Check for updates

Assessment of embodied visuospatial perspective taking in augmented reality: insights from a reaction time task

Aidan Renata[1,2]*, Renan Guarese[1,3], Marcel Takac[1] and Fabio Zambetta[1]

[1]RMIT University, Melbourne, VIC, Australia, [2]Deakin University, Melbourne, VIC, Australia, [3]KTH Royal Institute of Technology, Stockholm, Sweden

Understanding another person's visual perceptions is known as visuospatial perspective taking, with evidence to date demonstrating it is delineated across two levels, depending on how different that perspective is. Some strategies for visuospatial perspective taking have also been found to involve embodied cognition. However, the current generalisation of these findings is limited due to experimental setup and the use of computer monitors as the interface for experimental tasks. Augmented reality interfaces could possibly extend on the generalisation of these findings by situating virtual stimuli in the real environment, thus providing a higher degree of ecological validity and experimental standardisation. This study aimed to observe visuospatial perspective taking in augmented reality. This was achieved in participant experiments ($N = 24$) using the Left-Right behavioural speeded decision task, which requires participants to discriminate between target objects relative to the perspective of an avatar. Angular disparity and posture congruence between the avatar and participant were manipulated between each trial to delineate between the two levels of visuospatial perspective taking and understand its potentially embodied nature. Although generalised linear mixed modeling indicated that angular disparity increased task difficulty, unexpectedly findings on posture congruence were less clear. Together, this suggests that visuospatial perspective taking in this study can be delineated across two levels. Further implications for embodied cognition and empathy research are discussed.

KEYWORDS

visuospatial perspective taking, augmented reality, cognitive empathy, reaction time, emobdied cognition

1 Introduction

Understanding another individual's mental and emotional state is important for social and emotional functioning. This understanding is in part due to empathy. As a part of general social cognition (Saxe, 2006), empathy includes processes and abilities to perceive, sympathise, infer, and comprehend another individual's emotions or intentions (Decety and Lamm, 2006; McDonald, 2013; Singer and Lamm, 2009). Empathy also allows individuals to use this information to the prosocial benefit of another individual (Mathur et al., 2010). Current best definitions of empathy are complex and lack

consensus, and as such, accurate measurement has become increasingly important (Baumeister et al., 2007; Hall and Schwartz, 2019; Sunahara et al., 2022). Moreover, the propensity to which someone engages in an empathetic response is highly dependent on their personality, the situation, whether the other belongs to the same social group, and even the cost of helping the other (Graziano et al., 2007). Given the problems in defining and measuring empathy, Hall and Schwartz (2019) proposed that research focuses on its associated lower-level constructs and abilities for more precise measurement. Thus, a more accurate description of the processes related to empathy may provide further clarity on this integral part of human social cognition.

Visuospatial perscive taking (VSPT) is the mental representation and transformation involved in understanding the perceptions of another person (Gronholm et al., 2012; Samuel et al., 2023). Colloquially, this might be thought of as putting yourself in another person's shoes, where VSPT is part of this but concerns understanding their spatial and visual world. Some research suggests that VSPT is causally involved in psychological perspective taking, cognitive empathy, and closely related to theory of mind (Caldwell et al., 2022; McDonald, 2013; Sodian et al., 2007). Although alternative explanations exist (Schurz et al., 2021), according to the model of empathy by Decety and Jackson (2004) parallel and hierarchical mental processes work dynamically to produce three major components: self-other representations, self-other awareness, and mental flexibility or perspective taking. The latter of these three involves a cognitive capacity to perceive, infer, and understand the mental state of another person by assuming or reasoning the other's psychological perspective. Decety and Lamm, (2006) elaborated on this cognitive perspective taking, proposing that top-down information processing (reasoning and understanding the other's perspective) worked in parallel with bottom-up information processing (perceiving and attending to the other's perspective). Given this, understanding how VSPT operates may provide further clarity for how individual differences in empathy arise or even how VSPT may go on to affect other higher-level empathy processes.

1.1 Measuring visuospatial perspective taking

Visuospatial perspective taking has been observed using chronometric methods. This usually involves participants recording their reaction to a stimulus where speed and accuracy are the variables of interest. This measure has allowed researchers to infer the cognitive processes involved in VSPT from stimulus detection, stimulus identification, and response selection (Lindemann and Fischer, 2023). By manipulating the difficulty for these cognitive processes to occur, chronometric methods allow a certain degree of specificity as to which stage of the mental operation might be affected and even the nature of the operation (Lindemann and Fischer, 2023). The nature of VSPT is delineated across two levels: Level-1 being a line-of-sight computation where someone can understand what another may be perceiving, and Level-2 where someone can infer and understand the perspective of another that they may not immediately perceive

themselves (Flavell, 1977). Experimental evidence appears to support this delineation. Michelon and Zacks (2006) demonstrated that increasing the difference in perspective through an increase in angular disparity between the observer and an avatar also increased the difficulty of performing VSPT. These researchers found that longer reaction times (RT) were observed for angular disparity greater than 90°, concluding that angles greater than this threshold required more demanding mental operations, as in Level-2 VSPT (Michelon and Zacks, 2006). The exact nature of the mental operations involved in both levels of VSPT has also been a subject of investigation.

Evidence for the nature of Level-2 VSPT has suggested two major strategies for this mental operation. The first is similar to mental object rotation (Shepard and Metzler, 1971), wherein the visual scene is treated as an object to be rotated in mental imagery in order to understand it from a different reference frame (Hegarty and Waller, 2004; Zacks and Tversky, 2005). The second is embodiment, where an imagined physical self is mentally rotated to the desired reference frame in order to infer a desired viewpoint (Cavallo et al., 2017; Kessler and Thomson, 2010; Tversky and Hard, 2009; Zacks and Tversky, 2005). This imagined physical self is also known as the body schema, which is the dynamic mental representation of sensory and proprioceptive information about an individual's body in space (Holmes and Spence, 2004; Morasso et al., 2015). This embodiment strategy taps into larger theories about embodied cognition, supporting the idea that the nature and development of some cognition is informed by our lived physical reality (Fischer, 2024). Kessler and Thomson (2010) demonstrated that manipulating the observer's body position such that an increase in the difference between visual perspective with the agent was exaggerated, that this resulted in an increase in the difficulty of performing VSPT. These researchers concluded that body positions congruent with the direction and nature of the mental rotation facilitated faster and more accurate inferences about the perspective of an agent, and that the inverse was also true.

