



This is a peer-reviewed, final published version of the following document, COPYRIGHT: © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. and is licensed under Creative Commons: Attribution 4.0 license:

**Ford, Ben ORCID logoORCID: <https://orcid.org/0000-0001-7502-7505>, Monk, Rebecca, Litchfield, Damien and Qureshi, Adam (2024) Agent-Object Relationships in Level-2 Visual Perspective Taking: An Eye-Tracking Study. *Journal of Cognition*, 7 (1). pp. 1-17. doi:10.5334/joc.398**

Official URL: <http://dx.doi.org/10.5334/joc.398>  
DOI: <http://dx.doi.org/10.5334/joc.398>  
EPrint URI: <https://eprints.glos.ac.uk/id/eprint/14449>

#### **Disclaimer**

The University of Gloucestershire has obtained warranties from all depositors as to their title in the material deposited and as to their right to deposit such material.

The University of Gloucestershire makes no representation or warranties of commercial utility, title, or fitness for a particular purpose or any other warranty, express or implied in respect of any material deposited.

The University of Gloucestershire makes no representation that the use of the materials will not infringe any patent, copyright, trademark or other property or proprietary rights.

The University of Gloucestershire accepts no liability for any infringement of intellectual property rights in any material deposited but will remove such material from public view pending investigation in the event of an allegation of any such infringement.

PLEASE SCROLL DOWN FOR TEXT.



# Agent-Object Relationships in Level-2 Visual Perspective Taking: An Eye-Tracking Study

RESEARCH ARTICLE

**BEN FORD** 

**REBECCA MONK** 

**DAMIEN LITCHFIELD** 

**ADAM QURESHI** 

\*Author affiliations can be found in the back matter of this article

 ubiquity press

## ABSTRACT

Visual perspective taking (VPT) generates a shared frame of reference for understanding how the world appears to others. Whilst greater cognitive and neurophysiological demands are associated with increasing angular distance between the self and other is well documented, accompanying attentional characteristics are not currently understood. Furthermore, although age and group status have been shown to impact task performance, other important cues, such as the relationship between agents and objects, have not been manipulated. Therefore, 35 university students participated in an eye-tracking experiment where they completed a VPT task with agents positioned at a low or high angular disparity (45° or 135° respectively). The congruence between the age of the agent (child vs adult) and the object they are attending to (e.g., teddy-bear vs kettle) was also manipulated. Participants were required to respond to the direction of the object from the agent's position. The findings reveal more fixations and increased dwell-times on agents compared to objects, but this was moderated by the age of the task agent. Results also showed more attentional transitions between agents and objects at higher angular disparities. These results converge with behavioural and neurophysiological descriptions of task performance in previous studies. Furthermore, the congruency of the relationship between agents and objects also impacted attention shifting and response times, highlighting the importance of understanding how social cues and contexts can modulate VPT processes in everyday contexts and social interaction.

## CORRESPONDING AUTHOR:

**Ben Ford**

University of Gloucestershire,  
UK; Edge Hill University, UK  
bford6@glos.ac.uk

---

## KEYWORDS:

visual perspective taking;  
level-2; agent object  
relationships; object  
congruency; intentionality; eye-  
tracking; stereotype

## TO CITE THIS ARTICLE:

Ford, B., Monk, R., Litchfield, D.,  
& Qureshi, A. (2024). Agent-  
Object Relationships in Level-2  
Visual Perspective Taking: An  
Eye-Tracking Study. *Journal of  
Cognition*, 7(1): 72, pp. 1–17.  
<https://doi.org/10.5334/joc.398>

Representing the visual perspectives of others is a fundamental sociocognitive process. Evidence supports a distinction between the ‘level 1’ computations which are fast, involuntary, and required for visibility judgments (L1-VPT; Samson et al., 2010), and those for ‘level 2’, which are slower and involve a deliberate mental translation into the other’s position in order to represent *how* the world appears visually or spatially (L2-VPT; Flavell et al., 1981; Kessler & Thomson, 2010; Kessler & Rutherford, 2010; Kessler & Wang, 2012). Behavioural and neurophysiological studies of L2-VPT demonstrate that the computations required depend on the task agent’s degree of rotation away from a shared line-of-sight (Kessler & Thompson, 2010). At lower angles, participants can use their egocentric reference frame because the self and other perspectives are broadly similar. However, with angular disparity above 80°, participants perform a mental self-rotation into the position of the other person (Kessler & Thompson, 2010; Surtees et al., 2013a; Surtees et al., 2013b). This rotation required at higher angular disparities increases cognitive effort and is also evident in greater neurophysiological activity across a broad frontoparietal network (Seymour et al., 2018). However, despite cognitive and neurophysiological evidence, whether there are differences in attention allocation between low and high angular disparities is yet to be understood.

To-date, no eye-tracking experiments have been conducted with L2-VPT, yet eye-tracking data from L1-VPT and related processes such as gaze-cueing reveal significant effects of task manipulations. For example, work on gaze-cueing reveals preferential attention to in-group others (for a review see Kawakami et al., 2018) whether determined by minimal groups (Kawakami et al., 2014, study 2), race (Wu et al., 2012; Kawakami et al., 2014, studies 1 and 3), or age (Denkinger & Kinn, 2018; Strickland-Hughes et al., 2020). As for L1-VPT, when participants need to switch between self and other perspectives, compared to when perspective cues remained constant (e.g., always ‘YOU’ trials), participants seem to experience interference from the task agent’s differing perspective (Ferguson et al., 2017). Ferguson et al. (2017) also reveal a greater number of fixations on switching trials and suggested that the broader visual search required may modulate this type of interference. Further, Cane et al. (2017) found that under high working memory load and when perspectives differed, participants made earlier fixation shifts towards a cued object, shifts which were not evident when perspectives were shared. This suggests that participants were incorporating their privileged knowledge about the perspective of the other into subsequent processing. Whilst eye-tracking research reveals significant effects of agent cues (Kawakami et al., 2018), perspective cues (Ferguson et al., 2017) and cognitive load (Cane et al., 2017), no research has examined attentional allocation in a L2-VPT task, nor how this may be moderated by angular differences to the task agent’s position. Thus, the first aim of this study is to provide converging evidence for the distinction between low and high angular judgements by assessing whether there are differences in the number of fixations, the number of transitions between the object and agents, and the proportion of time spent looking at either the object or agent. Doing so will supplement our current understanding of cognitive and neurophysiological mechanisms in L2-VPT.

