to participants' own verbal assessments. Implications for the relationship between levels and conceptualization of Level-2 VPT are discussed along with suggestions for further research.

Keywords: level 1 visual perspective-taking, level 2 visual perspective-taking, spatial perspective taking, mental rotation, mental imagery, perceptual simulation, perceptual occlusion, perceptual decision making, social perception, theory of mind, social attention.

Introduction

Visual perspective taking (VPT) concerns the ability to effortlessly understand what another individual sees. Developmentally, VPT appears to emerge in two stages, as 12-months-old infants can determine whether another person can see an object (Lempers, Flavell & Flavell, 1977) but are unable to infer how features of the environment might look from another's location until 3½ - 4 years (Flavell, Everett, Croft & Flavell, 1981). These and other observations (Masangkay, et al, 1974) form the basis of the popular levelled hypothesis, which distinguishes between two 'levels' of VPT ability: understanding what another individual can see from their location (level-1 VPT, henceforth 'L1'), and understanding how features of the environment look from another's location (level-2 VPT, henceforth 'L2'). Due to this apparent developmental order, L1 and L2 abilities are often conceptualized as evolutions of the same cognitive process; differing only in degrees of complexity (with L2 being more complex than L1) and automaticity (where L1 is automatic but L2 requires conscious control). When making perceptual judgements therefore, it is believed that L1 VPT occurs earlier in the perceptual stream than L2 understanding, due to its simple and automatic nature. Following this, it has been hypothesised that a prior L1 judgement might be necessary in the establishment of L2 understanding. Indeed, this appears logical, as one might expect that knowing another individual can see an object (L1) is prerequisite to understanding how that object looks from their location (L2). Despite the fact that new VPT findings continue to be interpreted through this 'levelled' lens however, the nature of the relationship between L1 and L2 and their underlying processes have remained largely unclear.

Towards a better understanding of these underlying mechanisms, a growing body of work has utilized perceptual interference phenomena (Riecke, Cunningham & Bülthoff, 2007) as an investigative tool. In these paradigms the speed and/or accuracy of participants' visual judgements is inhibited by the presence of another individual who would make the same judgments differently. This operates on both levels of VPT. For example, regarding L1 VPT, in the well-established dot perspective task, people find it harder to judge the total number of objects in a scene when in the presence of another individual (hence an 'avatar') who can only see a proportion of the total objects (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). Similarly, in L2 VPT, people find it harder to judge the canonical orientation of objects (Surtees, Samson, & Apperly, 2016) or the spatial relationships between objects and observers (Tversky & Hard, 2009; Kessler & Thomson, 2010) when these objects look different or are differently spatially related from an avatar's perspective respectively. In the case of L2, these findings have been taken to suggest that participant's representations of the avatar's perspective, and participant's own perspective, share a similar perceptual format and therefore interfere with one another. Consequently, VPT has begun to be conceptualized as a form of "perceptual simulation" (Ward, Ganis, & Bach, 2019; Ward, Ganis, McDonough & Bach, 2020).

Recently, strong support for this conceptualization in the case of L2 VPT has been provided by a new experimental paradigm (Ward, Ganis & Bach, 2019), which builds upon classical mental rotation testing (Shepard and Metzler, 1971). In these tests, participants are tasked with identifying whether characters appearing on a table in front of them are mirror inverted or canonically presented ("R" vs. "Я") at different orientations. This produces a reliable finding that identification speed is directly related the degree of angular disparity from the upright position, such that upright characters (0° of rotation) are identified fastest whereas upside-down characters (180° of rotation) are identified slowest. These findings are taken to imply that participants must mentally rotate characters to the upright position in order to make accurate judgements (Cooper, 1975). A mental rotation profile, created by combining the average response time (RT) at each degree of character rotation, provides a clear depiction of this effect. Ward, Ganis and Bach (2019) added an avatar into this mental rotation setup to investigate VPT interference effects. This demonstrated that participants were not only able to identify characters more quickly if they appeared upright to the avatar (despite being rotated away from themselves), but also that they identified characters more slowly if they appeared rotated away from the avatar. These perceptual simulations occurred automatically (even though the avatar was task irrelevant), effectively 'painting' the avatar's perspective upon participants' own perceptual input (Ward, Ganis, & Bach, 2019; Roelfsema & de Lange, 2016; Le Bihan et al, 1993).

These studies have furthered understanding of L2 VPT but are unable to tell us anything about the relationship between L1 and L2 in their current form, as the avatar's line of sight was always unobstructed (e.g., L1 judgements are not manipulated). One effective way to begin investigation into this relationship would be to physically block the avatar's line of sight to the character, enabling direct observation of how participants' tendency to draw from the avatar's perspective (L2 understanding) is affected by the avatar's view of the character (L1 judgment). Indeed, blocking paradigms have been a staple feature of L1 VPT research since its conception. They are simply defined as instances where the participant has line of sight to a target object but the avatar's line of sight to this same object is occluded by an opaque barrier and have been implemented in a variety of different ways. For example, in Flavell, Everett, Croft & Flavell's (1981) study, participants viewed two-sided picture cards (e.g., on one face there was a picture of a dog and on the other a cat) that were held vertically between them and experimenter and tasked with stating what the experimenter saw as an indication of L1 VPT. This can be construed as an early form of blocking, as the participant cannot see the side of the card the experimenter sees when making judgements. Keysar, Barr, Balin and Brauner (2000) also used blocking in their 'director-task', using a shelf of 'cells' placed between the participant and an avatar. A proportion of these cells were missing back panels, such that objects placed in these locations were visible to both the avatar and the participant, whereas objects in cells with back panels were only visible to the participant. Participant's ability to distinguish what the avatar can/cannot see (L1 VPT) was central to their performance when following the instruction of the avatar to move certain objects. Furthermore, Cole, Atkinson, Le and Smith (2016) recently implemented blocking in Samson et al's (2010) dot-perspective task, by using two barriers which could be either completely opaque or could possess a transparent screen allowing the avatar visual access to the target stimuli.