Behavioural and neuroscientific evidence has highlighted the complex nature of embodied Level-2 VSPT, but suggests that further refinement of experimental conditions is needed. For example, neural activation for mental rotation, body representation, and self/other distinctions were observed during VSPT tasks (Gunia et al., 2021), but it is not clear as to whether this is due to an embodied or mental object rotation strategy (Ang et al., 2023). Furthermore, perspective taking is facilitated by the presence of social and directional cues, like the position of another body in the space (Gunalp et al., 2019; Kessler and Thomson, 2010). This is important as it situates VSPT as a spatial and social cognition that enables the understanding of another person's visual and spatial world in relation to the observer's. However, the spatial aspect of VSPT experiments poses a problem for current findings. Fiehler and Karimpur (2023) have suggested that research on spatial cognition is less generalisable when using stimuli in a pictorial plane on a 2D screen to examine 3D constructs and abilities. This is significant because most VSPT studies to date have utilised 2D computer monitors to display 3D images of stimuli (Samuel et al., 2023). This limits possible conclusions about VSPT, as individuals are asked to perform a 3D mental transformation on a 2D plane for a 3D image. Here, the common experimental setups for VSPT tasks are potentially confounded by their choice of experiment interface,

limiting their generalisability to real social situations that occur in the real 3D environment.

1.2 Assessing embodied VSPT through augmented reality

Immersive virtual technology (IVT) has emerged as a potentially revolutionary tool for investigating social and emotional psychological constructs (Foxman et al., 2021). This is due to the technology's unique ability to place people in 3D, ecologically valid environments while maintaining a high degree of experimental control over the situation or stimuli (Bombari et al., 2015; Kothgassner and Felnhofer, 2020). In fact, there is a growing but significant body of literature using IVT as an interface to investigate psychological perspective taking (Foxman et al., 2021; Paananen et al., 2023). These studies typically use self-report measures and tend to use virtual reality interfaces, where participants are completely immersed in the virtual environment (Foxman et al., 2021; Paananen et al., 2023). However, this may be problematic as some evidence suggests that self-report measures are weakly correlated with behavioural measures (Dang et al., 2020; Sunahara et al., 2022), and that augmented reality (AR) interfaces may strike a finer balance between ecological validity and experimental control (Ventura et al., 2018). This may be due to AR's ability to overlay virtual components or stimuli in the lived, real environment. Furthermore, none have observed VSPT using these technologies as interfaces, and given the proposed ecological and experimental benefits that they provide, there is an opportunity here to see if those benefits transfer to investigations into VSPT.

One type of task for assessing VSPT involves an observer distinguishing the position (left or right) of a target object in relation to another agent's visual perspective (Kessler and Thomson, 2010; Michelon and Zacks, 2006). A similar task assessing embodied VSPT has been adapted for virtual reality by Ueda et al. (2021). By increasing angular disparity to increase the difference in perspective, Ueda et al. (2021) found a delineation between angles above and below 90° using a virtual reality interface to display stimuli. This meant those stimuli were to-scale and placed within an environment that more closely replicates the real environment. To examine the embodiment of VSPT, Ueda et al. (2021) altered the body position of the avatar in the stimuli instead of the participant's. Although this may address the complete immersion in the virtual environment experienced in virtual reality, it has not established whether altering the observer's body position is a relevant factor for embodied VSPT. Doing so would allow a greater degree of replication of previous findings using similar tasks (Kessler and Thomson, 2010; Surtees et al., 2013; Tversky and Hard, 2009).

AR and other virtual interfaces have been useful tools for investigating embodied cognition (Felisatti and Gianelli, 2024). Some research has found that while using these interfaces, an individual's ability to incorporate and act on objects or scenes within these virtual environments reflects these actions in the real environment, especially in regards to the body schema (Berger et al., 2022; D'Angelo et al., 2018; Nakamura et al., 2014). As it stands, however, most research on cognitive empathy and its components has utilised virtual reality interfaces (Foxman et al., 2021; Herrera et al., 2018; Tay et al., 2023). Although virtual reality provides methodological advantages around the balance of ecological validity and experimental standardisation (Bombari et al., 2015; Kothgassner and Felnhofer, 2020), some have offered the critique that participants would be too unfamiliar with this novel interface (Pan and Hamilton, 2018). While this familiarity builds, one possible solution may be using AR as the interface to present stimuli. This interface adds virtual elements to the real environment, potentially striking a finer balance between ecological validity and experimental standardisation (Ventura et al., 2018). Given this, there is an opportunity to expand on the kinds of virtual interfaces used when investigating empathy and its components.

2 Research aims and hypotheses

Overall, there is a need to investigate empathy with more granularity by understanding its component processes, using more objective behavioural measures, and interfaces that promote both ecological validity and experimental standardisation. Research using IVT is headed in this direction, but has yet to expand on the kinds of virtual interfaces used and the components of empathy that are focused on. Therefore, this study aimed to observe VSPT in AR. Specifically, it aimed to observe whether Level-1 and Level-2 VSPT occurred in AR. The design adapted Kessler and Thomson (2010)'s Left-Right task for an AR interface, as it also allows manipulation of a participants body position. This version of the task was chosen because it demonstrates VSPT through increased RT and diminished accuracy when angular disparity and posture incongruency are used to increase the difference in perspectives. Therefore, four hypotheses were proposed:

H1: Higher angles would predict an increase in RT greater than lower angles.

H2: Incongruent postures would predict an increase in RT greater than neutral and congruent postures.
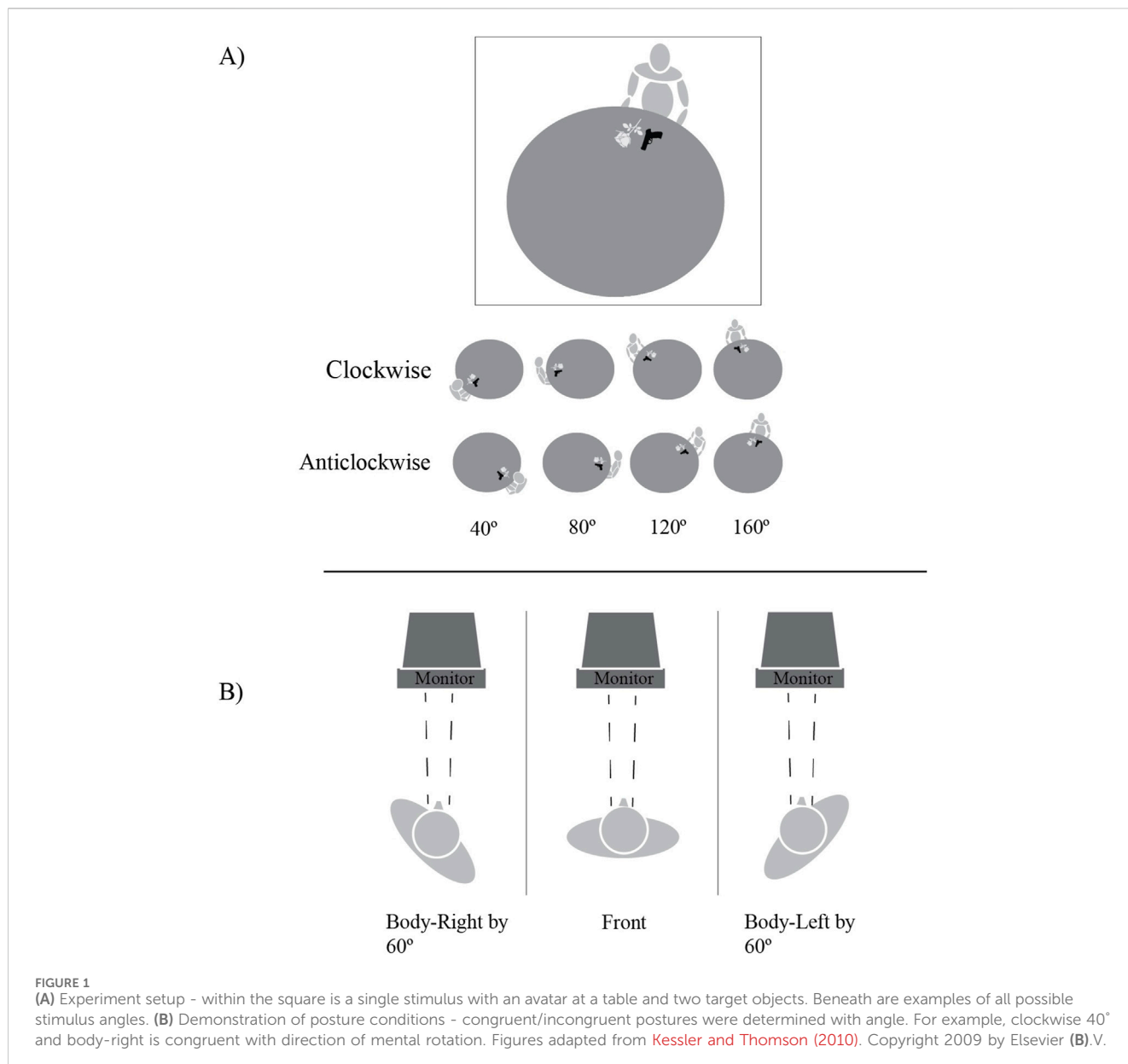
H3: Higher angles would predict a decrease in accuracy greater than lower angles.