Understanding the influence of social cues on perspective representation is also gaining increasing attention. To date, VPT has been shown to be affected by social factors such as group status (Savitsky et al., 2011; Todd et al., 2011; Simpson & Todd, 2017; Ferguson et al., 2018), humanness of agents (Ye et al., 2020) and avatar emotion (Monk et al., 2020). For example, divergent findings either report that similarity to task agents facilitates perspective representation (Ye et al., 2020) or creates egocentric interference (Savitsky et al., 2011; Simpson & Todd, 2017; Todd et al., 2011; Ferguson et al., 2018). More generally, these results suggest that the other person in VPT tasks is being considered as a social agent, with their attributes exerting a top-down effect on representing visual perspectives. However, while the extant literature to date has employed VPT tasks that present agents and their attentional-targets (e.g., the letters, objects or room features the agents are ‘looking’ towards), no research has examined whether the relationship between the agent and attentional-target impacts subsequent processing or attention allocation.

A close thematic relationship between agent and object (e.g., ‘doctor’ and ‘stethoscope’) results from the cooccurrence of one or multiple semantic features which are processed by a general representational mechanism (Xu et al., 2018). These thematic relationships are represented quickly and implicitly, are based on first-hand experience of objects and events (Estes et al.,

2011; Chaigneau et al., 2009; Kalénine et al., 2012) and have been shown to support learning, memory, and language comprehension (Estes et al., 2011). Therefore, it is possible that a close thematic relationship could also facilitate VPT performance. Furthermore, the close relationship between an agent and object could be a salient indicator of intentionality (Estes et al., 2011). For example, Chaigneau et al. (2009) found that compared to when objects are presented alone (e.g., projectiles), participants' ability to respond with the object's function is improved by providing contextual cues (e.g., a catapult) along with information about how they will be used. Intentionality judgments are important here because they come under the umbrella of Theory-of-Mind (ToM; Leslie, 1987) which plays a fundamental role in successful communication (Weiss & Harris, 2001; Hughes & Leekam, 2004; Brüne, 2005) and has previously been related to VPT (Hamilton et al., 2009; Dumontheil et al., 2010). In fact, observing others not only reveals their intentions but allows us to predict their future interaction with the environment (Frith & Frith, 2006) and shared intentionality is considered by some as the basis of ToM (Tomasello, 2018).

Crucially for VPT tasks, where an agent typically looks at an object in front of them, Becchio et al.'s (2008) review concluded that gaze alone can lead to the ascription of intentionality from an agent to an object. This 'intentional imposition' occurs when the target of another's gaze is endowed with properties it does not inherently show and is inferred from social affordances which arise from a combination of the automatic activations evoked by perceiving objects (object affordances, e.g., a toy affords being played with) and a context or intentional action (e.g., reaching for the toy; Gong et al., 2021). Whilst dynamic actions embedded in contexts yield a greater haemodynamic response in premotor mirror neurons (Iacoboni et al., 2005), an actual movement or vocalisation is not necessary to mentalise about others' intentions; gaze alone is sufficient (Castiello, 2003; Teufel et al., 2010a; Teufel et al., 2010b; Morgan et al., 2018; Morgan et al., 2021; although, see Cole et al., 2015 and Ward et al., 2020). Therefore, visual perspective tasks where participants simply view an agent looking at objects provide a befitting paradigm to manipulate inferred intentionality without having to make this potential action explicit (Nummenmaa & Calder, 2009). Furthermore, research examining semantically incongruent information, such as viewing a microscope on a kitchen worktop, has been related to attentional measures such as the number and duration of fixations (Henderson et al., 1999; Vö & Henderson, 2009). Yet, it remains unknown whether thematic relationships influence attention allocation in L2-VPT. As VPT requires attentional reorienting, incorporating eye-tracking to reveal the attentional impact agent-object relationships is pertinent and represents a novel and significant contribution to the existing literature on L2-VPT.

In summary, this pre-registered study (<https://osf.io/8gdm6>) project uses an established L2-VPT task manipulating angular disparity of the task agent whilst also manipulating whether the agent is an adult or child and whether the object attended is adult or child related. It seeks to understand whether the age of a task agent or agent-object relationship significantly affect cognitive demand, as measured by response times and accuracy, and gaze-behaviour, as measured by agent-object dwell-time proportion, overall number of fixations, and attentional transitions between objects and agents. Furthermore, eye-tracking will extend the current evidence that supports distinctive processing at lower and higher angular disparities to include attention allocation.

## **HYPOTHESES**

### **LOW VS HIGH ANGULAR DISPARITIES**

Replicating previous behavioural findings (Michelon & Zacks, 2006; Kessler & Thompson, 2010; Surtees et al., 2013a; Surtees et al., 2013b), we hypothesise that perspective representation at 135° compared to 45° will result in significantly slower response times and lower accuracy rates. To extend current understanding on attentional allocation in L2-VPT, we predict perspective judgements at 135° will require more fixations, more area-of-interest transitions (AOITs) between the agent and target object, and greater proportional focus on the avatar.