However, these past blocking paradigms are not capable of investigating the relationship between levels as L2 representations are not assessed. Altering these paradigms to involve both levels would require fundamental changes likely to undermine their capability. In addition, there are also some issues inherent in these existing blocking paradigms. For example, participants could be using cognitive heuristics instead of real perspective simulations (Santiesteban, Shah, White, Bird & Heyes, 2015), enabled by the consistency of the participant's and avatar's location in relation to the target stimulus during testing. Such heuristics could be as simple as knowing that the experimenter sees the opposite of what you see in the card-task (e.g., if I see a cat then they see a dog) or knowing that objects in a cell missing a back panel can always be seen by the avatar in the director-task. Another issue is the unknown influence of memorial factors in performance. For example, the card task is heavily reliant upon memory of the stimuli, as what the avatar sees is perceptually unavailable to the participant at the time of judgement. This is also somewhat true in the director task, as the participant and avatar view opposite faces of the same objects, thus participants are required to recall the opposite face when attempting to understand what the avatar sees (although memory is perhaps strongly cued by the visible side). Although the dot-perspective blocking manipulation avoids the previously mentioned issues (as the avatar's perspective is variable and the stimuli are always completely visible), it has been argued that interference effects in this task could result from attentional gaze cueing rather than perceptual representation (Langton, 2018; Cole, Atkinson, Le & Smith, 2016). Furthermore, as scenarios are presented two-dimensionally on a screen, it is difficult to ensure participants actually interpret the barrier as blocking the avatar's line of sight to the target. This is because participants can only use weak overlapping and retinal image size cues to interpret the 3D organisation of objects, rather than strong binocular parallax cues. Consequently, these past blocking paradigms are perhaps not an adequate basis upon which to further investigate the L1 and L2 VPT relationship.

Fortunately, however, implementing barriers in Ward et al's (2019) mental rotation paradigm would enable a direct observation of how L1 judgements affect L2 understanding, whilst avoiding these issues present in past and current blocking paradigms. For example, participants are unlikely to use cognitive heuristics, as they are exposed to a different combination of character, orientation, and avatar location on each trial. This paradigm would also isolate the effects of perceptual simulation from memorial influences, as all features of the character are always visible to both the avatar and participant (only the rotation of these features changes). Additionally, evidence of facilitation or inhibition would not be confounded by gaze cuing effects, as the target stimulus is always present in the same location, and the avatar always looks at it. The mental rotation profiles would also illustrate participants' use of the avatar's perspective in terms of character identification times, providing direct evidence of L2 representation and how it is affected by blocking.

Although a barrier manipulation in mental rotation has not yet been produced in the literature, research by Ward et al. (2020) recently came close by varying whether the avatar in the scene was looking at the character, as 'pseudo-blocking' is achieved when the avatar is looking away. Interestingly, they found that averted gaze had no effect on L2 facilitation and inhibition effects, leading them to conclude that L2 VPT is not sensitive to the avatars' sight (L1 VPT) but rather to their location in space. This contrasts the traditional levelled conception that L1 judgements are required for L2 understanding, instead suggesting that L2 VPT processes can operate in isolation. It would however be premature to accept these new findings on the basis of gaze alone, as participants might be representing what the avatar could see in principle from their position, rather than where the avatar was looking in a specific moment. Indeed, people frequently alter where they are looking in real-life scenarios, and only very small, imagined adjustments on the avatar's behalf would be required to enable line of sight to the character within this gaze manipulation.

Building upon these findings therefore, the following two experiments introduce barriers into Ward et al's (2019) mental rotation paradigm, in which participants discern the mirror/canonical presentation of characters at different orientations, in the presence of an avatar who offers a better or worse perspective of said character. In doing so, the author aimed to gain insight into two questions which have risen to importance: (a) does L2 VPT build upon L1 VPT in such a way that it should be considered an evolution of the L1 process (as suggested by the developmental theory), or do these 'levels' actually refer to separate cognitive processes, and (b) how can we best conceptualise the L2 VPT process?

Experiment 1 manipulated barriers as task irrelevant features, by blocking the avatar's line of sight to the character in 50% of trials and comparing the degree to which participants' performance was influenced by the presence of the avatar between blocked and unblocked trials. If simulations of another's visual perspective are dependent upon their line of sight, like the developmental theory suggests, participants should draw from the avatar's perspective less if the relevant object cannot be seen by the avatar in reality. Thus, participants should not benefit from perceptual facilitation effects (decreases in RT when the avatar's location offers a better view of the character than the participants own) nor not suffer perceptual interference effects (increases RT when the avatar's location offers a worse view of the character than the participants own). Alternatively, no difference between blocked and unblocked trials would support the notion that L2 representations are not analogous to one's own perceptual input. In Experiment 2, barriers were made task relevant by asking participants to verbally dictate at the start of each trial whether or not the avatar could see a white cross, which was positioned in the same location the character would appear. With this manipulation, we aim not only to make the barrier more salient to participants, but also to further tease apart the relationship between L1 and L2, by testing the possibility that making a L1 judgement might increase the salience of the following L2 representation.

Methodology

Methods for each experiment overlapped considerably, thus Experiment 1 and Experiment 2 are presented in a combined methods section.