H4: Incongruent postures would predict a decrease in accuracy greater than neutral and congruent postures.

3 Methods

3.1 Participants and recruitment

Twenty-four participants were recruited from the greater Melbourne area. Exclusion criteria were being under 18 years of age, blind or vision impaired, existing medical conditions involving seizures or heart conditions, or implanted medical devices. No participants were excluded on this basis. Participant ages ranged from 19 to 38 years ($M = 27.75$). Thirteen identified as female, 8 as male, and 3 as transgender or non-binary. All had normal or corrected-to-normal vision. There is some research suggesting that an individual's dominant eye modulates their visuospatial attention (Schintu et al., 2020). However, eye dominance was not collected in our participants as it was outside of the scope of the present study due to resources and our task focusing on speeded decision making. All participants were in the process of obtaining or had a university bachelor's degree or higher.

FIGURE 1
(A) Experiment setup - within the square is a single stimulus with an avatar at a table and two target objects. Beneath are examples of all possible stimulus angles. **(B)** Demonstration of posture conditions - congruent/incongruent postures were determined with angle. For example, clockwise 40˚ and body-right is congruent with direction of mental rotation. Figures adapted from Kessler and Thomson (2010). Copyright 2009 by Elsevier **(B)**.V.

Ethical approval for the experiment was provided by the RMIT Human Research Ethics Committee[1], and all participants provided written informed consent.

3.2 Materials

A pre-session Qualtrics[2] survey was used to gather demographic information including age, gender, education level, and for the screening of exclusion criteria.

3.2.1 Spatial perspective taking

The computerised Spatial Orientation Test (SOT; (Friedman et al., 2020)) was used to understand individual differences in spatial perspective taking. The SOT requires determining the orientation of a target object while assuming a spatial perspective. This determines an individual's ability to imagine different spatial perspectives.

3.2.2 AR interface

The Microsoft HoloLens 2[3] was used as the platform and user interface for the experiment. Unity 2020.3.29f1[4] was used to develop the AR application and deploy the experiment on the HoloLens 2.

---

1 Under Research Ethics Application Project ID 26382

2 https://www.qualtrics.com/

3 https://www.microsoft.com/en-au/hololens/hardware

4 https://unity.com/

Modelling and rendering the virtual objects for AR, including the avatar, table, gun, and flower[5], was completed using Blender 3.6.1 LTS[6]. Mapping them to the lab environment was done by identifying a visual 2D marker on the table with the Vuforia SDK[7], which anchored and scaled these objects to be proportionate to the room dimensions.

3.2.3 Stimuli and design

A Left-Right task was used to measure VSPT (Kessler and Thomson, 2010; Michelon and Zacks, 2006; Surtees et al., 2013; Tversky and Hard, 2009). This task requires that participants make a speeded decision between which side of an avatar a target object is positioned. To test the mental rotation involved in VSPT, the difference in visual perspectives between the avatar and participant is increased by manipulating their angular disparity in different trials. In this sense, an avatar sitting at a round table was used as a stimulus, appearing at different angles in relation to the participant, who was standing stationary 2 m away from the table. These angles (40°, 80°, 120°, 160°) included both clockwise and counterclockwise rotations around the table. The task design and setup can be seen in Figure 1 which also depicts the original Kessler and Thomson (2010) task setup where participants viewed the stimuli on a computer monitor. This was included to demonstrate how the current study adapted this task for a new AR interface.

Adapting the stimuli from Kessler and Thomson (2010) for an AR inference involved creating and rendering stimuli objects to the lab environment, which in turn meant scaling these to be as close as possible to life size. Both of the clockwise and counterclockwise versions of each angle were adapted form the original task for the current study in AR (see Figure 1). Each stimulus depicted an avatar sitting at the table and two target objects placed in front of them. The target objects were a gun and a flower. These were chosen because they are semantically and visually distinct, minimising chances that participants would confuse them. Since the stimuli appeared *in situ* and to-scale, the orientation angle of 0° would have placed the participant directly behind the stimuli, occluding the view of the target objects. Therefore, this angle was excluded from the trials. Regarding occlusions due to the HoloLens two viewport size, standing 2 m away from the table helped to compensate for this, ensuring that the whole stimulus was visible at all times even with minimal participant movement. An example of what participants actually saw in the lab through the HoloLens 2 can be seen in Figure 2.

