My Right, Your Left: A Speeded Reaction Time Task for Visuospatial Perspective Taking in Augmented Reality

Aidan Renata

Discipline of Psychology, Royal Melbourne Institute of Technology

BESC1427: Thesis 2

Prof Fabio Zambetta, Dr Marcel Takac, Mr Renan Guarese

October 30, 2023

Word Count: 4996

Abstract

Augmented reality technologies (AR) are used as intervention tools in social cognition and empathy. While mounting evidence suggests their effectiveness, the underlying social cognitive mechanisms remain unclear leaving these claims of effect unsubstantiated. To address this, some have called for research on empathy to focus on behavioural and empirical measures of lower level contributing processes and abilities. One such lower-level ability is visuospatial perspective taking (VSPT) – inferring another's visual perspective through mental transformations. This study aimed to observe VSPT in AR. This was achieved using the speeded decision Left-Right task, which requires participants to discriminate between objects from an avatar's perspective. Twenty-four participants aged 19 to 38 years old (M = 27.75, SD = 3.84), completed 24 trials each. By increasing the difference in perspective between the participant and avatar, VSPT becomes more difficult to perform as denoted by reaction time and accuracy. This difference was increased through angular disparity (40º, 80º, 120º, 180º) and posture congruency (congruent, incongruent, and neutral) between the avatar and participant. As expected, generalised linear mixed modelling indicated that angular disparity increased the difficulty. However, an unexpected finding was posture congruency being less important for determining difficulty. Overall, the findings suggest that VSPT does occur in AR. The significance being that understanding another's visual perspective in using this interface is possible, which is important for higher-level social cognition and empathy. Further implications for experiments on VSPT are discussed, especially in relation to embodied cognition.

Keywords: visuospatial perspective taking, augmented reality, cognitive empathy, reaction time.

My Right, Your Left: A Speeded Reaction Time Task for Visuospatial Perspective Taking in Augmented Reality

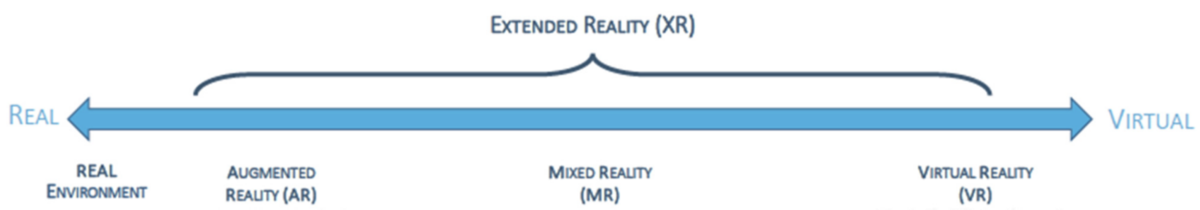
Technology interfaces such as augmented reality (AR) have been proven to be effective pro-social empathy intervention tools (Foxman et al., 2021). Because AR blends virtual components with the real environment (Arena et al., 2022), researchers are able to provide social experiences that people may not have had otherwise (Guarese et al., 2023). This has led many to address cognitive empathy, the ability to understand someone's psychological perspective, in a bid to increase individual differences in this ability through AR (Schrier & Farber, 2021). However, understanding someone's psychological perspective requires an understanding of their visual perspective as well (Gronholm et al., 2012). This is called visuospatial perspective taking (VSPT), which is the mental transformation involved in understanding the perceptions of another person (Samuel et al., 2023). Further, VSPT has been found to be important in theory of mind (Caldwell et al., 2022), an ability involved in the mental representation of the internal states of others - which is closely related to cognitive empathy (Premack & Woodruff, 1978; Schurz et al., 2021; Wimmer & Perner, 1983). There is not enough evidence to establish whether VSPT can occur in AR and, without knowing this, the mechanisms involved in the AR interventions are not fully understood, leaving the claims of effect unsubstantiated. For this reason, further investigation into whether VSPT occurs in AR is required.

Immersive Virtual Technology, Empathy, and Social Cognition

Unique insights into empathy may be gained through immersive virtual technology. Lab research into social behaviours sacrifice ecological validity for experimental standardisation, which virtual technology may remedy (Kothgassner & Felnhofer, 2020). For example, using real people as stimuli may confound or distract participants, but highly controlled stimuli may be too abstracted from the natural environment. Immersive virtual technology offers a solution to this problem with its ability to blend the natural environment with controlled virtual stimuli (Bombari et al., 2015). This results in a greater degree of generalisation to findings. Further, there is a range of immersive virtual technology, each offering a different degree of immersion in the virtual environment. The sense of immersion is the experience of a virtual world as if it were real, partly mediated by a feeling of being there (Bowman & McMahan, 2007). This degree of immersion can be defined on an axis from the real environment to a complete virtual environment in virtual reality (Milgram et al., 1995). AR positions interactive virtual elements or components in the real environment in real time (Arena et al., 2022). Figure 1 depicts where AR belongs on this axis.

Figure 1

Axis of Immersion for Immersive Virtual Technology



Note. Extended reality is a catch-all phrase for these technologies. From "An Overview of Augmented Reality", by F. Arena, M. Collotta, G. Pau, and F. Termine, 2022, *Computers, 11*(2), Article e28 (https://doi.org/10.3390/computers11020028).Copyright 2022 by the authors.

As a part of general social cognition (Saxe, 2006), empathy includes processes and abilities to perceive, sympathise, infer, and comprehend another's emotions or intentions (McDonald, 2013). Decety and Lamm (2006) proposed a dual process model of empathy, where top-down information processing involves perspective taking, amongst other things. This perspective taking is based on a self/other distinction and allows the active consideration of another's psychological perspective (Decety & Lamm, 2006; Singer & Lamm, 2009). This is also known as theory of mind but can be generally thought of as cognitive empathy (Zaki & Oschner, 2012). Cognitive empathy is important for overall empathic understanding (Decety & Jackson, 2004). In Decety and Lamm's model information processed at the bottom-up level feeds into cognitive empathy, which includes understanding the perceptions and behaviours of others. Although some have argued for a different conceptual organisation (Schurz et al., 2021), this highlights the component processes and abilities involved in cognitive empathy.