### **AVATAR-AGE**

Extending previous ingroup effects on VPT (Savitsky et al., 2011; Todd et al., 2011; Simpson & Todd, 2017; Ferguson et al., 2018; Ye et al., 2020), we expect the visual perspective of adult (ingroup) avatars to be more easily represented, leading to significant quicker response times and significantly less fixations required to process the visuospatial information. We do not expect any avatar-age effects on accuracy, AOITs or dwell-times.

## OBJECT CONGRUENCE

We hypothesise that congruent age-object relationships will result in significantly quicker response times (akin to Xu et al., 2018), less fixations (akin to Henderson et al., 1999; Vö & Henderson, 2009), significantly less AOITs, and less time spent looking at the object relative to the agent. We do not hypothesise any accuracy effects.

## INTERACTION EFFECTS

To supplement the pre-registered hypotheses, and for clarity, there is currently no evidence for any outcome measure to hypothesise an interaction between angular disparity and object congruence, angular disparity and avatar-age, object-congruence and avatar-age, or a three-way interaction.





## METHOD

### PARTICIPANTS

35 participants (27 female) from XXXXX participated for course credit or were recruited via advertisements placed around campus and paid £5.00. 1 participant was removed due to low overall accuracy (<66%). The final sample included 34 participants (26 female;  $M_{age} = 21.06$ ,  $SD_{age} = 4.33$ , Range = 18–35). As per the pre-registration, recruitment was restricted by budget and time but was in line with previous research using similar design and dependent measures (Fergusson et al., 2017). As per Westfall et al. (2014) with a stimuli-within-condition design, a power of .8 and 34 participants, the current experiment was powered to detect a main effect of  $d = .50$ .

### DESIGN

A  $2 \times 2 \times 2$  repeated measures design with three independent variables of angle (low vs high), avatar-age (child vs adult) and object-relevance (relevant vs irrelevant) as factors. Example stimuli can be seen in Figure 1. Dependent measures included accuracy, response times, number of fixations, area of interest transitions and proportionate dwell time.

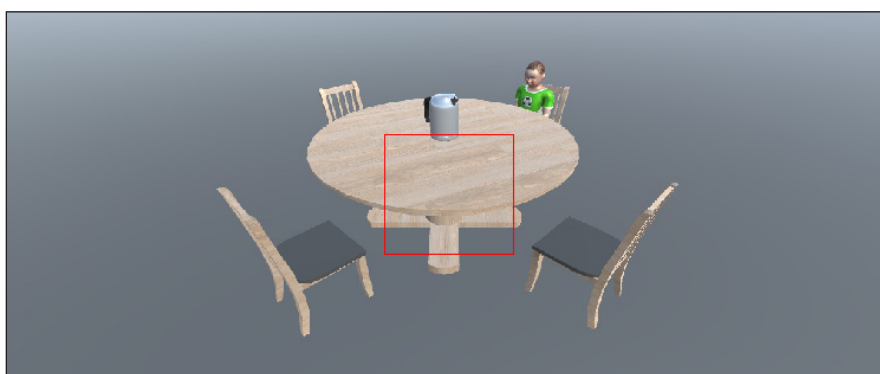
Avatar-Age	Object-Relevance	Angle	
Adult	Congruent	135	
	Incongruent	45	
Child	Congruent	135	
	Incongruent	45	

**Figure 1** Example stimuli to demonstrate manipulations of avatar-age, object congruence and angular disparity.

Participants completed an adapted version of the Kessler and Thomson (2010) task. Stimuli were designed using Unity game engine (Unity Technologies, 2020) and the experiment was programmed in Experiment Builder (SR Research). Original stimuli dimensions were  $1920 \times 1080$ . Stimuli portrayed a circular table from above, with the sight-angle set at  $35^\circ$ . At the table were 4 chairs, positioned at  $45^\circ$ ,  $135^\circ$ ,  $225^\circ$ , and  $315^\circ$ .  $0^\circ$  was the line-of-sight of participants and angular disparity was determined by their rotation around an imaginary arc that borders the target object in the centre of the table. On each stimulus, a human avatar was seated at one of these positions. The avatars gender always matched the participant's gender. Realistic human avatars were used because the distinction between children and adults was integral to the manipulation. Each stimulus had a single object placed in the middle of the table. This object could be child-related (a teddy-bear or toy train), or adult related (a laptop or kettle). These objects were selected because of their spatial properties. Participants were told that left-facing meant a teddy-bear's face pointing leftward, a train's locomotive pointing leftward, a laptop screen facing leftward, and a kettle's spout on the left-side as this is the direction it would pour (see Figure X). Across all stimuli, the child and adult sat in each of the tables positions equally, they were also paired with each object an equal number of times, and all had an equal number of leftward and rightward facing correct responses. All stimuli are available at <https://osf.io/8gdm6>.

## PROCEDURE

Participants rested their heads on a fixed chin rest 60 cm from the 19 inch ( $360 \times 270$  mm) ViewSonic Graphic Series G90fB coloured display screen. The left eye of participants was recorded using an EyeLink 1000+ eye-tracker running at 1000 Hz. Participants were shown example images of the objects facing left and right (see Figure X for some examples). Participants were then shown example experimental stimuli (see Figure X) and told that they had to respond to whether the object was facing right or left from the avatars position using the 'Q' for leftward and 'P' for rightward. Each trial began with a fixation cross lasting 750 ms and was followed by the experimental stimuli that remained on screen until a response was made or 8 seconds had elapsed. The experimental stimuli were  $1024 \times 460$  pixels and appeared centrally on the screen; this corresponds to  $33.4 \times 15.35^\circ$  of visual angle which should mean participants had to move their gaze to focus on the individuals sat at the seating positions (a graphic with the standard  $5^\circ$  of visual angle ( $149.03 \times 149.03$  pixels) laid over an example stimulus can be seen in Figure 2). Participants completed 8 practice trials, before completing 128 experimental trials (16 trials per condition) presented randomly across two equal blocks. Participants completed a standard EyeLink 9-point calibration at the beginning of the experiment, after the practice trials and in the break between the two blocks of trials.