To examine the potentially embodied mental rotation involved, the participant's body posture was manipulated to be either neutral, congruent, or incongruent with that of the avatar's body posture. Standing at a mark near a real table, participants were asked to turn their bodies front-on, right by 60°, or left by 60°, while always keeping their heads facing 0° and their gaze focused on the table. These angles were determined by markings
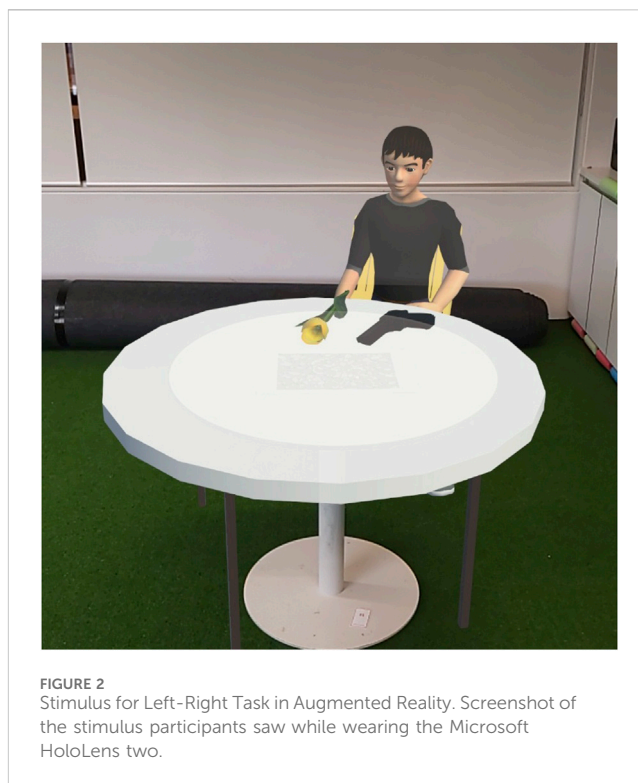

FIGURE 2
Stimulus for Left-Right Task in Augmented Reality. Screenshot of the stimulus participants saw while wearing the Microsoft HoloLens two.

on the ground that were measured relative to the 0° point of the table, where participants would align their feet to match these markings. This resulted a change in their left-right orientation relative to the avatar at a given angle. A neutral posture was defined as the participant's body facing directly forward; congruent as the participant having their body orientation match the current orientation of the avatar, and incongruent as their body orientation match the opposite orientation of the avatar. For example, if turning their body right and presented with an avatar at clockwise angles of 40°, 80°, 120°, or 160°, a participants posture was deemed congruent as they had a matched left-right orientation to the avatar. Whereas if their body was turned right and they were presented with avatar angles anticlockwise 40°, 80°, 120°, or 160°, their left-right orientation does not match the avatar and is therefore incongruent. The same is also true of the reverse. This use of angle and posture to determine posture congruence was taken from Kessler and Thomson (2010), where these authors suggested that incongruence would mean extra effort in mentally aligning postures with an avatar.

Participants wore the HoloLens 2 and through it were presented with the virtual stimulus in AR, overlaying the real table. Between trials, the angle at which the avatar was seated in relation to the participant would change, and the positions of the target objects would swap between being on the left or right of the avatar. This was a repeated measures design, so every participant experienced all angle and posture conditions. The dependent variables were RT (ms), defined as the time from stimulus onset to the moment of response indication, and the accuracy of their response, recorded as either correct or incorrect.

3.3 Procedure

After providing written informed consent, the SOT (Friedman et al., 2020) was administered using a computer monitor. A time limit of 5 min to complete 12 items was set, and the average angular error was calculated for each participant. Task order between SOT and the Left-Right task was alternated across participants.

A short induction to the AR hardware was provided with three practice trials of the Left-Right task. This also involved calibrating the HoloLens 2 to the lab environment using the anchor point, ensuring that the participants were able to see the virtual components in their field of view and in proportion. Participants were verbally instructed to take the visual perspective of the avatar and decide on whether a target object lay on the left or right of the avatar. They were then instructed to respond as quickly and accurately as possible by indicating their response with either a left or right arrow key on a computer keyboard they held. Before each trial, participants received instructions on which posture they should maintain and which target object they should focus on. They were also instructed to keep their heads and gaze facing forward at the centre of the table and to only turn their bodies to face the left or right-facing angled body position.

Participants verbally indicated when they were ready to begin the trial. Then, a fixation cross was displayed on the table for 500 ms, followed immediately by the experimental stimulus. Participants were naive to the order of the AR stimuli conditions, only receiving prior instruction about body-position conditions prior to every trial. Thus, they were not aware of the position the avatar would appear at around the table prior to beginning the trial. Feedback about accuracy was provided for practice trials, but not for the 24 experimental trials. The target object, left/right position of the gun and flower, posture, and angular disparity were counterbalanced across all experimental trials. As each participant experienced all conditions, counterbalancing was achieved using a Latin square wherein each condition was allocated to each participant the same number of times with each participant experiencing a different sequence. This was done to counterbalance learning or task order effects.

3.4 Data analysis

For the main analysis, reaction times less than 100 ms or greater than two times the standard deviation of an individual's mean RT were excluded from the analysis to account for accidental responses and attentional issues (Ratcliff, 1993). Generalised linear mixed-modelling (GLMM) was used to analyse RT and accuracy data, in line with best practice protocols set out by Meteyard and Davies (2020). Cut-offs were not used here in order to preserve the usually positively skewed RT distribution as they may over-transform the data (Lo and Andrews, 2015; Ratcliff, 1993). For both RT and accuracy dependent variables, a GLMM was fit by maximum likelihood using the *lme4* package (Bates et al., 2015) in *R* version 2023.6.2.561[8]. Predictor variables for both were avatar-

---

8 http://www.posit.co

participant angular disparity (40°, 80°, 120°, 160°) and participant body posture (neutral, congruent, incongruent). These predictor variables and their conditions were fixed effects for GLMM, and random effects included participant. Participant was deemed a random effect to account for individual variance in performance (Lo and Andrews, 2015).

In addition, participants were excluded from the main analysis if they were an outlier on both the main task and the SOT. This was done to assist in identifying participants who may have deficits in spatial abilities, as this test is commonly applied in research on spatial cognition (Hegarty and Waller, 2004; Friedman et al., 2020). However, none were excluded from the analysis on this basis.

4 Results

For the Left-Right task, four single data points were treated as missing due to equipment malfunction. Outliers of less than 100 ms or greater than two standard deviations of an individual's mean RT were removed, with 554 data points remaining out of the 576 originally collected. A Wilcoxon Rank-Sum test determined there was no significant difference between clockwise or counterclockwise angles on RT ($W = 29680$, $p = .376$), and so these conditions were treated as being the same. Participants' average angular error on the SOT ranged from 5.70° to 97.83° ($M = 32.18°$, $SD = 22.89°$).

4.1 Reaction time

Mean reaction times are represented in Figure 3. These indicate that RTs were longer on average for higher angles, neutral postures, and incongruent postures. Summary scores for model comparison can be found in Table 1. Models were compared based on their complexity, with less complexity across summary scores indicating a better fit. Given the nested models, the comparison for complexity method was a Likelihood Ratio Test. In comparison to the Null model, Angle Only ($\chi^2(3) = 173.18$, $p < .001$) fit the data better than Posture Only ($\chi^2(2) = 8.21$, $p = .016$). This led to a comparison between Angle Only and a model with interaction of Angle × Posture, however, the model failed to converge ($\chi^2(8) = 15.20$, $p = .055$). Therefore, the interaction term was simplified with Angle + Posture fitting the data best ($\chi^2(2) = 10.11$, $p = .006$). Angle + Posture was chosen as the final model.