Participants

Forty-four and thirty participants were recruited via the University of Plymouth participation pool in Experiment 1 and 2 respectively, following approval by the university Ethics Committee. All participants were adults and gave written informed consent (Appendix A). Those who chose to participate were compensated with course credit. Four participants in Experiment 1 and fourteen participants in Experiment 2, with accuracy scores below 80% across all conditions, were excluded from final analysis (as in Ward et al., 2019). A single participant in Experiment 2 was removed as an outlier. Hence data from forty participants in Experiment 1 (31 female, age M=21.13, SD=5.28) and sixteen participants in Experiment 2 (14 female, age M=19.50, SD=1.86) was used in final analysis.

Apparatus

Both experiments were conducted in campus labs at the University of Plymouth. Experiment 1 and 2 were displayed on a 19" LED computer monitor (Resolution: 1900x1200; Refresh rate: 60Hz) using Presentation® software (Version 18.0, www.neurobs.com) and PsychoPy software (Version 3.0., www.psychopy.org, Peirce, 2007) respectively. Participants responded to stimuli using a standard keyboard, with the three keys 'UP', 'DOWN', and 'SPACE'. Green and red stickers were placed on the 'UP' and 'DOWN' keys for clarity. In Experiment 2, Participants made their verbal responses using a headset. In both experiments an example sheet (Appendix B) was provided prior to experimentation.

Design, Stimuli and Procedure

After completion of the consent form, participants were given an example sheet (depicting the characters that would appear on the screen) and completed a short instructional procedure. This was followed by 10 practice trials with feedback. If a participant scored under 80% in these 10 trials, then the same 10 trial sequence was repeated, up to a maximum of 5 times. If a participant failed to score 80% or higher 5 times over, they were excluded from the experiment (no participants were excluded under these circumstances). Upon achieving a score of 80% or over on the 10 practice trials, participants began the experiment proper. Experiment 1 consisted of 384 trials, each composed of two frames (fig.1), whereas Experiment 2 consisted of 192 trials, each consisting of five frames (fig.2), plus another feedback frame (a black screen with centred red text: "Incorrect") which appeared only if an incorrect answer or no answer was entered.

In Experiment 1, the first frame depicted a scene for 500ms (a view onto a corner of a square table in a grey room), containing an avatar looking downwards towards the centre of the table (see Figure 1).

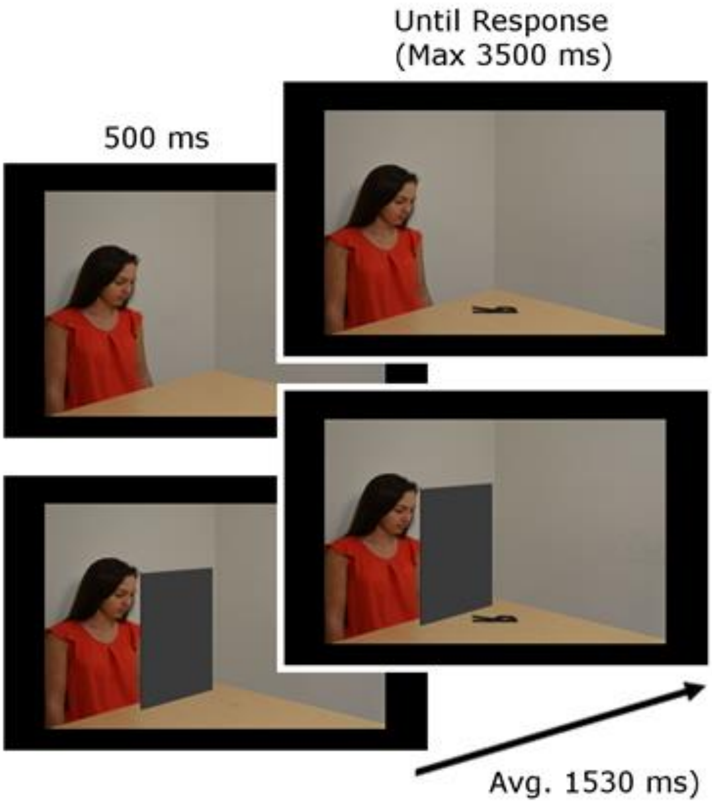

Figure 1: Example stimuli used in Experiment 1. Blocking (Avatar LoS to character uninterrupted or interrupted) varied between trials.

The Avatar was either male or female and could be seated either to the left (Avatar-left trials) or right (Avatar-right trials) of the table, so that they were at roughly 270° or 90° angle to the centre of the table, relative to the participant. Both these variables were balanced across all 384 trails (in 50% of trials the avatar was male and 50% of trials were Avatar-left trials). A barrier occluded the avatar's line of sight to the centre of the table in 50% of trials.

The second frame was identical to the first in all aspects (sex, position and barrier presence) as to maintain scene continuity, aside from the appearance of a character in the centre of the table where the avatar is looking. Said character could be either the number '4', or letters 'P' or 'R', which would appear either canonically (e.g., "R") or in mirror inverted form (e.g., "Я") at one of eight orientations (0°, 45°, 90°, 135°, 180°, 225°, 270° or 315°), totalling 48 possible items. From 0° (appearing upright to the participant) a character's orientation increases in a clockwise fashion. The character always appeared in the same location in the centre of the table (as rotation of the character occurred around the character's centre point), such that the avatar's line of sight to the character was perpendicular to the line of sight of the participant, when either seated to the left or right of the table (90° or 270° respectively, relative to the participants position).

In each trial, participants task was to indicate, by pressing either the green or red key using their dominant hand, whether the presented character was in its canonical or mirror-inverted form ("R" vs "Я"). Speed of response was measured between the appearance of the character and the participant's key response, up to a maximum duration of 3500ms.

In Experiment 2, to reduce the total time of the experiment, the sex of the avatar was randomised for each trial, so that the total number of trials was 192 instead of 384. Being a crucial variable however, the position of the avatar remained balanced across all 192 trails (50% of trials were Avatar-left trials). Again, a barrier occluded the avatar's line of sight to the centre of the table in 50% of trials. A white fixation cross was also present in the middle of the table in all slides aside from when it was replaced by the appearance of the character.