The measurement of cognitive empathy's component processes and abilities are highly debated. Calls have been made for research on social phenomenon to pursue behavioural observations

(Baumeister et al., 2007). Hall and Schwartz (2019) suggest studies on empathy focus on its lower-level components with an emphasis on behavioural and empirical operationalisations. However, research on immersive virtual technology and cognitive empathy has heavily relied on self-report measures (Foxman et al., 2021). Additionally, there is growing evidence that self-report and behavioural measures of empathy are weakly correlated (Dang et al., 2020; Sunahara et al., 2022). This casts doubt on virtual technology as an intervention tool in social cognition and empathy as there is a lack of empirical data.

Virtual reality interfaces are most used in research on cognitive empathy (Foxman et al., 2021; Herrera et al., 2018; Tay et al., 2023), with the benefit being a balance between ecological validity and experimental standardisation (Bombari et al., 2015; Kothgassner & Felnhoffer, 2020). Even so, some researchers have critiqued this use of virtual reality citing interfaces are too new to market, and so participants' unfamiliarity with the interfaces may confound findings (Pan & Hamilton, 2018). Although work is underway to bridge this gap (Cipresso et al., 2018), AR's ability to add virtual elements to the real environment may strike a finer balance between ecological validity and experimental standardisation (Ventura et al., 2018). Furthermore, virtual reality is a potentially revolutionary tool for clinical interventions in social cognition (Gainsford et al., 2020; Lohaus et al., 2023). However, AR may be more beneficial because it is less immersive, offering clients the opportunity to situate their learned therapy skills in the real environment (Giglioli et al., 2015). For these reasons, it may be beneficial for research to expand on cognitive empathy by utilising AR to investigate lower-level cognitive processes with empirical data.

Visuospatial perspective taking

Visuospatial perspective taking is important for cognitive empathy (Gronholm et al., 2012) with well-established neural overlap (Schurz et al., 2013) and may be causally involved in its development (Caldwell et al., 2022; Sodian et al., 2007). VSPT can also be measured through the behavioural decision Left-Right task (Michelon & Zacks, 2006; Samuel et al., 2023). This measure applies a chronometric method, using reaction times (RT) and accuracy, revealing the cognitive processes involved in the decision from stimulus detection, stimulus identification, and response selection (Lindemann & Fischer, 2023). The indication or a decision between left or right as to the location of a target object is the behaviour measured in the Left-Right task (Michelon & Zacks, 2006).

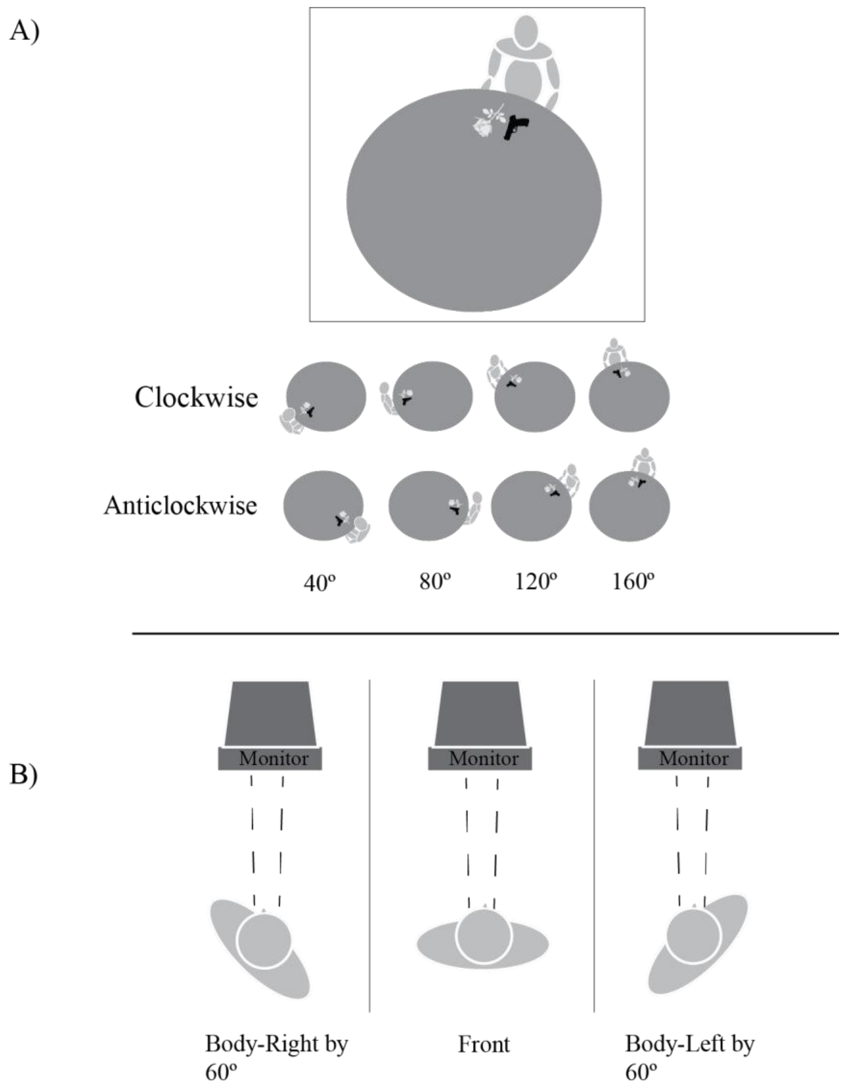
Chronometric methods work to increase the difficulty of a cognitive process involved in this decision making by manipulating some relevant variable involved in the mental process (Lindemann & Fischer, 2023). Michelon and Zacks (2006) found that increasing the difference in perspective through angular disparity between a participant and an avatar increased the difficulty of VSPT. The researchers concluded VSPT has two levels, at angles below 90º participants would use line-of-sight computations (Level 1) and above 90º they would use mental rotation (Level 2; Michelon & Zacks, 2006). Michelon and Zacks posit that mental rotation at Level 2 was necessary because the difference between perspectives is much greater and must involve more complex mental transformations.

This use of chronometric methods to infer the involvement of mental rotation in VSPT is contested. Cole and Millet (2019) offer the critique that the time from stimulus onset to an indication of a response may be confounded by some other unknown mental process. Although this is a fair critique, neuroscientific evidence has demonstrated brain activation in areas for mental rotation, body representation, and self/other distinctions during VSPT tasks (Gunia et al., 2021). In fact, mental rotation is thought to be involved in how an observer may imagine different spatial perspectives generally (Kozhenikov & Hegarty, 2001; Hegarty & Waller, 2004). To account for these potential unknowns, some have attempted to explain mental rotation in VSPT in greater detail.