R^2 for GLMM was calculated (Nakagawa and Schielzeth, 2013). Marginal R^2 for the selected Angle + Posture model was .07, indicating approximately 7% of the variance in RT is explained by the fixed effects in the model. The conditional R^2 was .16, indicating approximately 16% of the variance in RT is explained by both fixed and random effects in the model. As seen in Table 2, the model intercept set as 40° and neutral body posture was statistically significant. The angle of 80° was not statistically significant. Angle 120° was significant with a 0.13 s increase in RT, and angle 160° was a significant predictor with a 0.31 s increase in RT. Incongruent postures were not statistically significant, but congruent postures had a 0.04 s decrease in RT. Variance for the random effect of participant was 0.01 ($SD = 0.13$).
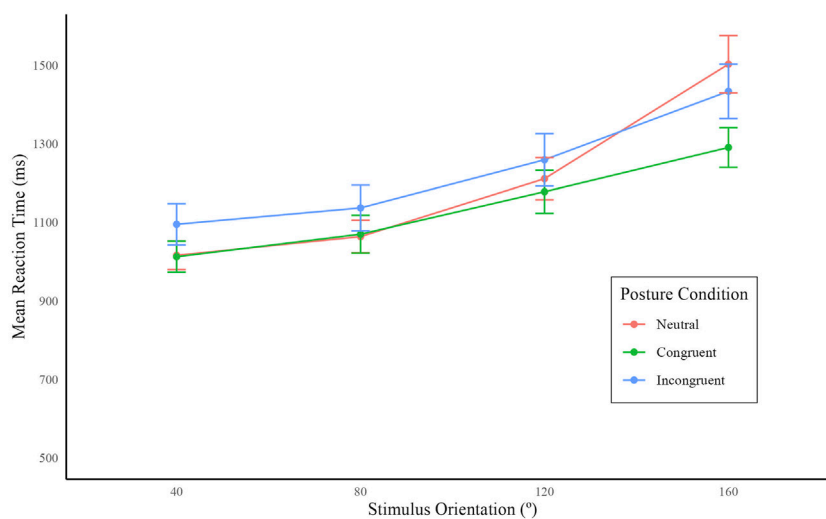
FIGURE 3
Mean reaction times according to stimulus orientation grouped by participant posture. Clockwise and counterclockwise for each angle orientation are counted as one. Error bars represent standard errors.

TABLE 1 Generalised linear mixed modelling summary scores for model comparison and model building RT = reaction time, AIC = Akaike information Criterion, BIC = Bayesian information Criterion, LL = Log likelihood.

Model specification	Model name	Nested model	Fixed effects added	Random effects	Model fit		
					AIC	BIC	LL
RT Only	Null	-	-	Participant	9.42	22.37	−1.71
RT Main Effect	Angle Only	Null	+Angle	"	−157.77	−131.87	84.88
RT Main Effect	Posture Only	Null	+Posture	"	5.22	26.81	2.39
Interaction	Angle × Posture	Angle Only	+Angle × Posture	"	−156.96	−96.52	92.48
Interaction	Angle +Posture	Angle Only	+Angle +Posture	"	−163.88	−129.34	89.94

TABLE 2 Selected Model Reporting for Reaction Time Fixed Effects Upper and lower confidence intervals (CI) have been calculated using the Wald method Model equation: RT ~ Angle + Posture + (1 | Subject).

Fixed effect	β	SE	95% CI	t	p
Intercept	1.17	0.08	1.01, 1.33	14.41	<.001
Angle 80	0.03	0.01	−0.003, 0.07	1.77	.076
Angle 120	0.13	0.02	0.09, 0.17	6.33	<.001
Angle 160	0.30	0.02	0.25, 0.35	12.34	<.001
Congruent	−0.04	0.01	−0.07, −0.005	−2.25	.024
Incongruent	0.01	0.01	−0.02, 0.05	0.84	.397

4.2 Accuracy

The overall percentage correct is represented in Figure 4. On average, participant accuracy was above 90 percent. However, accuracy decreased for higher angles of stimulus orientation. Summary scores for model comparison can be seen in Table 3. In comparison to the Null model, Angle Only ($\chi^2(3) = 11.67$, $p = .008$) significantly fit the data better than Posture Only ($\chi^2(2) = 0.89$, $p = .641$). Posture Only also failed to converge. This led to a comparison of interaction terms between Angle Only with Angle × Posture ($\chi^2(8) = 5.19$, $p = .737$) and Angle + Posture ($\chi^2(2) = 0.84$, $p = .657$), where both were not significant. Thus, Angle Only was determined to be the best fit for the data.

Marginal R^2 calculated for the chosen Angle Only model was .02, indicating that approximately 2% of the variance in accuracy was explained by the fixed effects of the model. The conditional R^2 was .07, indicating that approximately 7% of the variance in accuracy was explained by both fixed and random effects in the model. As seen in Table 4, the model intercept set as 40° was statistically significant. Angle 80° was not a significant predictor of accuracy. However, Angle 120° was a significant predictor with a 1.69 decrease in accuracy score (%). Angle 160° was not a significant
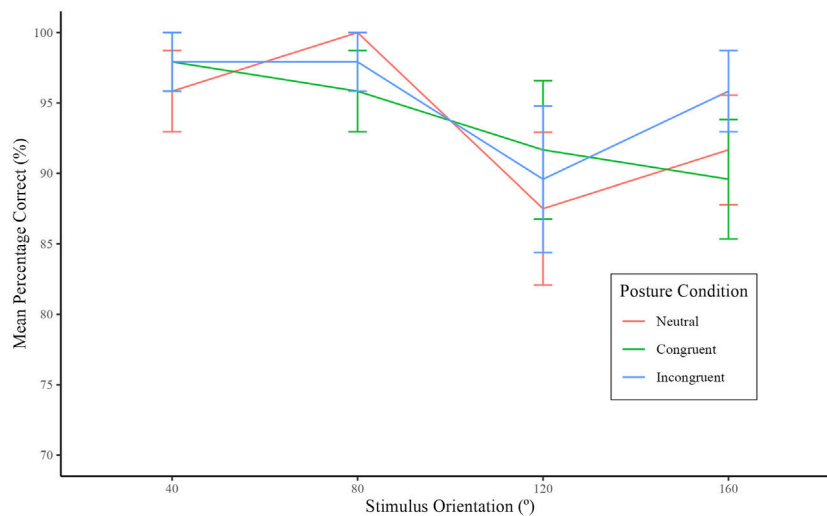
FIGURE 4
Mean percentage correct according to stimulus orientation grouped by participant posture. Clockwise and counterclockwise for each angle orientation are counted as one. Errors bars represent standard errors.

TABLE 3 Generalised linear mixed modelling summary scores for model comparison and model building ACC = accuracy, AIC = Akaike information Criterion, BIC = Bayesian information Criterion, LL = Log likelihood.

Model specification	Model name	Nested model	Fixed effects added	Random effects		Model fit	
					AIC	BIC	LL
ACC Only	Null	-	-	Participant	207.34	215.97	−101.67
ACC Main Effect	Angle Only	Null	+Angle	”	201.67	223.26	−95.84
ACC Main Effect	Posture Only	Null	+Posture	”	210.45	227.72	−101.22
Interaction	Angle ×Posture	Angle Only	+Angle ×Posture	”	212.49	268.61	−95.42
Interaction	Angle +Posture	Angle Only	+Angle +Posture	”	204.83	235.05	−95.42

predictor. The variance for the individual variance in performance for each participant as a random effect in the model was 1.21 ($SD = 1.10$).