The first frame presented was the same as in Experiment 1, however it lasted for 1000ms and depicted the avatar looking outwards towards the participant (see Figure 2). The second frame, lasting for 500ms, was identical to the first in all aspects (sex, position and barrier presence) but now showed the avatar looking at the cross on the table. This was proceeded by a third frame, which displayed the message "Can the person see the White cross" above the scene (identical in all aspects to the last slide) and "(Press Space to lock in your answer)" below, as well as a small recording icon which appeared in the top left of the screen. Participants made their answers by stating either 'yes' or 'no' out loud into the microphone, depending on whether a barrier was between the avatar and cross. Participants then pressed the space bar to 'lock in' their answer and stop the recording, after which the text and icon disappeared and there was a 1500ms pause before a character appeared. Unknown to participants, no voice data was recorded, as the ease of the task and presence of experimenter made it so that participants interacted with the dictation task properly.

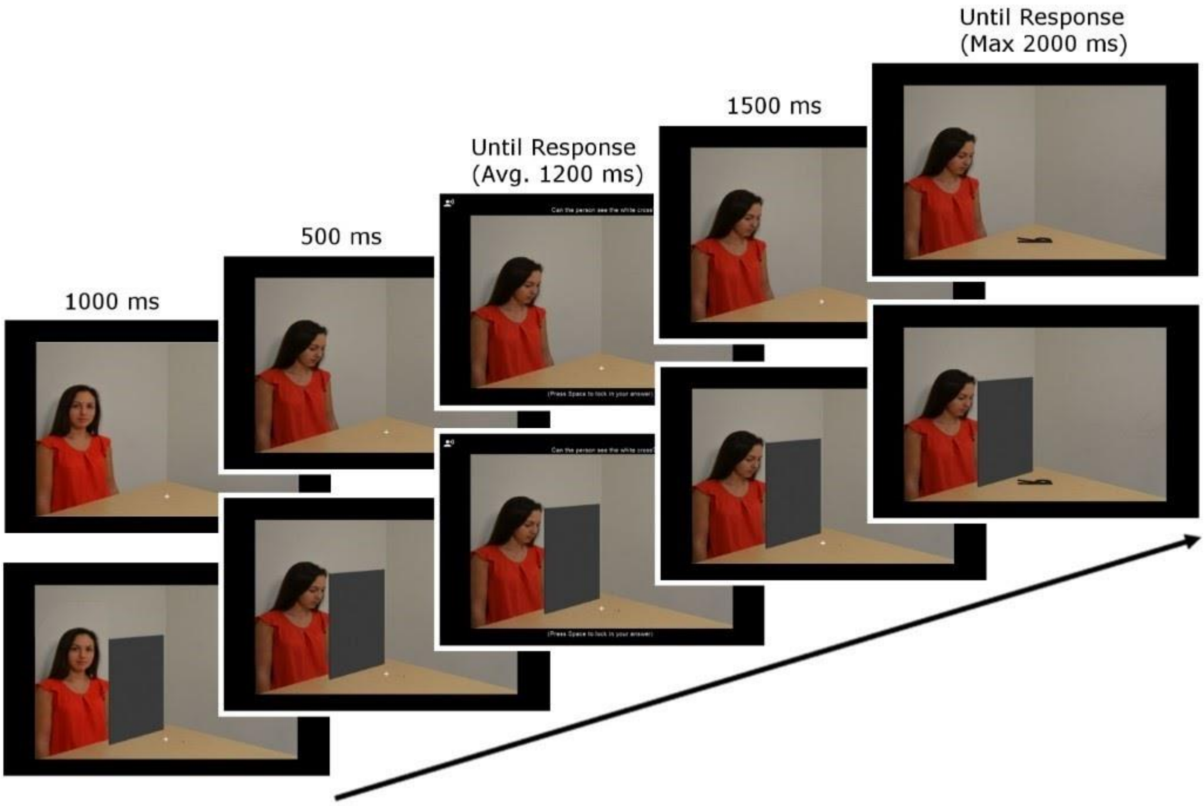

Figure 2: Example stimuli used in Experiment 2. Blocking (Avatar LoS to character uninterrupted or interrupted) varied between trials. Participants required to make a verbal level 1 judgment proceeding each character judgement.

In the fifth frame the white cross disappeared and was replaced by a character, which like in Experiment 1, could be one of three alphanumeric characters (4, P, or R), appearing at one of eight orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) and could be either presented canonically or mirror-inverted. The location of this character was identical to in Experiment 1. Participants made the same canonical or mirror-inverted judgements using the red and green keys within 2000ms of the character's appearance. If participants either answered incorrectly, or after the 2000ms duration had elapsed, then a black screen with the red text "Incorrect" or "Respond Faster" would appear respectively for 40ms. If a correct answer was entered, then no such screen was shown, and the participant proceeded to the next trial. Participants were debriefed after completion of all trails (see Appendix C).

Results

Data processing and analysis was conducted using Microsoft Excel (version 2101) and JASP (Version 0.14.1, 2018). Recognition times for each character orientation (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315), depending on the avatar's location (Avatar-left, Avatar-right) and barrier presence (Blocked, Unblocked) were the dependent measures. Trials with recognition times longer than 2000ms or shorter than 150ms were excluded from analysis, along with data from participants whose rate of identification error was above 20% (Exp.1, N=4; Exp.2, N=14). One participant in Experiment 2 was removed as an outlier. In an identical manner to Ward et al. (2019), the main effects for mental rotation and perspective taking were derived from two statistically independent summary measures.