Embodied cognition theories posit that some cognitions are restrained to representations or processes that mirror our physical reality such as the body, or parts of it (Fischer, 2023; Goldman & de Vignemont, 2009; Lindblom, 2020). Kessler and Thomson (2010) suggest that second level VSPT is embodied by the body schema, which is the dynamic mental representation of sensory and proprioceptive information about our body in space (Holmes & Spence, 2004; Morasso et al., 2015). Kessler and Thomson found that while performing the Left-Right task, the incongruence between participant and avatar posture affected task difficulty indicated by slower RTs and diminished accuracy. The researchers concluded that participants must be mentally rotating their body schema from their egocentric reference frame to the target reference frame of the avatar (Kessler & Thomson, 2010). This is a convincing explanation considering the brain activation for mental rotation, body representation, and self/other distinctions (Gunia et al., 2021). It also goes some way to address the critique of other unknown mental processes involved in VSPT (Cole & Millet, 2019). However, Kessler and Thomson's findings are limited by the fact the task was completed using computerised stimuli on a monitor. The problem lies in participants being required to perform VSPT across different spatial scales, where the supposed mental rotation is used to infer the perspective of a 2D monitor in response to an image that depicts 3D space. Fiehler and Karimpur (2021) have noted that research on spatial cognitions may be less generalisable when using stimuli in a pictorial plane on a 2D monitor to examine 3D constructs. Therefore, AR and other virtual immersive interfaces may be used to investigate embodied cognitions (Felisatti & Gianelli, 2023).

Figure 2

Experiment Setup and Demonstration of Posture Conditions



Note. A) 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) depicts experiments setup and participant posture manipulations. Congruent/incongruent postures were determined with angle. For example, clockwise 40º and body-right is congruent with direction of mental rotation. Adapted from "The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference", by K. Kessler, and L.A. Thomson, 2010, *Cognition, 114*(1), p. 72 (https://doi.org/10.1016/j.cognition.2009.08.015). Copyright 2009 by Elsevier B.V.

The Kessler and Thomson (2010) version of the Left-Right task was adapted for virtual reality by Ueda et al. (2021). These researchers showed VSPT was made more difficult, the greater the difference between avatar and participant perspective was. However, the complete immersion in the virtual environment meant participant posture was irrelevant, so avatar posture was altered (Ueda et al., 2021). Ueda et al. demonstrated that VSPT did occur in VR, but the avatar posture being altered does not speak as clearly to the embodied nature of Level 2 VSPT. Recent evidence suggests the body schema may be highly relevant to how individuals interact with and understand virtual environments (Berger et al., 2022; D'Angelo et al., 2018; Nakamura et al., 2016). If this is the case, then the embodied nature of VSPT is relevant to understanding how this social cognition operates in AR.

Although immersive technologies have been used to observe and intervene in high-level abilities like empathy and cognitive empathy, not all their lower-level processes have been established in this environment. With a lack of behavioural measures of empathy in AR, the cognitive mechanisms involved in these technologies as intervention tools are not fully understood. Visuospatial perspective taking is a lower-level component of cognitive empathy with well-established behavioural measures, yet it has not been directly observed in AR. Furthermore, current methods for VSPT have relied on computerised experiments, which limits ecological validity. Augmented reality offers a unique opportunity to place experimental stimuli in situ, thus increasing this ecological validity while maintaining experimental standardisation.

Research Aims and Hypotheses

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's (2010) Left-Right task for an AR interface. This version of the task was employed 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.

Method

Participants and Recruitment

Twenty-four participants were recruited from the greater Melbourne area. Pre-session surveys showed all participants had normal or corrected-to-normal vision and used technology in their work or study. All were in the process of obtaining or had a bachelor's degree or higher. Their ages ranged from 19 to 38 years (M = 27.75). Thirteen identified as female, 8 as male, and 3 as transgender or non-binary. Exclusion criteria were included being under 18 years of age, blind or vision impaired, medical conditions involving seizures or heart conditions, or implanted medical devices. No participants were excluded in this study.

A G*Power analysis based on a VSPT study by Kessler and Thomson (2010) showed to achieve sufficient power (0.80) at a significance level of .05, 24 participants were recommended (Faul et al., 2007). Ethical approval for the experiment was provided by the RMIT Human Research Ethics Committee, and all participants provided written informed consent.

Materials

A pre-session Qualtrics survey was used to gather demographic information including age, gender, education level, technology use in work or study, and exclusion criteria.

Spatial Perspective Taking

The computerised Spatial Orientation Test (SOT; Friedman et al., 2021) 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 various spatial perspectives. An item example can be seen in Appendix A. This test is commonly applied in research on spatial cognition (Hegarty & Waller, 2004; Friedman et al., 2021). To assist with interaction of outliers arising from inherent deficits in spatial ability, if a participant was deemed as an outlier here as well as in the Left-Right task, they were excluded from the final analyses.

Left-Right Task

AR Interface. The Microsoft HoloLens 2 (Microsoft Corporation, 2023) was used as the hardware. Unity (Unity Technologies, 2022) was used to develop the stimuli and deploy the experiment on the HoloLens 2. A demonstration of the interface and its capabilities can be seen in Appendix B.

Stimuli and Design. The current study employed a repeated measures experimental design. The independent variables were avatar-participant angular disparity (40º, 80º, 120º, 160º) and participant body posture (neutral, congruent, incongruent). High angles were defined as being above 90º and low angles below 90º. The stimuli were adapted from Kessler and Thomson (2010), an example adapted for AR can be seen in Appendix B. Because stimuli appeared to scale and in situ, the avatar angle of 0º would place the participant directly behind the stimuli with no view of the target objects. Therefore, this angle was excluded. Standing at 0º at a table, participants turned their body front on, right by 60º, or left by 60º, while keeping their head facing 0º. Neutral postures were all front on, right and left body postures were deemed congruent if the participant and avatar left/right matched and incongruent if not. The dependent variables were RT (ms) defined as the time of stimulus onset to the moment of response indication, and accuracy recorded as correct or incorrect responses.

Each trial, the participants saw the avatar sitting at the table with two target objects. The target objects were a flower and gun. These were chosen because they are semantically and visually distinct, minimising the chances participants would confuse them. Participants pressed a corresponding left/right arrow keystroke indicating which side the target was on from the avatar's perspective.