4.3 Post-hoc analysis

A *post hoc* analysis to understand the relationship between RT and accuracy across participants was conducted. A Spearman's Rank Order Correlation between mean RT and mean percentage correct for all conditions revealed a significant weak negative relationship ($r_s(70) = −.18, p = .002$).

5 Discussion

A clearer understanding of the nature of both Level-1 and Level-2 VSPT is important in understanding how mental processes operate in complex abilities like empathy. This clarity is sometimes impeded by the nature of experimental setups,

TABLE 4 Selected Model Reporting for Accuracy Fixed Effects Upper and lower confidence intervals (CI) have been calculated using the Wald method Model equation: Accuracy ~ Angle + (1 | Subject).

Fixed effect	β	SE	95% CI	z	p
Intercept	4.40	0.69	3.04, 5.75	6.35	< .001
Angle 80	0.00001	0.84	−1.65, 1.65	0.00	> .99
Angle 120	−1.69	0.67	−3.02, −0.35	−2.48	.012
Angle 160	−1.14	0.71	−2.54, 0.25	−1.60	.108

however, AR offers a potential new avenue to gaining a clearer understanding of these mental processes by increasing the ecological validity of controlled experimental tasks. The aim of this study was to observe VSPT in AR, specifically aiming to observe whether the delineation of Level-1 and Level-2 still emerged within this new experimental setup. Choosing to adapt the Kessler and Thomson (2010) task where the strategy of Level-2 VSPT was embodied, this study proposed four hypotheses to investigate this embodiment strategy for Level-2 and also the delineation with Level-1. The

present results support H1 as we observed larger angles (120° and 160°) predicting an increase in RTs greater than in lower angles (40° and 80°). This goes part of the way to demonstrate the delineation between Level-1 and Level-2 VSPT. Partial support for H2 was found, where incongruent postures predicted an increase in RTs greater than congruent postures but, unexpectedly, not greater than neutral postures. This is unexpected because incongruent postures were thought to disadvantage the observer, whereas neutral postures were thought to neither disadvantage nor advantage them. Similarly, this study only found partial support for H3, with only the angle of 120° predicting a significant decrease in accuracy, but not 160°, when compared to lower angles. This was also unexpected, as a difference in accuracy for angles above and below 90° has been found elsewhere (Kessler and Thomson, 2010; Michelon and Zacks, 2006). Finally, the present results did not support H4. This too was unexpected, as there was no significant difference between incongruent, congruent, or neutral postures in predicting accuracy. Whereas Kessler and Thomson (2010) had found posture congruence to affect accuracy.

Regarding the delineation of Level-1 and Level-2 VSPT, the present study converged with previous findings on this (Kessler and Thomson, 2010; Michelon and Zacks, 2006; Surtees et al., 2013; Tversky and Hard, 2009), providing evidence that angles above 90° produced long RTs compared to angles below. The present study differs from Kessler and Thomson (2010), who found a significant decrease in accuracy at higher angles. Here, only the angle of 120° produced a significant error rate which recuperated at the higher angle of 160°. This is different again from Michelon and Zacks (2006), who found angle had no effect on accuracy. Thus, the present results suggest there was something particularly difficult about the condition of 120°, especially considering it also increased RT. Because participants still responded with an accuracy greater than chance at 120°, it is reasonable to conclude they were applying some method of VSPT to make their inference. However, given the recuperation at 160°, 120° may be a threshold for difference in perspectives, and differences beyond this require more careful inference. Further, although not within the scope of the current study to assess the relationship between speed and accuracy directly, a speed-accuracy tradeoff may have been present here. This speed-accuracy tradeoff involves participants sacrificing either one, in order to be able to meet the standards of the other by either being as quick as or as accurate as possible (Proctor and Vu, 2003). Participants in this study were told to focus on both, and recently published work has found that ambivalent instructions like these may affect error rates in decision tasks (Selimi and Moeller, 2024). Overall, it appears that the condition of increasing angular disparity to increase the difference in visuospatial perspective was able to be adapted for AR. As the experiment occurred to-scale and in the real environment, this evidence provides support that VSPT may indeed be organised across two levels.

As for the assertion that Level-2 VSPT is embodied, the current study is less confident. It could be said that a facilitation effect is present, as congruent postures resulted in faster RTs. However, the unexpected finding for incongruent postures casts doubt on this. A possible explanation for this is that participants in this study were standing, and viewing the entire stimuli on a human scale. As previously stated, the use of AR as the stimuli interface was where this study differed from previous research (Kessler and Thomson, 2010; Tversky and Hard, 2009). Ueda et al. (2021) did find a difference between posture positions on a similar task in their study using virtual reality, but the posture being altered in their study was the avatar's. With the addition of the present study's findings on posture, future research may need to consider altering posture as a method of detecting embodiment with greater caution.

Further, the unexpected finding of posture congruence not predicting accuracy adds to the mixed findings within the literature. Kessler and Thomson (2010) found that congruence affected accuracy in certain conditions, whereas Ueda et al. (2021) found no effect of posture congruence on accuracy in virtual reality. In the context of a reaction time task and the present results, this could indicate that posture may not interfere with a participants ability to complete the task. However, with the present mixed findings on posture and RT and the fact that participants were standing for our study, it may be that posture conditions need more consideration when applied using an AR interface for stimuli.

Considering both data together, angular disparity and posture only explain a small percentage of the variance in RT and accuracy data. This is also the case when accounting for individual differences in performance. Further, the weak negative correlation between RT and accuracy also suggests a speed-accuracy trade-off. This suggests that although these differences in perspective are important for VSPT, there are other contributing factors that are not explained here. Given the nature of chronometric methods like RT (Lindemann and Fischer, 2023; Lo and Andrews, 2015; Ratcliff, 1993), these other factors may be cognitive processes like attention or cognitive load. For example, the novelty of AR interfaces could also contribute to increased RTs as the unfamiliarity with the interfaces may put more demand on working memory and attention.

The results of this study provide initial evidence that AR can be used to investigate VSPT, a component of empathy. By adapting a usually computerised task for this interface, participants were able to experience stimuli on a human scale situated in the real environment. Although this study did not directly compare AR to traditional computer monitors, it suggests that AR is also an appropriate avenue for constructs like the embodied cognition of VSPT. Furthermore, the findings of this study also reflect Decety and Jackson (2004)'s model, where top-down abilities like psychological perspective taking are informed by bottom-up information processing (VSPT in this case). Establishing that an observer can infer the visual perspective of another in AR will be important for future experiments on cognitive empathy. This study has also demonstrated that behavioural measures of components of empathy can be observed in AR. This is also relevant for the use of virtual technologies as intervention tools for social cognition in mental disorders. There may be a chance to target VSPT in these interventions, as VSPT has been causally linked to the development of more complex theory of mind and cognitive empathy abilities (Caldwell et al., 2022; Sodian et al., 2007).