Mental Rotation

The first of these summary measures compared the recognition speed for characters oriented away from participants (225°, 180°, 135°) to characters oriented towards them (315°, 0°, 45°), known as the Towards/Away bias. The relative towards/away contribution of different orientations was extracted by treating each angle of rotation as a vector, possessing both a 'latitudinal' (towards/away = 0°/180°) and 'longitudinal' (left/right = 90°/270°) component. In this way, characters at 0° and 180° of rotation contribute maximally to the towards/away measure (possessing a wholly latitudinal component), whereas characters at 90° and 270° do not contribute (possessing a wholly longitudinal component). Calculation of contribution was achieved in the same way as Ward et al. (2019), by multiplying recognition time by the negative of the cosine for each orientation angle, for each participant and condition (Avatar-left Blocked, Avatar-left Unblocked, Avatar-right Blocked, and Avatar-right Unblocked). Consequently, the average of these values is indicative of mental rotation in the case of a positive result (faster recognition for characters oriented towards rather than away from the participant), thus confirming Shepard & Meltzer's (1971) traditional mental rotation effect. Overall (across conditions) towards/away bias was compared in both experiments using a simple t-test against zero, illustrating the difference in towards/away identification time in ms. An alpha level of .05 was utilized. A Shapiro-Wilk test confirmed a normal distribution in both Experiment 1, $W(39)>0.962$, $p>.193$, and 2, $W(14)>0.897$, $p=>.087$. As expected, bias was significantly positive in both Experiment 1 $M=56.21$; SD6.35, p2.15, $p<.001$ and 2, $M= 52.54$; $SD<77.05$, $t(14)>2.15$, $p<.05$. for all conditions (as demonstrated in Figure 3), confirming the mental rotation effect (character format identification is quicker the more they are oriented towards rather than away from participants).

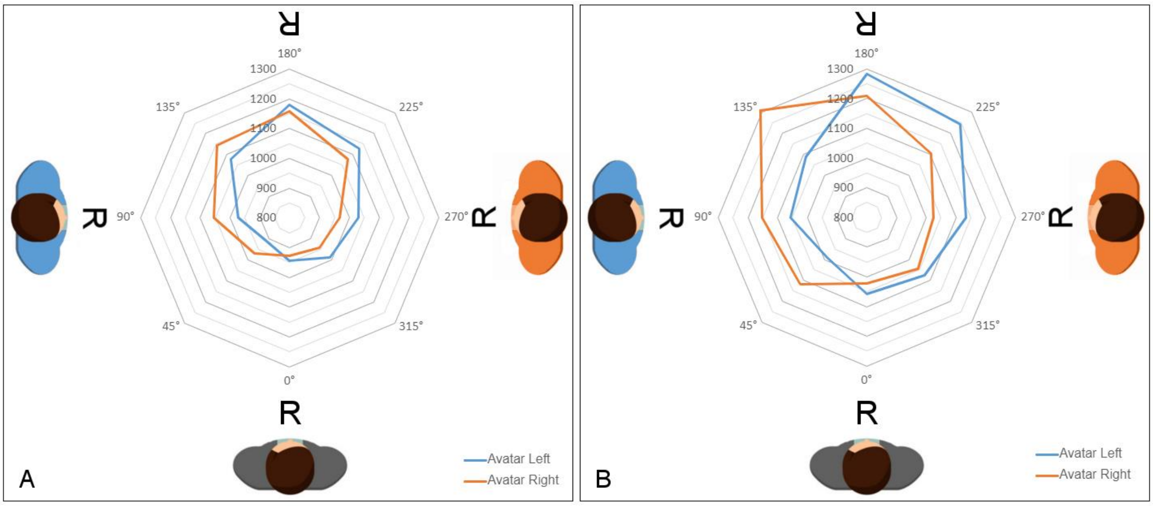

Figure 3: Mean recognition times (ms) for mirror-inverted/canonical judgements by degree of character rotation in Experiment 1 (A) and 2 (B). Data indicates faster recognition times in instances when the character is oriented upright to the avatar, either when they are positioned to the right (orange) or the left (blue) of the participant (grey).

Perspective Taking

The second summary measure compares the recognition speed for characters oriented towards the left (45°, 90°, 135°) and right (225°, 270°, 315°) of participants, known as the Left/Right bias. The relative contributions of different orientations to the Left/Right measure is orthogonal to the Towards/Away measure, such that characters at 90° and 270° of rotation contribute maximally (possessing a wholly latitudinal component), whereas characters at 0° and 180° do not contribute (possessing a wholly longitudinal component). The Calculation of each angle's contribution to the Left/Right measure was therefore analogous to the Towards/Away measure, being the recognition time multiplied by the sine of the orientation angle, for each participant in each condition. Hence rightward facing characters (45°, 90°, 135°) contribute negatively and leftward facing characters (225°, 270°, 315°) contribute positively to the final summary score. The difference between these values for Avatar-left and Avatar-right conditions can therefore index the VPT effect. For example, if participants are faster at identifying the character when it appears more upright to the avatar, then we should find significantly positive Left/Right summary scores in overall Avatar-Left trials and a significantly negative Left/Right summary scores in overall Avatar-Right trials. Alternatively, equally quick character format identification described by a zero value (non-significant) would suggest no VPT effect. From these calculations we are able to detect whether VPT is affected by occluding the avatar's line of sight to the character using a barrier, by comparing average Left/Right measures between blocked and unblocked trials. We therefore calculated each participants' Left/Right bias for each condition, with the resulting left/right-bias summary scores being entered into a 2 x 2 repeated measures ANOVA, with within-subjects factors of Avatar Location (Avatar-left, Avatar-right) and Occlusion (Blocked, Unblocked).