Procedure

Participants completed a pre-session survey to obtain demographic information. The SOT was administered using a computer monitor. A time limit of 5 minutes to complete 12 items was set and average angular error was calculated for each participant. Task order between SOT and Left-Right task were balanced across participants.

A short induction to the AR hardware was then provided with three practice trials of the Left-Right task. Participants were verbally instructed to take the 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. Trials began with posture and target object instructions. Participants verbally indicated they were ready to begin the trial. Trials began with a fixation cross displayed on the table's centre for 500ms, followed immediately by the experimental stimulus. 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.

Results

Participants' average angular error on the SOT ranged from 5.70º to 97.83º (M = 32.18º, SD = 22.89º). No participant was an outlier on both the SOT and Left-Right task, so none were excluded on this basis. For the Left-Right task, four individual datapoints were treated as missing due to equipment malfunction. Outliers were removed where RT was less than 100ms or greater than two times the standard deviation of an individual's mean RT. This was done as it is theoretically impossible to react in under 100ms and to account for attentional issues (Ratcliff, 1993). Of the 576 data points this left 554. A multivariate analysis of variance (MANOVA) was planned to assess the effect of angular disparity and posture on RT and accuracy.

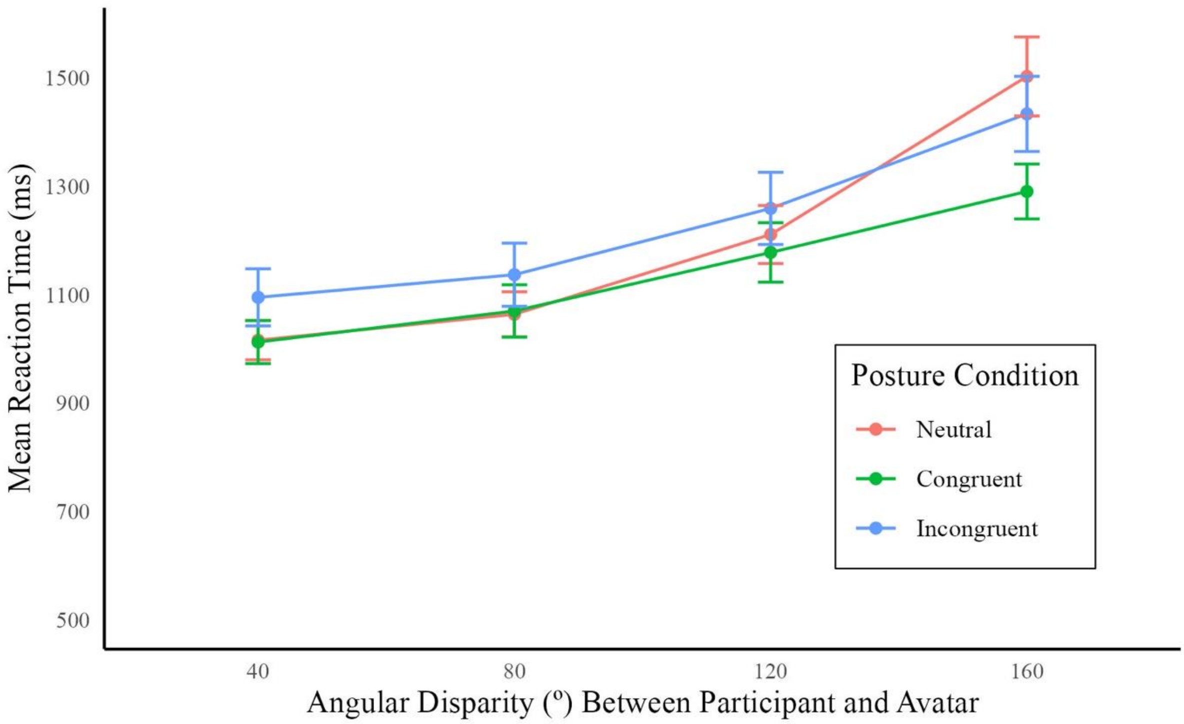
Assumptions for a MANOVA were tested which indicated the data were not normally distributed. Additionally, the repeated measures violated the independent observations assumption. Because RT data is usually positively skewed, representing a combination of Gaussian and exponential distributions, some have suggested cut-offs or transformations to normalise the distribution (Ratcliff, 1993). However, other literature suggests including the whole distribution is more statistically and theoretically sound (Lo & Andrews, 2015). The latter approach was chosen to preserve as much of the data as possible. Lo and Andrews (2015) have suggested the use of generalised linear mixed models (GLMM) as it allows for non-normal dsitrbutions and repeated measures. Meteyard and Davies (2020) and Bolker et al. (2009) have set out best practice protocols for reporting these models. These protocols were followed here.

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 (Posit Team, 2023). Predictor variables for both were avatar-participant angular disparity (40º, 80º, 120º, 160º) 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 & Andrews, 2015).

It was determined whether clockwise or anticlockwise avatar-participant angular disparity affected RT. A Wilcoxon Rank-Sum test determined neither direction to have a significant effect on RT ($W = 29680$, $p = .376$), and so were treated as being the same. As seen in Figure 3, mean RTs were longer at higher angles, neutral postures, and incongruent postures, indicating participants took longer to decide.