6 Limitations and future directions

A strength of the present study is that it provided novel insight into how VSPT operates within AR. However, it is limited

by its small sample size and relatively small amount of datapoints. Future studies should endeavour to increase both. This would aid in statistical analyses like drift models, which have been shown to provide more insight into the time course of the cognitive processes involved in behavioural decision tasks (Lindemann and Fischer, 2023). Especially in the context of an emerging technology like AR, these cognitive processes may include individual differences in how information is perceived and attended to in the virtual environment.

This study did not investigate what information about another's perspective is used in VSPT. Thus, a possible future direction is the use of eye-tracking data, which has proven useful in studies on virtual technology and social cognition (Geraets et al., 2021). This may provide further explanation of what information is related to error rates at higher angles. Understanding how an observer attends to this information may also be informed by individual differences in perception, like eye dominance. As previously stated eye dominance was not within the scope of this study, but future research may benefit from understanding what role it plays in VSPT, especially as we come to understand how eye dominance modulates visuospatial attention. This may also help to clarify how information about posture and body position is relevant to Level-2 VSPT emobdied strategies. Although VSPT was observed using an apparatus with high ecological validity such as AR, it did not make a direct comparison to the computer monitor version of the task. Thus, future studies may include comparing performance across these two apparatuses.

7 Conclusion

The present study suggests that AR can be used to assess VSPT, meaning that people can take the visual perspective of another agent while using this immersive virtual technology. This was achieved by adapting the Left-Right task for an AR interface. The findings suggest that in the real environment, where stimuli are to-scale, altering body posture may not be sufficient to infer the body schema as the mechanism of embodiment in Level-2 VSPT. Even so, this study helps to confirm that the greater the difference between visual and spatial perspectives, the more difficult it is to assume and thus infer that other perspective. Furthermore, the knowledge that VSPT occurs in AR has potential implications for research on cognitive empathy and social cognition studies using virtual interfaces. The implication being that understanding an individual's visual perspective is part of understanding their psychological perspective.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the RMIT Human Research Ethics Committee. The studies were conducted in accordance with the local legislation and institutional requirements.

The participants provided their written informed consent to participate in this study.

Author contributions

AR: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Writing–original draft, Writing–review and editing. RG: Conceptualization, Methodology, Resources, Software, Visualization, Writing–original draft, Writing–review and editing. MT: Conceptualization, Methodology, Project administration, Supervision, Writing–review and editing. FZ: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing–review and editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

---

9   https://assetstore.unity.com/packages/3d/characters/humanoids/fantasy/relaxed-man-character-32665

10 https://www.cgtrader.com/free-3d-models/plant/flower/low-poly-tulip-free-sample

11 https://www.cgtrader.com/free-3d-models/military/gun/desert-eagle-648d0830-72b8-47af-a111-e3bf0663f451

References

Ang, N., Brucker, B., Rosenbaum, D., Lachmair, M., Dresler, T., Ehlis, A.-C., et al. (2023). Exploring the neural basis and modulating factors of implicit altercentric spatial perspective-taking with fnirs. *Sci. Rep.* 13 (1), 20627. doi:10.1038/s41598-023-46205-w

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi:10.18637/jss.v067.i01

Baumeister, R. F., Vohs, K. D., and Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspect. Psychol. Sci.* 2, 396–403. doi:10.1111/j.1745-6916.2007.00051.x

Berger, C. C., Lin, B., Lenggenhager, B., Lanier, J., and Gonzalez-Franco, M. (2022). Follow your nose: extended arm reach after pinocchio illusion in virtual reality. *Front. Virtual Real.* 3, 712375. doi:10.3389/frvir.2022.712375

Bombari, D., Schmid Mast, M., Canadas, E., and Bachmann, M. (2015). Studying social interactions through immersive virtual environment technology: virtues, pitfalls, and future challenges. *Front. Psychol.* 6, 869. doi:10.3389/fpsyg.2015.00869

Caldwell, M. P., Cheung, H., Cheung, S.-K., Li, J.-B., and Carrey Siu, T.-S. (2022). Visuospatial perspective-taking in social-emotional development: enhancing young children's mind and emotion understanding via block building training. *BMC Psychol.* 10, 264. doi:10.1186/s40359-022-00976-5

Cavallo, A., Ansuini, C., Capozzi, F., Tversky, B., and Becchio, C. (2017). When far becomes near: perspective taking induces social remapping of spatial relations. *Psychol. Sci.* 28, 69–79. doi:10.1177/0956797616672464

Dang, J., King, K. M., and Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends cognitive Sci.* 24, 267–269. doi:10.1016/j.tics.2020.01.007

D'Angelo, M., di Pellegrino, G., Seriani, S., Gallina, P., and Frassinetti, F. (2018). The sense of agency shapes body schema and peripersonal space. *Sci. Rep.* 8, 13847. doi:10.1038/s41598-018-32238-z

Davis, M. H. (1980). Interpersonal reactivity index

Decety, J., and Jackson, P. L. (2004). The functional architecture of human empathy. *Behav. cognitive Neurosci. Rev.* 3, 71–100. doi:10.1177/1534582304267187

Decety, J., and Lamm, C. (2006). Human empathy through the lens of social neuroscience. *Sci. World J.* 6, 1146–1163. doi:10.1100/tsw.2006.221

Felisatti, A., and Gianelli, C. (2024). "A taxonomy of methods in embodied cognition research," in *Experimental methods in embodied cognition* (Taylor & Francis Group), 19–34.

Fiehler, K., and Karimpur, H. (2023). Spatial coding for action across spatial scales. *Nat. Rev. Psychol.* 2, 72–84. doi:10.1038/s44159-022-00140-1

Fischer, M. H. (2024). "The embodied cognition approach: principles and research questions," in *Experimental methods in embodied cognition* (Taylor & Francis Group), 3–18.

Flavell, J. H. (1977). The development of knowledge about visual perception. *Neb. Symposium Motivation* 25, 43–76.