This revealed strong evidence for the main effect of Avatar Location in both Experiment 1 $F(1,39)=26.629$, $p=<.001$, $\eta p2=.406$, and 2, $F(1,14)=15.592$, $p=<.001$, $\eta p2=.527$ (as demonstrated in Figure 4). Fully replicating Ward et al's (2019) VPT findings that participants recognized leftwards rotated characters more quickly than rightwards oriented ones when an avatar was sitting on the left, but they recognized rightwards rotated characters more quickly than leftwards rotated ones when an avatar was present on the right (See Figures 3 and 4). Conversely, no significant effect was found for Occlusion, in either Experiment 1, $F(1,39)=1.317$, $p=.258$, $\eta p2=.033$, or 2, $F(1,14)=0.410$, $p=.532$, $\eta p2=.028$. Furthermore, no interaction between Avatar Location and Occlusion was detected in either Experiment 1, $F(1,39)=.003$, $p=.959$, $\eta p2=<.001$, or 2, $F(1,14)=0.698$, $p=.417$, $\eta p2=.048$. Error rates followed the same pattern as RTs (e.g., participants made more errors when identifying character's oriented away from them) but were not significantly different between conditions. Therefore, these findings suggest VPT is dependent upon the avatar's location but not their line of sight to the character. A further paired samples t-test for the Left/Right bias between blocked and unblocked conditions in both Experiment 1 (unblocked: $t(39)= 4.704$, $p=<.001$, $d=.744$, and blocked: $t(39)= 3.693$, $p=<.001$, $d=.584$) and Experiment 2 (unblocked: $t(14)= 3.390$, $p=.004$, $d=.875$, and blocked: $t(14)= 3.013$, $p=.009$, $d=.778$), confirmed that the VPT effect was reliably present both when the avatar could see and could not see the character.

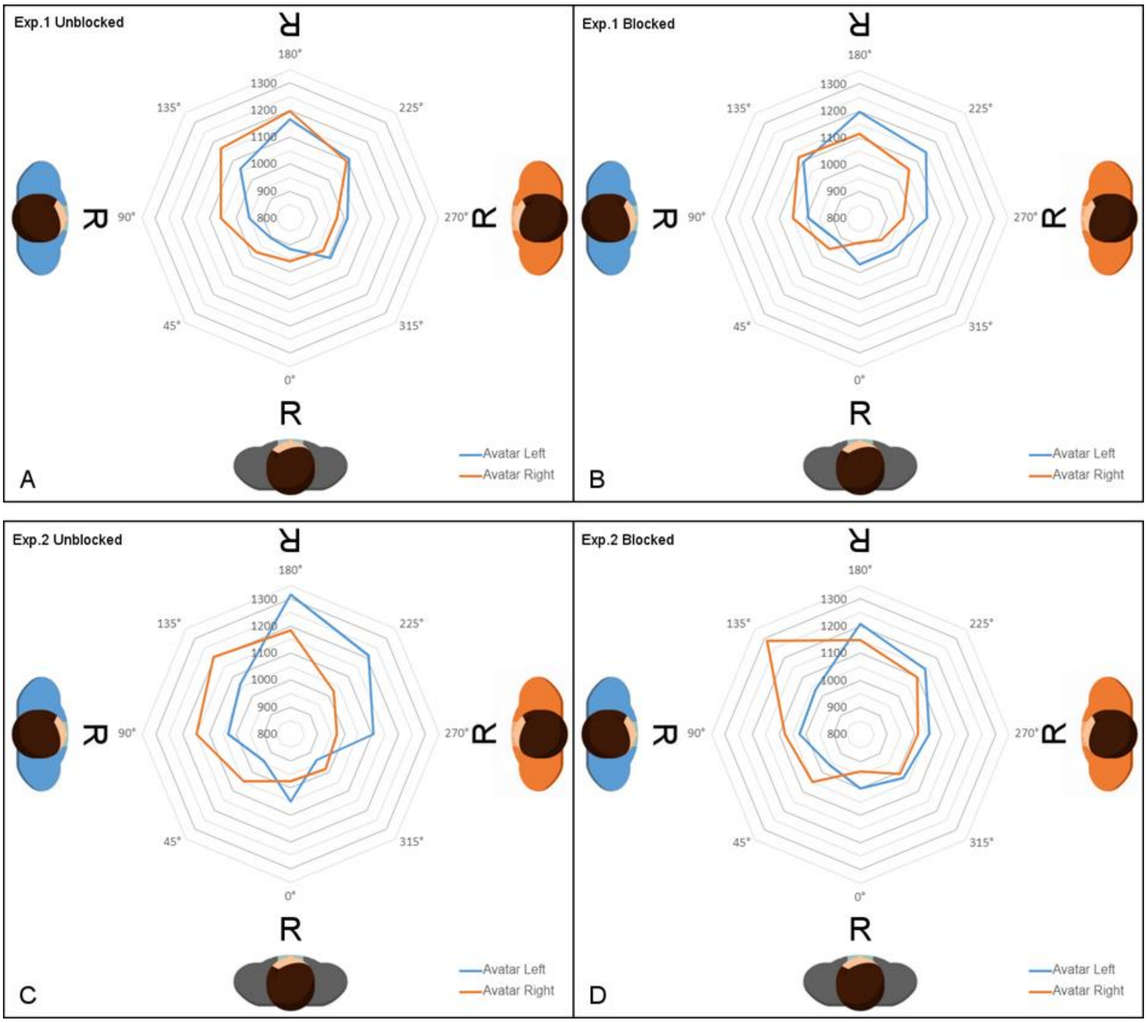

Figure 4: Mean recognition times (ms) for mirror-inverted/canonical judgements by degree of character rotation in Experiment 1 (A and B) and 2 (C and D), when the view of the avatar is unobstructed (A and C) and obstructed (B and D) by the barrier. Data indicates no difference in recognition times when avatar's view is obstructed, either when they are positioned to the right (orange) or the left (blue) of the participant (grey).