Figure 3

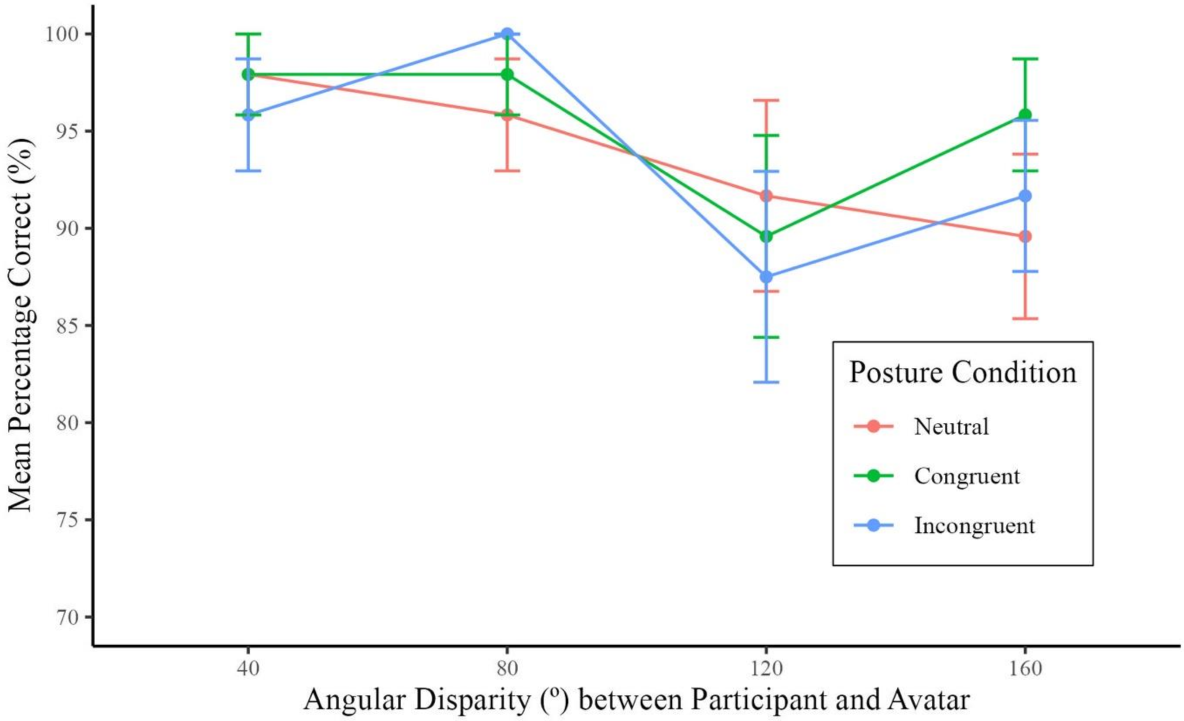
Mean Reaction Time for Angular Disparity by Posture



Overall percentage correct was calculated for each angle and posture condition. Figure 4 depicts shows that participants response accuracy was generally high. It also appears there is a decrease at higher angles, meaning that participants were less accurate in their responses.

Figure 4

Mean Percentage Correct for Angular Disparity by Posture



Reaction Time

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 x 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.

Table 1

Generalised Linear Mixed Modelling Summary Scores for Model Comparison and Model Building

Model Specification	Model name	Nested Model	Fixed effects added	Random Effects	Model Fit		
					AIC	BIC	LL
RT Only	Null	-	-	Subject	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 x Posture	Angle Only	+Angle x Posture	"	-156.96	-96.52	92.48
Interaction	Angle + Posture	Angle Only	+Angle + Posture	"	-163.88	-129.34	89.94

Note. AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, LL = Log Likelihood, df = Degrees of Freedom, X2 = Chi-Square Statistic, LRT = Likelihood Ratio Test.

R^2 for GLMM was calculated (Nakagawa & Schielzeth, 2012). Marginal R^2 for the selected Angle + Posture model was .07, indicating approximately 7% of the variance in RT was explained by the fixed effects in the model. The conditional R^2 was .16, indicating approximately 16% of the variance in RT was explained by both fixed and random effects in the model. As seen in Table 2, the model intercept set as 40º and neutral body posture was statistically significant. The angle of 80º was not statistically significant. Angle 120º was significant with a 0.13 second increase in RT, and angle 160º was a significant predictor with a 0.31 second increase in RT. Incongruent postures were not statistically significant, but congruent postures were with a 0.04 second decrease in RT. Variance for the random effect of participant was 0.01 (SD = 0.13).

Table 2

Selected Model Reporting for Reaction Time Fixed Effects

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

Note. Upper and lower confidence intervals (CI) have been calculated using the Wald method. Model equation: RT ~ Angle + Posture + (1 | Subject).

Accuracy

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 x 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.

Table 3

Generalised Linear Mixed Modelling Summary Scores for Model Comparison and Model Building

Model Specification	Model name	Nested Model	Fixed effects added	Random Effects	Model Fit		
					AIC	BIC	LL
ACC Only	Null	-	-	Subject	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 x Posture	Angle Only	+Angle x Posture	"	212.49	268.61	-95.42
Interaction	Angle + Posture	Angle Only	+Angle + Posture	"	204.83	235.05	-95.42

Note. ACC = Accuracy, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, LL = Log Likelihood, df = Degrees of Freedom, X2 = Chi-Square Statistic, LRT = Likelihood Ratio Test.

Marginal R^2calculated 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 40º was statistically significant. Angle 80º was not a significant predictor of accuracy. However, Angle 120º was a significant predictor with a 1.69 decrease in accuracy score. Angle 160º was not a significant predictor. Variance for the random effect of participant was 1.21 (SD = 1.10).

Table 4

Selected Model Reporting for Accuracy Fixed Effects

Fixed Effects	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

Note. Upper and lower confidence intervals (CI) have been calculated using the Wald method. Model equation: Accuracy ~ Angle + (1 | Subject).

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 ($r_s(70) = -.18$, $p = .002$).

Discussion

Current studies on empathy in AR have focused on self-report measures (Foxman et al., 2021). Adapting a behavioural decision measure for the lower-level VSPT is important for further understanding of AR as an intervention tool in higher-level abilities like empathy. Further, AR presents an opportunity to increase the ecological validity of current VSPT tasks. The main goal of this study was to observe VSPT in AR. This was achieved through adapting Kessler and Thomson's (2010) version of the Left-Right task for AR.

The present results support H1, with higher angles predicting a greater increase in RTs. This suggests participants found these angles more difficult. This study found partial support for H2, with no significant difference between incongruent and neutral postures predicting RT, but congruent postures showing a significant decrease in RT. This suggests that congruent postures facilitated the decision-making process for participants. H3 was only partially supported by these results, with only 120º showing a significant decrease in accuracy. This indicates participants found the angle of 120º more difficult. H4 was not supported by the present results, with no posture condition significantly predicting accuracy. This suggests posture did not interfere with participants' ability to decide. Overall, the main finding of this study is that VSPT can be observed in AR. This is demonstrated by an increase in difficulty with longer RTs at angles above 90º, and a facilitation effect for congruent postures. These results are similar to Kessler and Thomson (2010). The implication here is that this lower-level component of cognitive empathy might be utilised in virtual environments, adding further understanding to studies on empathy, social cognition, and virtual technology.