**Figure 2** The red square represents  $5^\circ$  of visual angle from the centre of the screen. The target objects and avatar positions are located beyond the boundaries to encourage eye-movements.

## EYE-TRACKING MEASURES

On each trial, a total of two *Area of Interest Transitions* (AOITs) were specified; one surrounding the object with a constant size of  $82 \times 82$  pixels for all objects and one surrounding the task agents which at higher angular disparities was  $86 \times 123$  pixels and at lower angles  $144 \times 185$  pixels. AOITs are measured as the number of times participants fixated in one AOI before leaving the AOI boundary and fixating in the other AOI. For example, if participants fixated on the agent AOI, shifted attention outside the AOI boundary and fixated on something other than the object AOI before returning to the agent AOI, this would not count as a transition. However, if participants fixating on the agent AOI shifted attention outside the boundary and then fixated on the object

AOI, this would count as one AOIT; if they immediately shifted back to the agent AOI, this would count as a second AOIT. *Proportionate dwell time* is calculated as time fixating on avatar AOI subtracted from total time fixating within object AOI, divided by total trial time and multiplying the outcome by 100. Thus, it acts as a proportionate measure of attention allocation between task agents and task objects relative to total trial time, with positive values indicating greater attention to objects and negative values indicating greater attention to agents. *Total number of fixations* are the sum of all fixations irrespective of where in the scene they were made. Thus, total fixations were not restricted to AOI fixations and therefore inherently linked to response times.

## RESULTS

Factors of angle (45°/135°), avatar-age (child/adult) and age-object relevance (relevant/irrelevant) were entered into a linear mixed effects model for response times, number of fixations, area-of-interest transitions (AOITs) and proportionate dwell times, and logistic regression for accuracy data using the lme4 package (Bates et al., 2015) for R (R Core Team, 2013). Random intercepts were included to control for variation across individuals and stimuli,<sup>1</sup> reducing the potential for type 1 error that can arise if they are treated as fixed (Barr, 2013; Barr et al., 2013). For mixed effects models on number of fixations and area of interest transitions, maximal models and models including a random effect of stimuli led to models with a singular fit. Therefore, for these models, only random effects of individual were included. Post-hoc tests were computed using the ‘emmeans’ package (Lenth et al., 2019) for R and multiple comparisons were controlled using a Bonferroni adjustment. All means and standard deviations for each outcome measure can be found in Table 1 and all statistical outcomes in Tables 2 and 3.

INDEPENDENT VARIABLES			DEPENDENT VARIABLES				
ANGLE	OBJECT RELEVANCE	AGENT AGE	M (sd) RESPONSE TIME (ms)	M (sd) ACCURACY (%)	M (sd) TOTAL FIXATIONS	M (sd) PROPORTIONATE DWELL TIME	M (sd) AREA OF INTEREST TRANSITIONS
45	Relevant	Child	1153 (630)	98.0 (14.1)	3.83 (1.88)	65.2 (36.0)	1.52 (0.54)
		Adult	1180 (568)	97.6 (15.4)	3.84 (1.70)	64.8 (38.0)	1.50 (0.51)
	Irrelevant	Child	1267 (643)	98.3 (12.8)	4.10 (1.85)	61.2 (37.2)	1.58 (0.52)
		Adult	1170 (542)	99.1 (9.56)	3.80 (1.70)	66.3 (34.2)	1.55 (0.51)
135	Relevant	Child	1642 (959)	90.8 (29.0)	4.80 (2.49)	57.1 (36.1)	1.62 (0.53)
		Adult	1650 (917)	93.5 (24.7)	4.68 (2.27)	54.9 (38.9)	1.59 (0.54)
	Irrelevant	Child	1678 (971)	91.2 (28.3)	4.77 (2.43)	52.2 (38.4)	1.64 (0.54)
		Adult	1699 (942)	91.5 (27.9)	4.93 (2.49)	57.2 (35.9)	1.64 (0.51)

**Table 1** Means and standard deviations on all dependent measures for all conditions.

	ACCURACY			RESPONSE TIMES			
	ODDS RATIO	95% CI	p	COEFFICIENTS	95% CI	t-value	p
Fixed Effect: Angle	0.09	0.03, 0.22	<.001	544.13	465.94, 622.33	13.64	<.001
Fixed Effect: Avatar-age	0.55	0.16, 1.67	.289	86.91	10.62, 163.20	2.23	0.026
Fixed Effect: Object Relevance	0.37	0.11, 1.04	.063	10.49	-66.01, 86.99	0.27	0.788
Interaction: Angle × Avatar-age	1.69	0.50, 6.23	.393	-93.70	-204.30, 16.89	-1.66	.097
Interaction: Angle × Object Relevance	3.64	1.13, 13.15	.030	-64.58	-175.12, 45.95	-1.15	0.252
Interaction: Avatar-age × Object Relevance	2.15	0.53, 9.43	0.277	-112.14	-220.34, -3.94	-2.03	0.042
Interaction: Angle × Avatar-age × Avatar-gender	0.33	0.06, 1.63	0.161	124.28	-32.10, 280.65	1.56	0.119
(Intercept)	219.84	87.33, 696.311	<.001	1180.64	990.81, 1370.48	12.19	<.001

**Table 2** Results from Mixed Effects Logistic Regression on Accuracy Data and Linear Mixed Effects Analysis on Response Times.