Discussion

In two experiments, I manipulated whether an avatar could see an alphanumeric character by blocking their line of sight (LoS) using an opaque barrier, in a version of Ward et al's (2019) mental rotation paradigm. Participants reported whether characters in various orientations were presented in their normal or mirror inverted form, in the presence of an avatar for whom these characters appeared either more upright or less upright than the participant, and whose line of sight to the character was either blocked or not blocked. Overall data from both experiments replicated the classic mental rotation effect: that participants are slower to identify the mirrored/canonical nature of a character the more it is oriented away from them (Shepard & Metzler, 1971). Data from avatar left and avatar right unblocked trials in both experiments successfully replicated Ward et al's (2019) original perspective taking finding: that the presence of an avatar automatically facilitates/inhibits (dependent upon the avatar's location) the speed of participants' character judgements, despite being task irrelevant. In other words, participants identified rotated away characters more quickly when they appeared upright to another person in the scene, but more slowly when the characters were rotated even further away from this other person. Interestingly, this finding was also replicated in the data from blocked trials in both experiments, such that VPT effects were similar across both blocked and unblocked conditions, even when participants made a verbal L1 judgement ("Can the avatar see the character location?") prior to L2 measurement. Consequently, participant's automatic tendency to draw from the avatar's perspective occurred irrespective of whether the avatar could see the character from their location in reality, and regardless of whether the barriers were task relevant. This finding consequently has implications for both the theoretical relationship between L1 and L2 VPT and our conceptualization of L2 VPT itself.

First, the findings here do not fit with the claims of the traditional levelled hypothesis. This is because if L1 and L2 were indeed different 'levels' of the same VPT process (such that L1 is a prerequisite for L2 understanding), we would expect that an L1 judgement should exert some effect upon the contents of the L2 representation (such as reducing the salience of objects which cannot be seen from the other individual's location, or not representing these objects at all). However, the current experiment identified neither of these interactions. Instead, establishing whether another can see an object (L1) is not necessary in the formation of a conception of how they see the object (L2), and the information obtained from the L1 process (e.g., what another person can see) had no bearing upon the formed L2 representation (e.g., occluded objects are represented), even when participants made the L1 judgments explicitly. These results therefore suggest little integration between L1 and L2. Consequently, it appears unlikely that L1 and L2 VPT processes are fundamentally related in a levelled sense.

Second, the endurance of L2 VPT in these circumstances also has implications for our conception of the process itself. Imagine, for example, if we had observed that the presence of the barrier removed perceptual facilitation/inhibition effects. In this case we might conclude that L2 representations are analogous to our own perception, simulating what one might literally see if seated in the same location as the avatar, as objects which are occluded from this location in reality would similarly be occluded in L2 representation. Although there is a general and understandable reluctance to confirm the nature of VPT, this indeed appears to be the view of some authors, who have discussed L2 VPT as a "snapshot" of another's perspective (Moll and Kadipasaoglu, 2013) or a visual computation of "what other people can see" (Samson et al., 2010; Furlanetto et al, 2016). In contrast to this conception however, the current results suggest that visual occlusion does not factor in participant's formed L2 VPT representations at all. This is supported by observation of mental rotation profiles, which indicated that participant's representations of the avatar's location (indicated by the point of greatest facilitation) were consistent between blocked and unblocked conditions. Hence, retained VPT effects in blocked trials were unlikely to result from participants' imagining the avatar looking around the barrier. Instead, and in contrast to participants' own appraisal of the scene (e.g., that the avatar's line of sight to the character was occluded), the occlusion of the character was completely ignored in L2 representation. Explanation of this finding consequently requires a new conceptualization of L2 VPT, one in which the formed representation is not analogous to one's own perceptual input.

To this effect, an alternative idea is that L2 VPT is perhaps rooted in more general visuo-spatial and navigational mechanisms (Ward, Ganis, McDonough & Bach, 2020), which spontaneously represent the relative positions (Tversky & Hard, 2009) and orientations (Elekes, Varga, & Király, 2016, 2017) of relevant features of the environment (e.g., the avatar, the barrier, the character and oneself). Indeed, there is correlation between VPT and navigational ability (Gunalp, Moossaian, & Hegarty, 2019; Kozhevnikov, Motes, Rasch, & Blajenkova, 2006) and it is well known that, during navigation, people can dynamically transform the alter-centric representation of space into egocentric representations (Kovács, Téglás & Endress, 2010; Capozzi, Cavallo, Furlanetto & Becchio, 2014; Nielsen, Slade, Levy & Holmes, 2015). Following from this, it has recently been proposed that we perhaps possess a natural disposition towards these altercentic states (Southgate, 2020). These navigational transformations appear to be supported by specific hippocampal-posterior parietal networks such as the Temporo-parietal junction (TPJ) (Committeri, Piccardi, Galati, & Guariglia, 2015; Wang, Callaghan, Gooding-Williams, McAllister & Kessler, 2016). Interestingly, the TPJ is also implicated in the theory of mind abilities such as false belief reasoning (Schurz, Aichhorn, Martin, & Perner, 2013) and self-other distinction (Quesque & Brass, 2019; Lamm, Bukowski, & Silani, 2016). Disruption leads to distortions of one's experience of space, inducing out of body experiences in the most extreme case (Blanke, Landis, Ortigue & Seeck 2002; Blanke & Arzy, 2005; Tsakiris, Costantini & Haggard, 2008; Lopez & Elziere, 2018).