Converging with the literature (Michelon & Zacks, 2006; Kessler & Thomson, 2010), a difference in RT between angles above and below 90º emerged. This suggests the two levels of VSPT were occurring in AR, where participants used Level 1for lower angles and Level 2 for higher angles to make inferences. This delineation between the levels of VSPT shows that greater mental transformations are required for greater differences in perspectives. The significance of this occurring in AR is that it demonstrates social cognitive abilities like VSPT can replicate in virtual environments in similar ways to monitor based experiments. For congruent postures to facilitate faster RTs suggests that more similar postures between an observer and another make it easier to infer their perspective, namely because their body is already in the required direction of mental rotation. This reflects findings by Kessler and Thomson (2010), supporting the explanation that a mental rotation of the body schema is a method for Level-2 VSPT. However, Kessler and Thomson (2010) found incongruent postures to increase RTs compared to neural postures. With incongruent and neutral postures not being significantly different in this study, this suggests incongruent postures may not have been incongruent enough. Especially in the context of AR, where participants had more lifelike and to scale information about an avatar's perspective compared to an image on monitor.

The present study differs from Kessler and Thomson (2010) who found a significant decrease in accuracy at higher angles. Here, only angle 120º produced a significant error rate which recuperated at the higher angle of 160º. 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 120º, especially considering it also increased RT. Because participants still responded greater than chance at 120º, it is reasonable to conclude they were applying some method of VSPT to make their inference. However, given the recuperation at 160º, 120º may be a threshold for difference in perspectives, and differences beyond this require more careful inference.

Overall, 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. 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 and their distributions (Lindemann & Fischer, 2023; Lo & 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 implications of these findings for research applications are that they aid in substantiating claims of effect for the use of AR as an intervention tool in cognitive empathy and broader social cognition. In experiments that have asked people to take the psychological perspective of another in AR, these findings aid in describing how this operates. This also reflects Decety and Lamm's (2004) model, where top-down abilities like psychological perspective taking are informed by bottom-up information processing. 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 it 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).

The implications of these findings for research on VSPT include an increase in ecological validity. By examining this ability in a 3D, real environment, while maintaining experimental control of the stimuli there is a greater degree of generalisation for VSPT. This study suggests that both levels of VSPT and the embodied mental rotation occur in the real environment. However, it may also reveal that altering body posture in this context may not be sufficient to confidently infer the involvement of the body schema. This is due to incongruent postures not being different from neutral postures here. This may be because participants have more accurate information about body position in this 3D environment. However, a facilitation effect was observed for congruent postures, which indicates body position matters in understanding another's perspective.

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 & Fischer, 2023).

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. Further, it may also clarify how information about posture and body position is used in VSPT. 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 include comparing performance across these two apparatuses.

Conclusion

The current research suggests that VSPT occurs in AR, meaning that people can take the visual perspective of another person 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, that 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 confirms 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, and that this may be highly relevant in the effectiveness of these interfaces as intervention tools.

References

Alutaroma. (2019, August). *Low poly tulip free sample.* CGTrader. https://www.cgtrader.com/free-3d-models/plant/flower/low-poly-tulip-free-sample

Arena, F., Collotta, M., Pau, G., & Termine, F. (2022). An overview of augmented reality. *Computers, 11*(2), Article e28. https://doi.org/10.3390/computers11020028

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Baumeister, R.F., Vohs, K.D., & Funder, D.C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspective on Psychological Science, 2*(4), 396 – 403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Berger, C.C., Lin, B., Lenggenhager, B., Lanier, J. & Gonzalez-Franco, M. (2022). Follow your nose: Extended arm reach after Pinocchio illusion in virtual reality. *Frontiers in Virtual Reality, 3*, Article e712375. https://doi.org/10.3389/frvir.2022.712375

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., & White, J.S.S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution, 24*(3), 127 – 135. https://doi.org/10.1016/j.tree.2008.10.008

Bombari, D., Mast, M.S., Canadas, E., & Bachmann, M. (2015). Studying social interactions through immersive virtual environment technology: Virtues, pitfalls, and future challenges. *Frontiers in Psychology, 6,* Article e869. https://doi.org/10.3389/fpsyg.2015.00869

Bowman, D.A., & McMahan, R.P. (2007). Virtual reality: How much immersion is enough? *Computer, 40*(7), 36 – 43. https://doi.org/10.1109/MC.2007.257

Caldwell, M.P., Cheung, H., Cheung, S.K., Li, J.B., & Siu, T.S.C.(2022). Visuospatial perspective-taking in social-emotional development: enhancing young children's mind and emotion understanding via block building training. *BMC Psychology, 10*(1), Article e264. https://doi.org/10.1186/s40359-022-00976-5

Cipresso, P, Giglioli, I.A.C., Raya, M.A., & Giuseppe, R. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology, 9*, Article e2086. https://doi.org/10.3389/fpsyg.2018.02086

Dang, J., King, K.M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences, 24*(4), 267 – 269. https://doi.org/10.1016/j.tics.2020.01.007

D'Angelo, M., Di Pellegrino, G., Seriani, S., Gallina, P., & Frassinetti, F. (2018). The sense of agency shapes body schema and peripersonal space. *Scientific Reports, 8,* Article e13847. https://doi.org/10.1038/s41598-018-32238-z

Decety, J., & Jackson, P.L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 71 – 100. https://doi.org/10.1177/1534582304267187

Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *The Scientific World Journal, 6,* 1146 – 1163. https://doi.org/10.1100/tsw.2006.221

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 176-191. https://doi.org/10.3758/bf03193146

Fiehler, K., & Karimpur, H. (2023). Spatial coding for action across spatial scales. *Nature Reviews Psychology, 2*, 72-84. https://doi.org/10.1038/s44159-022-00140-1

Fischer, M.H. (2023). The embodied cognition approach: Principles and research questions. In A. Felisatti, & M.H. Fischer (Eds), *Experimental methods in embodied cognition: How cognitive psychologists approach embodiment* (pp. 3 – 18). Taylor & Francis Group. https://doi.org/10.4324/9781003290698-2

Foxman, M., Markowitz, D.M., & Davis, D.Z. (2021). Defining empathy: Interconnected discourses of virtual reality's prosocial impact. *New Media & Society, 23*(8), 2167-2188. https://doi.org/10.1177/1461444821993120