<sup>1</sup> For clarity, including stimuli here controls for the potential variability introduced by individual stimuli.

	NUMBER OF FIXATIONS				AREA OF INTEREST TRANSITIONS				DWELL TIME PERCENTAGE			
	COEFFICIENTS	95% CI	t-value	p	COEFFICIENTS	95% CI	t-value	p	COEFFICIENTS	95% CI	t-value	p
<b>Fixed Effect: Angle</b>	1.17	0.96, 1.37	11.21	<.001	.09	0.04, 0.14	3.53	<.001	-9.77	-14.34, -5.19	-4.19	<.001
<b>Fixed Effect: Avatar-age</b>	0.26	0.06, 0.45	2.53	.012	0.02	-0.03, 0.07	0.77	.444	-5.56	-10.05, -1.07	-2.43	.015
<b>Fixed Effect: Object Relevance</b>	0.03	-0.17, 0.22	0.26	.798	-0.05	-0.10, 0.0001	2.03	.043	-1.79	-6.29, 2.71	-0.78	0.44
<b>Interaction: Angle × Avatar-age</b>	.040	-0.69, 0.11	2.72	.007	-0.03	-0.10, 0.05	0.68	0.495	-0.06	-6.53, 6.41	-0.02	0.99
<b>Interaction: Angle × Object Relevance</b>	.027	-0.56, 0.02	1.83	.067	0.003	-0.07, 0.08	0.08	0.933	-0.41	-6.88, 6.06	-0.12	0.90
<b>Interaction: Avatar-age × Object Relevance</b>	.026	-0.54, 0.02	1.83	.068	.004	-0.07, 0.08	0.12	0.907	5.53	-0.83, 11.90	-1.71	0.09
<b>Interaction: Angle × Avatar-age × Avatar-gender</b>	0.54	0.14, 0.95	2.62	.009	.04	-0.07, 0.14	0.67	.501	1.46	-7.68, 10.61	0.31	0.75
(Intercept)	3.83	3.36, 4.31	15.82	<.001	1.55	1.44, 1.67	27.19	<.001	66.38	58.16, 74.60	15.84	<.001

## OBJECT ANALYSIS

Before proceeding to hypothesis testing, it was important to confirm that there was no effect of object. Linear mixed effects analysis found no significant effect of object on response times, fixations, AOITs or proportionate dwell time (all  $p$ s > .515). Logistic regression found no effect of object on accuracy ( $p$  = .851).

## RESPONSE TIMES

Response time analysis found a significant effect of angle ( $\beta$  = 544.13,  $t(4042)$  = 13.64,  $p$  < .001; Std.  $\beta$  = 0.66) with participants slower at higher compared to lower angles. There was also a significant effect of avatar-age ( $\beta$  = 86.91,  $t(4042)$  = 2.23,  $p$  = 0.026; Std.  $\beta$  = 0.11) with slower response times for child avatars. However, no significant effects of object relevance were found ( $p$  = 0.79). Analysis found a significant interaction between avatar-age and avatar-object relevance ( $\beta$  = ,112.14,  $t(4042)$  = 2.03,  $p$  = 0.042; Std.  $\beta$  = ,0.14); Figure 3 and related post-hoc comparisons report no significant effect of object relevance for adult avatars ( $p$  = .44) but do reveal a significant effect of object relevance for child avatars ( $p$  = .013) with slower response times for irrelevant compared to relevant objects (see Figure 3). No other interactions were significant (all  $p$ s > .097).