Such a conception would explain these findings, as the represented content would be spatially rather than visually coded, hence visual properties such as occlusion would not feature. Within this conception, VPT effects could arise out of spatial rather than visual interference. For example, the location of other individuals could act as a highly salient cue, biasing one's representation of the environment (objects and their rotations) and navigation through it. Consequently, L2 VPT could be thought of as 'quasi-perceptual' in the sense that the spatial reasoning employed is not itself visually/pictorially coded, but the result is an understanding of an alternative visual form of an object. Interestingly, if one were to accept this spatial conception of L2 VPT, they might draw an inverse conclusion about the relationship between levels. This is because L1 judgements require one to determine whether a line of sight (between another's eyes and an object) in three-dimensional space passes through an opaque entity. In order to achieve this therefore, one must first possess an accurate understanding of the relationship between objects and observers. Given that L2 VPT (as measured by our paradigm) quickly and automatically represents this spatial information about the environment, one might conclude that L1 line of sight judgements can only be accurate when placed in the context of wider L2 spatial understanding. Hence, L1 line of sight judgements might in fact require a prior L2 representation.

Conclusion

In conclusion, whilst the true nature of the processes underlying L2 representation is beyond the scope of this experiment, the present findings present convincing evidence against L2 VPT involving a literal representation of another's viewpoint. This challenges the traditional levelled conception that L2 VPT is related to L1 VPT. Further research is needed to confirm that the barrier was indeed represented in a location which occludes the avatar's line of sight. This is because two-dimensional screen testing is monocular and thus provides only weak locational cues (such as the retinal size of the barrier or the fact it slightly overlaps the avatar). Consequently, it is possible that participants could have distorted the location of the barrier such that it allowed the avatar line of sight in blocked trials, maintaining VPT effects. A future three-dimensional set-up in virtual reality (providing strong binocular locational cues), varying barrier size and location (to avoid the development of cognitive heuristics), would be sufficient to address these issues. As a final point, it has recently been suggested that a theory of VPT should be “unambiguously concerned with what another person can see'' (Cole, Millett, Samuel & Eacott, 2020). If one follows this definition, the findings presented here would not be considered VPT phenomena. Despite this, ‘VPT’ has been used in this experiment, in line with the terminology used in the paradigm upon which the present manipulation is based.

Future work

The author recognises however the degree to which the current findings overlap with similar ‘spatial perspective-taking’ notions in the field (Surtees, Apperly & Samson, 2013), and thus accepts that a semantic change might be necessary in light of future research.

Acknowledgements

I would like to thank my two project supervisors Dr Giorgio Ganis and Dr Patric Bach, and the members of Dr Bach’s lab, for all their help, support and guidance throughout this project. I have really valued the time I got to spend with such a professional and friendly research team and wish you all the best with your continuing projects and achievements.

References

Blanke, O., & Arzy, S. (2005). The out-of-body experience: disturbed self-processing at the temporo-parietal junction. *The Neuroscientist*, *11*(1), 16-24.

Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Stimulating illusory own-body perceptions. *Nature*, *419*(6904), 269-270.

Capozzi, F., Cavallo, A., Furlanetto, T., & Becchio, C. (2014). Altercentric intrusions from multiple perspectives: Beyond dyads. *PLoS One*, *9*(12), e114210.

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

Cole, G. G., Millett, A. C., Samuel, S., & Eacott, M. J. (2020). Perspective-taking: in search of a theory. *Vision*, *4*(2), 30.

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.

Elekes, F., Varga, M., & Király, I. (2017). Level-2 perspectives computed quickly and spontaneously: Evidence from eight-to 9.5-year-old children. *British Journal of Developmental Psychology*, *35*(4), 609-622.

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.

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.

Gunalp, P., Moossaian, T., & Hegarty, M. (2019). Spatial perspective taking: Effects of social, directional, and interactive cues. *Memory & cognition*, *47*(5), 1031-1043.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32-38.

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

Kovács, Á. M., Téglás, E., & 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.

Langton, S. R. (2018). I don't see it your way: The dot perspective task does not gauge spontaneous perspective taking. *Vision*, *2*(1), 6.

Le Bihan, D., Turner, R., Zeffiro, T. A., Cuenod, C. A., Jezzard, P., & Bonnerot, V. (1993). Activation of human primary visual cortex during visual recall: a magnetic resonance imaging study. *Proceedings of the National Academy of Sciences*, *90*(24), 11802-11805.

Lempers, J. D., Flavell, E. R., & Flavell, J. H. (1977). The development in very young children of tacit knowledge concerning visual perception. *Genetic Psychology Monographs*.

Lopez, C., & Elzière, M. (2018). Out-of-body experience in vestibular disorders–A prospective study of 210 patients with dizziness. *Cortex*, *104*, 193-206.

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.

Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in human neuroscience*, *7*, 558.

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

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1-2), 8-13.

Quesque, F., & Brass, M. (2019). The role of the temporoparietal junction in selfother distinction. *Brain topography*, 32(6), 943-955.

Riecke, B. E., Cunningham, D. W., & Bülthoff, H. H. (2007). Spatial updating in virtual reality: the sufficiency of visual information. *Psychological research*, *71*(3), 298-313.

Roelfsema, P. R., & de Lange, F. P. (2016). Early visual cortex as a multiscale cognitive blackboard. *Annual review of vision science*, *2*, 131-151.

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

Santiesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic bulletin & review*, *22*(3), 844-849.

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.

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

Southgate, V. (2020). Are infants altercentric? The other and the self in early social cognition. *Psychological Review*, *127*(4), 505.

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

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.

Tsakiris, M., Costantini, M., & Haggard, P. (2008). The role of the right temporo-parietal junction in maintaining a coherent sense of one's body. *Neuropsychologia*, *46*(12), 3014-3018.

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

Wang, H., Callaghan, E., Gooding-Williams, G., McAllister, C., & Kessler, K. (2016). Rhythm makes the world go round: An MEG-TMS study on the role of right TPJ theta oscillations in embodied perspective taking. *Cortex*, *75*, 68-81.

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

Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (2020). Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze. *Cognition*, *199*, 104241.