Friedman, A., Kohler, B., Gunalp, P., Boone, A.P., & Hegarty, M. (2020). A computerised spatial orientation test. *Behavior Research Methods, 52*(2), 799-812. https://doi.org/10.3758/s13428-019-01277-3

Gainsford, K., Fitzgibbon, B., Fitzgerald, P.B., & Hoy, K.E. (2020). Transforming treatments for schizophrenia: Virtual reality, brain stimulation and social cognition. *Psychiatry Research, 288*, Article e112974. https://doi.org/10.1016/j.psychres.2020.112974

Geraets, C.N.W., Tuente, S.K., Lestestuiver, B.P., van Beilen, M., Nijman, S.A., Marsman, J.B.C., & Veling, W. (2021). Virtual reality facial emotion recognition in social environments: An eye-tracking study. *Internet Interventions, 25,* Article e100432. https://doi.org/10.1016/j.invent.2021.100432

Giglioli, I.A.C., Pallavicini, F., Pedroli, E., Serion, S., & Riva, G. (2015). Augmented reality: A brand new challenge for the assessment and treatment of psychological disorders. *Computational and Mathematical Methods in Medicine,* Article e862942. https://doi.org/10.1155/2015/862942

Goldman, A., & de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences, 13*(4), 154 – 159. https://doi.org/10.1016/j.tics.2009.01.007

Gronholm, P.C., Flynn, M., Edmonds, C.J., & Gardner, M.R. (2012). Empathic and non-empathic routes to visuospatial perspective-taking. *Consciousness and Cognition, 21*(1), 494 – 500. https://doi.org/10.1016.j.concog.2011.12.004

Guarese, R., Pretty, E., Fayek, H., Zambetta, F., & van Schyndel, R. (2023). Evoking empathy with visually impaired people through an augmented reality embodiment experience. *Proceedings of 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR),* 184 – 193. https://doi.org/10.1109/VR55154.2023.00034

Gunia, A., Moraresku, S., & Vlček, K. (2021). Brain mechanisms of visuospatial perspective-taking in relation to object mental rotation and the theory of mind. *Behavioural Brain Research, 407,* Article e113247. https://doi.org/10.1016/j.bbr.2021.113247

Hall, J.A., & Schwartz, R. (2018). Empathy present and future. *The Journal of Social Psychology, 159*(3), 225-243. https://doi.org/10.1080/00224545.2018.1477442

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking abilities. *Intelligence, 32*(2), 175-191. https://doi.org/10.1016/j.intell.2003.12.001

Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zaki, J. (2018). Building long-term empathy: A large-scale comparison of traditional virtual reality perspective-taking. *PLoS ONE*, *13*(10), Article e0204494. https://doi.org/10.1371/journal.pone.0204494

Holmes, N.P., & Spence, C. (2004). The body schema and the multisensory representation(s) of peripersonal space. *Cognitive Processing, 5*(2), 94-105. https://doi.org/10.1007/s10339-004-0013-3

Kessler, K., & Thomson, L.A. (2010). The embodied nature of spatial perspective-taking: Embodied transformation versus sensorimotor interference. *Cognition, 114*(1), 72-88. https://doi.org/10.1016/j.cognition.2009.08.015

Kothgassner, O.D., & Felnhofer, A. (2020). Does virtual reality help to cut the Gordian knot between ecological validity and experimental control? *Annals of the International Communication Association, 44*(3), 210-218. https://doi.org/10.1080/23808985.2020.1792790

Kozhevnikov, M., & Hegarty, M. (2001). A dissociated between object manipulation spatial ability and spatial orientation ability. *Memory and Cognition, 29*(5), 745-756. https://doi.org/10.3758/bf03200477

Lindblom, J. (2020). A radical reassessment of the body in social cognition. *Frontiers in Psychology, 11*, Article e987. https://doi.org/10.3389/fpsyg.2020.00987

Lindemann, O., & Fischer, M.H. (2023). Chronometric methods: Reaction times and movement times. In A. Felisatti, & M.H. Fischer (Eds), *Experimental methods in embodied cognition: How cognitive psychologists approach embodiment* (pp. 37 – 48). Taylor & Francis Group. https://doi.org/10.4324/9781003290698-5

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontier in Psychology, 6*, Article e1171. https://doi.org/10.3389/fpsyg.2015.01171

Lohaus, T., Rogalla, S., & Thoma, P. (2023). Use of technologies in the therapy of social cognition deficits in neurological and mental diseases: A systematic review. *Telemedicine and e-Health, 29*(3), 331-351. https://doi.org/10.1089/tmj.2022.0037

McDonald, S. (2013). Impairments in social cognition following severe traumatic brain injury. *Journal of International Neuropsychological Society, 19*(3), 231-246. https://doi.org/10.1017/S1355617712001506

Meteyard L., & Davies, R.A.I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language, 112,* Article e104092. https://doi.org/10.1016/j.jml.2020.104092

Michelon, P., & Zacks, J.M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics, 68*(2), 327-337. https://doi.org/10.3758/BF03193680

Microsoft Corporation. (2023, April 20). Microsoft HoloLens 2. https://www.microsoft.com/en-au/hololens

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. *Proceedings of SPIE Telemanipulator and Telepresence Technologies, 2351,* 282 – 292. https://doi.org/10.1117/12.197321

Morasso, P., Casadio, M., Mohan, V., Rea, F., & Zenzeri, J. (2015). Revisiting the body-schema concept in the context of whole-body postural-focal dynamics. *Frontiers in Human Neuroscience, 9*, Article e83. https://doi.org/10.3389/fnhum.2015.00083

Nakamura, S., Mochizuki, N., Konno, T., Yoda, J., & Hashimoto, H. (2016). Research on updating of body schema using AR limb and measurement of the updated value. *IEEE Systems Journal, 10*(3), 903-911. https://doi.org/10.1109/JSYST.2014.2373394

Pan, X., & Hamilton, A.F.C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology, 109*(3), 395 – 417. https://doi.org/10.1111/bjop.12290

Posit Team. (2023). RStudio: Integrated development environment for R (2023.6.2.561 version). http://www.posit.co

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, *1*(4), 515-526. https://doi.org/10.1017/S0140525X00076512

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510-532. https://doi.org/10.1037/0033-2909.114.3.510

Samuel, S., Cole, G.G., & Eacott, M.J. (2023). It's not you, it's me: A review of individual differences in visuospatial perspective taking. *Perspectives on Psychological Science, 18*(2), 293-308. https://doi.org/10.1177/17456916221094545