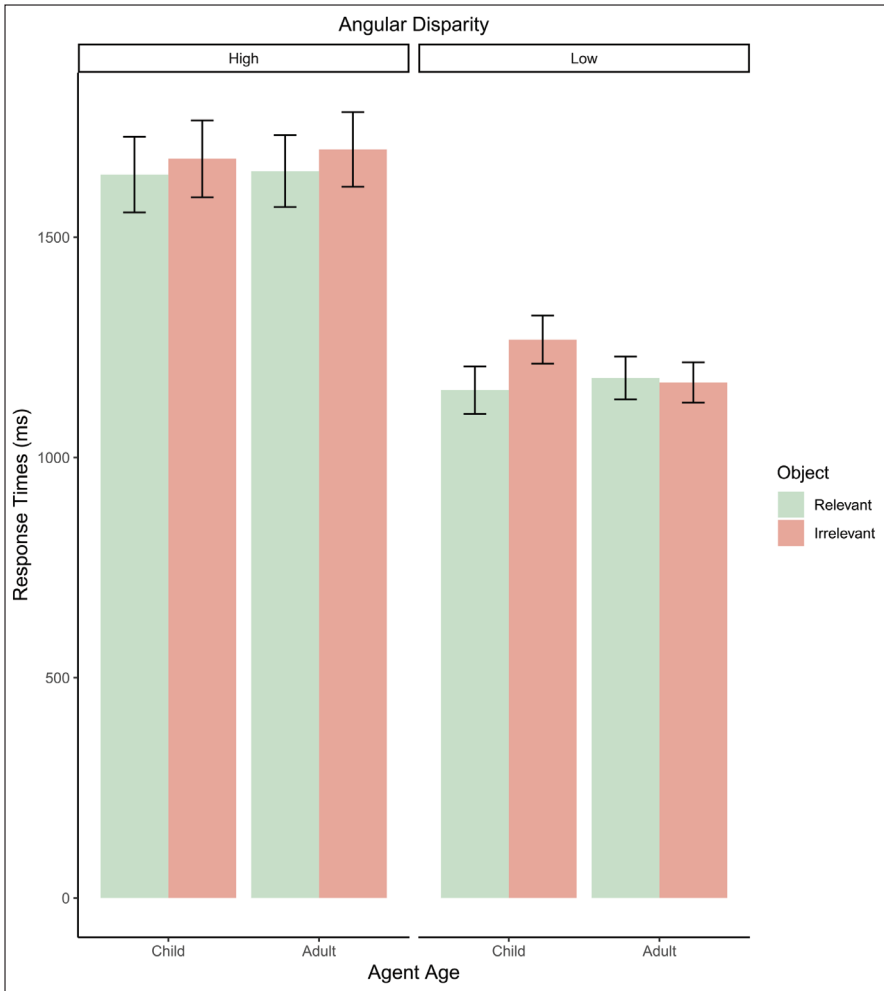
## ACCURACY

Accuracy analysis found a significant fixed effect of angle ( $OR$  = .09,  $p$  < .001) with participants less accurate at 135° than at 45°. No effects of age ( $p$  = 0.29) or object relevance were found ( $p$  = .063). Analysis revealed significant interaction between angle and relevance ( $OR$  = 3.64,  $p$  = 0.03) with higher accuracy for relevant objects at low angular disparity but higher accuracy for irrelevant objects at high angular disparity. Though, post-hoc comparisons reveal no significant differences between object relevance conditions at low ( $p$  = .09) or high angles ( $p$  = .49) (see Figure 4). No other interactions were significant (all  $p$ s > .16).

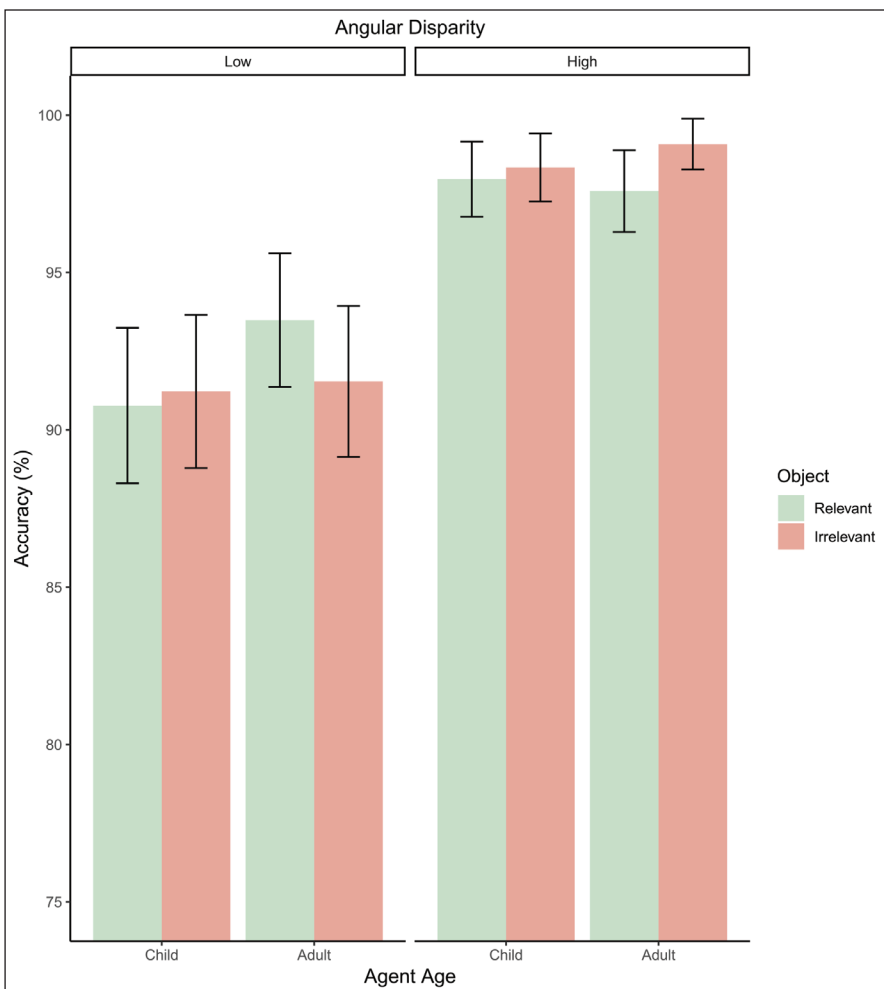
## TOTAL FIXATIONS

Number of fixation analysis found a significant main effect of angle ( $\beta$  = 1.17,  $t(4043)$  = 11.21,  $p$  < .001; Std.  $\beta$  = 0.54) with more fixations at higher angles compared to lower angles. Furthermore, a significant effect of avatar-age was found ( $\beta$  = 0.26,  $t(4043)$  = 2.53,  $p$  = 0.012; Std.  $\beta$  = 0.12) with more fixations made when representing the perspective of a child. However, there was no significant effects of object relevance ( $p$  = 0.80). Analysis found an interaction between angular disparity and avatar-age ( $\beta$  = ,0.40,  $t(4043)$  = 2.72,  $p$  = 0.007; Std.  $\beta$  = ,0.19) with more fixations for child avatars at low angles but no difference at high angles. No other two-way interactions were significant (both  $p$ s > .067). However, analysis revealed a significant

**Table 3** Results from Linear Mixed Effects Analysis on Number of Fixations, Area of Interest Transitions and Dwell Time Proportions.