Foxman, M., Markowitz, D. M., and Davis, D. Z. (2021). Defining empathy: interconnected discourses of virtual reality's prosocial impact. *New Media and Soc.* 23, 2167–2188. doi:10.1177/1461444821993120

Friedman, A., Kohler, B., Gunalp, P., Boone, A. P., and Hegarty, M. (2020). A computerized spatial orientation test. *Behav. Res. methods* 52, 799–812. doi:10.3758/s13428-019-01277-3

Geraets, C., Tuente, S. K., Lestestuiver, B., Van Beilen, M., Nijman, S., Marsman, J., et al. (2021). Virtual reality facial emotion recognition in social environments: an eye-tracking study. *Internet interv.* 25, 100432. doi:10.1016/j.invent.2021.100432

Graziano, W. G., Habashi, M. M., Sheese, B. E., and Tobin, R. M. (2007). Agreeableness, empathy, and helping: a person × situation perspective. *J. personality Soc. Psychol.* 93, 583–599. doi:10.1037/0022-3514.93.4.583

Gronholm, P. C., Flynn, M., Edmonds, C. J., and Gardner, M. R. (2012). Empathic and non-empathic routes to visuospatial perspective-taking. *Conscious. cognition* 21, 494–500. doi:10.1016/j.concog.2011.12.004

Gunalp, P., Moossaian, T., and Hegarty, M. (2019). Spatial perspective taking: effects of social, directional, and interactive cues. *Mem. and cognition* 47, 1031–1043. doi:10.3758/s13421-019-00910-y

Gunia, A., Moraresku, S., and Vlček, K. (2021). Brain mechanisms of visuospatial perspective-taking in relation to object mental rotation and the theory of mind. *Behav. Brain Res.* 407, 113247. doi:10.1016/j.bbr.2021.113247

Hall, J. A., and Schwartz, R. (2019). Empathy present and future. *J. Soc. Psychol.* 159, 225–243. doi:10.1080/00224545.2018.1477442

Hegarty, M., and Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* 32, 175–191. doi:10.1016/j.intell.2003.12.001

Herrera, F., Bailenson, J., Weisz, E., Ogle, E., and Zaki, J. (2018). Building long-term empathy: a large-scale comparison of traditional and virtual reality perspective-taking. *PloS one* 13, e0204494. doi:10.1371/journal.pone.0204494

Holmes, N. P., and Spence, C. (2004). The body schema and multisensory representation (s) of peripersonal space. *Cogn. Process.* 5, 94–105. doi:10.1007/s10339-004-0013-3

Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi:10.1016/j.cognition.2009.08.015

Kothgassner, O. D., and Felnhofer, A. (2020). Does virtual reality help to cut the gordian knot between ecological validity and experimental control? *Ann. Int. Commun. Assoc.* 44, 210–218. doi:10.1080/23808985.2020.1792790

Lindemann, O., and Fischer, M. H. (2023). "Chronometric methods: reaction times and movement times," in *Experimental methods in embodied cognition* (Taylor & Francis Group), 37–48.

Lo, S., and Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front. Psychol.* 6, 1171. doi:10.3389/fpsyg.2015.01171

Mathur, V. A., Harada, T., Lipke, T., and Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *Neuroimage* 51, 1468–1475. doi:10.1016/j.neuroimage.2010.03.025

McDonald, S. (2013). Impairments in social cognition following severe traumatic brain injury. *J. Int. Neuropsychological Soc.* 19, 231–246. doi:10.1017/s1355617712001506

Meteyard, L., and Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *J. Mem. Lang.* 112, 104092. doi:10.1016/j.jml.2020.104092

Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective taking. *Percept. and Psychophys.* 68, 327–337. doi:10.3758/bf03193680

Morasso, P., Casadio, M., Mohan, V., Rea, F., and Zenzeri, J. (2015). Revisiting the body-schema concept in the context of whole-body postural-focal dynamics. *Front. Hum. Neurosci.* 9, 83. doi:10.3389/fnhum.2015.00083

Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* 4, 133–142. doi:10.1111/j.2041-210x.2012.00261.x

Nakamura, S., Mochizuki, N., Konno, T., Yoda, J., and Hashimoto, H. (2014). Research on updating of body schema using ar limb and measurement of the updated value. *IEEE Syst. J.* 10, 903–911. doi:10.1109/jsyst.2014.2373394

Paananen, V., Kiarostami, M. S., Lik-Hang, L., Braud, T., and Hosio, S. (2023). From digital media to empathic spaces: a systematic review of empathy research in extended reality environments. *ACM Comput. Surv.* 56, 1–40. doi:10.1145/3626518

Pan, X., and Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape. *Br. J. Psychol.* 109, 395–417. doi:10.1111/bjop.12290

Proctor, R. W., and Vu, K.-P. L. (2003). "Action selection," in *Handbook of psychology*, 293–316.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychol. Bull.* 114, 510–532. doi:10.1037/0033-2909.114.3.510

Samuel, S., Cole, G. G., and Eacott, M. J. (2023). It's not you, it's me: a review of individual differences in visuospatial perspective taking. *Perspect. Psychol. Sci.* 18, 293–308. doi:10.1177/17456916221094545

Saxe, R. (2006). Uniquely human social cognition. *Curr. Opin. Neurobiol.* 16, 235–239. doi:10.1016/j.conb.2006.03.001

Schintu, S., Chaumillon, R., Guillaume, A., Salemme, R., Reilly, K., Pisella, L., et al. (2020). Eye dominance modulates visuospatial attention. *Neuropsychologia* 141, 107314. doi:10.1016/j.neuropsychologia.2019.107314

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., et al. (2021). Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychol. Bull.* 147, 293–327. doi:10.1037/bul0000303

Selimi, S., and Moeller, B. (2024). Instructed speed and accuracy affect binding, *Instr. speed accuracy affect Bind.* 88, 1203–1211. doi:10.1007/.s00426-024-01927-y

Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171, 701–703. doi:10.1126/science.171.3972.701

Singer, T., and Lamm, C. (2009). The social neuroscience of empathy. *Ann. N. Y. Acad. Sci.* 1156, 81–96. doi:10.1111/j.1749-6632.2009.04418.x

Sodian, B., Thoermer, C., and Metz, U. (2007). Now i see it but you don't: 14-month-olds can represent another person's visual perspective. *Dev. Sci.* 10, 199–204. doi:10.1111/j.1467-7687.2007.00580.x

Sunahara, C. S., Rosenfield, D., Alvi, T., Wallmark, Z., Lee, J., Fulford, D., et al. (2022). Revisiting the association between self-reported empathy and behavioral assessments of social cognition. *J. Exp. Psychol. General* 151, 3304–3322. doi:10.1037/xge0001226

Surtees, A., Apperly, I., and Samson, D. (2013). The use of embodied self-rotation for visual and spatial perspective-taking. *Front. Hum. Neurosci.* 7, 698. doi:10.3389/fnhum.2013.00698

Tay, J. L., Xie, H., and Sim, K. (2023). Effectiveness of augmented and virtual reality-based interventions in improving knowledge, attitudes, empathy and stigma regarding people with mental illnesses—a scoping review. *J. Personalized Med.* 13, 112. doi:10.3390/jpm13010112

Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129. doi:10.1016/j.cognition.2008.10.008

Ueda, S., Nagamachi, K., Nakamura, J., Sugimoto, M., Inami, M., and Kitazaki, M. (2021). The effects of body direction and posture on taking the perspective of a humanoid avatar in a virtual environment. *PloS one* 16, e0261063. doi:10.1371/journal.pone.0261063

Ventura, S., Baños, R. M., and Botella, C. (2018). Virtual and augmented reality: new frontiers for clinical psychology. *InTech*. doi:10.5772/intechopen.74344

Zacks, J. M., and Tversky, B. (2005). Multiple systems for spatial imagery: transformations of objects and bodies. *Spatial Cognition Comput.* 5, 271–306. doi:10.1207/s15427633scc0504_1