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16*(2). 235 – 239. https://doi.org/10.1016/j.conb.2006.03.001

Schrier, K., & Farber, M. (2021). A systematic literature review of 'empathy' and 'games'. *Journal of Gaming & Virtual Worlds, 13*(2), 195-214. https://doi.org/10.1386/jgvw_00036_1

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,* Article e712. https://doi.org/10.3389/fnhum.2013.00712

Schurz, M., Radua, J., Tholen, M.G., Milske, L., Marguiles, D.S., Mars, R.B., Sallet, J., & Kanske, P. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin, 147*(3), 293-327. http://doi.org/10.1037/bul0000303

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences, 1156,* 81 – 96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science, 10*(2), 199-204. https://doi.org/10.1111/j.1467-7687.2007.00580.x

Sunahara, C.S., Rosenfield, D., Alvi, T., Wallmark, Z., Lee, J., Fulford, D., & Tabak, B.A. (2022). Revisiting the association between self-reported empath and behavioral assessments of social cognition. *Journal of Experimental Psychology: General, 151*(12), 3304 – 3322. https://doi.org/10.1037/xge0001226

Tay, J.L., Xie, H., & 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. *Journal of Personalized Medicine, 13*(1), Article e12. https://doi.org/10.3390/jpm13010112

Ueda, S., Nagamachi, K., Nakamura, J., Sugimoto, M., Inami, M., & 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*(12), Article e0261063. https://doi.org/10.1371/journal.pone.0261063

Unity Technologies. (2022). *Unity 3D* (Version 2020.3.29f1) [VR Software]. https://unity.com/releases/editor/whats-new/2020.3.29

Ventura, S., Banos, R.M., & Botella, C. (2018). Virtual and augmented reality: New frontiers for clinical psychology. In N. Mohamudally (Ed), *State of the Art Virtual Reality and Augmented Reality Knowhow* (pp. 99 – 118). IntechOpen. https://directory.doabooks.org/handle/20.500.12854/59983

Vu Studios. (2015, May). *Relaxed man character.* Unity Asset Store. https://assetstore.unity.com/packages/3d/characters/humanoids/fantasy/relaxed-man-character-32665#publisher

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128. https://doi.org/10.1016/0010-0277(83)90004-5

Youtipie. (2022, September). *Desert eagle*. CGTrader. https://www.cgtrader.com/free-3d-models/military/gun/desert-eagle-648d0830-72b8-47af-a111-e3bf0663f451

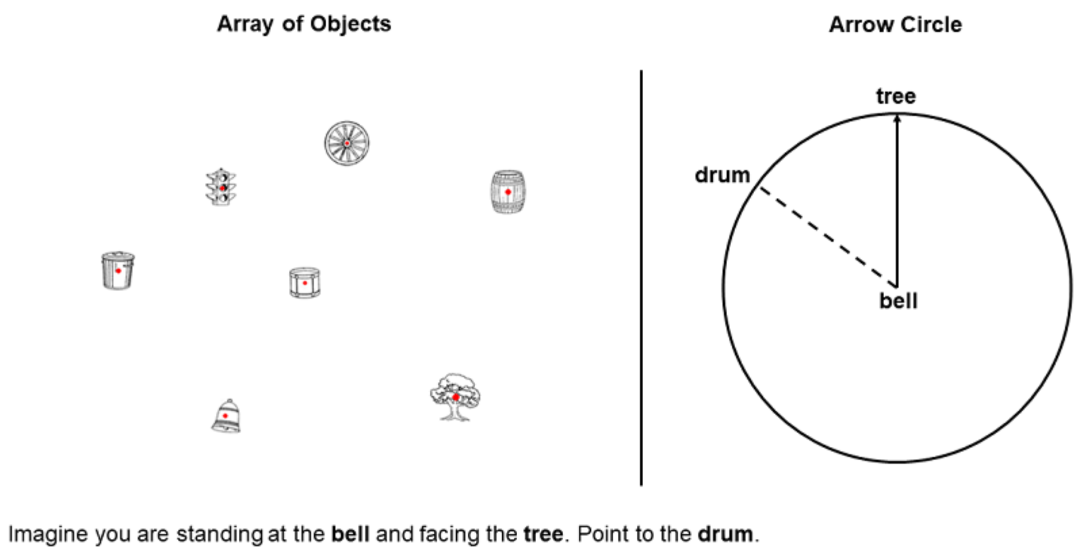
Zaki, J., & Ochsner, K. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience, 15,* 675 – 680. https://doi.org/10.1038/nn.3085

Appendix A

Spatial Orientation Test

Figure 1

Sample of Computerised Spatial Orientation Test



Note. The object array was displayed in this configuration in all trials. The array and arrow circle were displayed side-by-side each trial and is how participants indicated their answer. The difference between their indicated answer the real angle was the angular error. From "A computerised spatial orientation test", by A. Friedman, B. Kohler, P. Gunlap, A.P. Boone, and M. Hegarty, 2020, *Behavior Research Methods, 52*(2), p. 799 (https://doi.org/10.3758/s13428-019-01277-3). Copyright 2019 by The Psychonomic Society Inc.

Appendix B

HoloLens

Figure 1

Image of the Microsoft HoloLens 2

Note. When the device is worn virtual elements appear in the field of view. Elements and interactions with them are tracked by the device. Adapted from *Microsoft HoloLens 2,* by Microsoft Corporation, 2023 (https://www.microsoft.com/en-au/hololens). Copyright 2023 by Microsoft.

Figure 2

Stimulus for Left-Right Task in Augmented Reality

Note. Screenshot of the stimulus participants saw while wearing the Microsoft HoloLens 2. 3D models adapted from *Relaxed Man Character,* by VuStudios, 2015 (https://assetstore.unity.com/packages/3d/characters/humanoids/fantasy/relaxed-man-character-32665). Copyright 2023 by Unity Technologies; *Low poly tulip free sample,* by Alutaroma, 2019 (https://www.cgtrader.com/free-3d-models/plant/flower/low-poly-tulip-free-sample). Copyright 2023 by CGTrader; *Desert Eagle,* by Youtipie, 2022 (https://www.cgtrader.com/free-3d-models/military/gun/desert-eagle-648d0830-72b8-47af-a111-e3bf0663f451). Copyright 2023 by CGTrader.