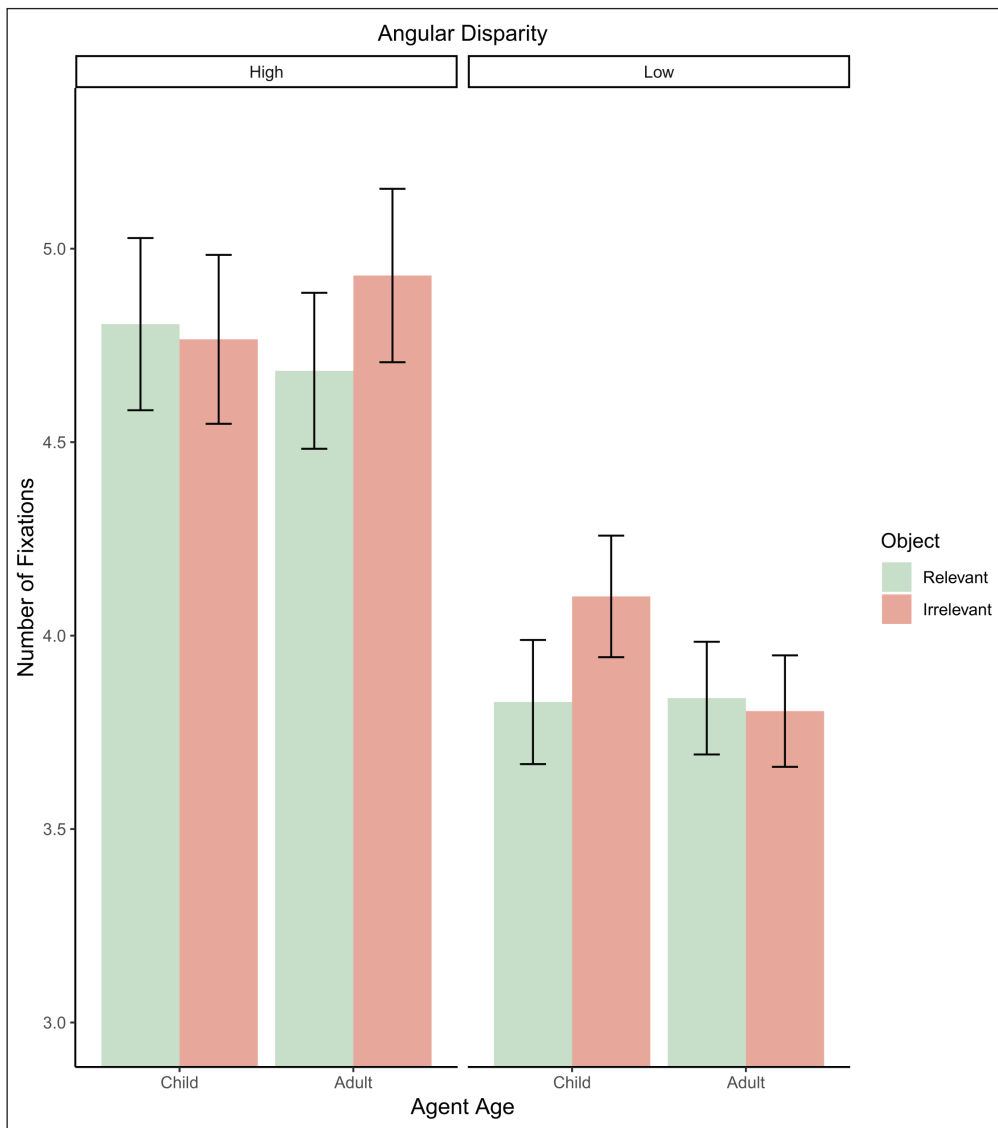


**Figure 3** Bar plot depicting response times of all conditions. Error bars = 95% confidence intervals.



**Figure 4** Bar plot depicting accuracy as a percentage for all conditions. Error bars = 95% confidence intervals.

three-way interaction effect ( $\beta = 0.54$ ,  $t(4043) = 2.62$ ,  $p = 0.009$ ; Std.  $\beta = 0.25$ ). Figure 5 and post-hoc comparisons reveal that at low angles object relevance affects child avatars ( $p = .020$ ) with increased fixations when the object is irrelevant but has no impact on adult avatars ( $p = .798$ ). Conversely, at higher angles object relevance does not impact child avatars ( $p = .710$ ) but does affect adult avatars ( $p = .023$ ) with more fixations when an irrelevant object is presented.



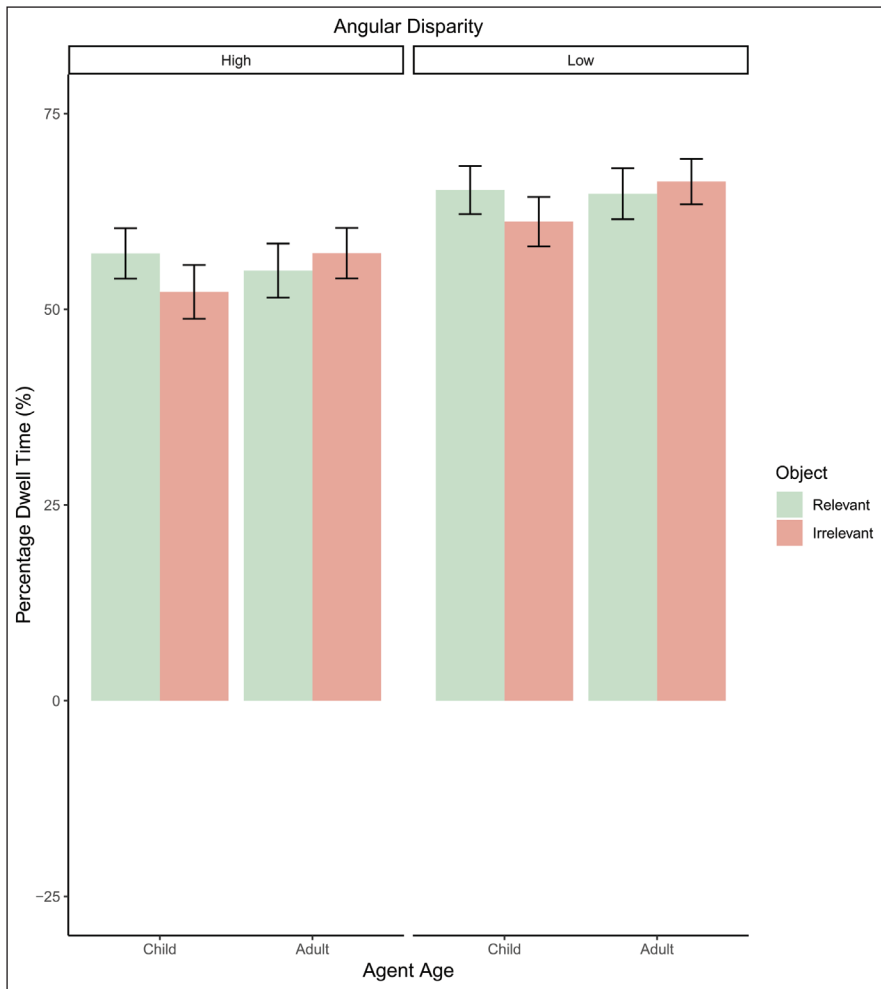
**Figure 5** Bar plot depicting the three-way interaction between avatar-age and object-relevance on number of fixations. Error bars = 95% confidence intervals.

### PROPORTIONATE DWELL TIME

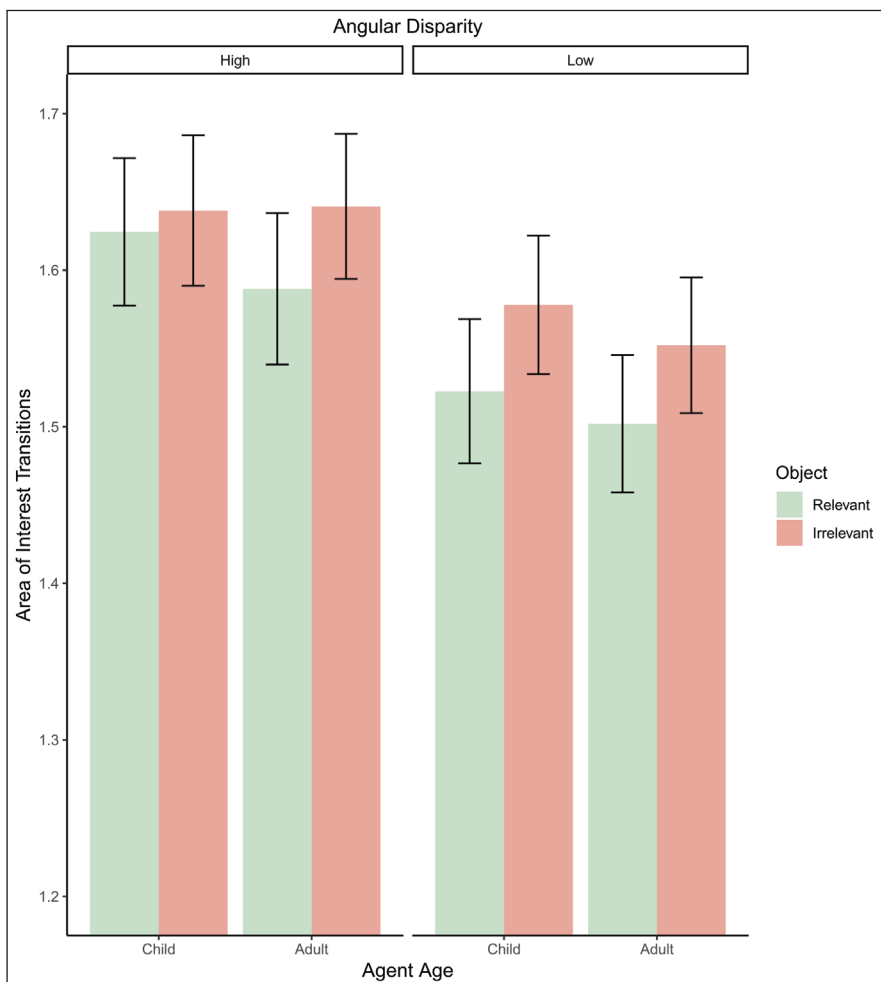
Dwell time percentage analysis (depicted in Figure 6) found a significant effect of angle, with more time spent looking at the avatar during high angular disparity judgements ( $\beta = -9.77$ ,  $t(4042) = -4.19$ ,  $p < .001$ ; Std.  $\beta = -0.26$ ). There was also a significant effect of avatar-age, with more time spent looking at the object with adult-avatars than child-avatars ( $\beta = -5.56$ ,  $t(4042) = -2.43$ ,  $p = 0.015$ ; Std.  $\beta = -0.15$ ). No significant effect of object relevance ( $p = .44$ ) or any significant interaction effects were observed (all  $ps > .09$ ).

### AREA OF INTEREST TRANSITIONS

Area of interest transition analysis (depicted in Figure 7) found a significant effect of angle ( $\beta = 0.09$ ,  $t(4043) = 3.53$ ,  $p < .001$ ; Std.  $\beta = 0.18$ ) with more AOITs at higher angles than at lower angles (see Figure 2). A significant effect of object relevance was observed ( $\beta = -0.05$ ,  $t(4043) = 2.03$ ,  $p = 0.043$ ; Std.  $\beta = 0.10$ ) with more AOITs for irrelevant than relevant objects (see Figure 3). There was no observed effect of avatar-age ( $p = .44$ ). No interactions were significant (all  $ps > .50$ ).



**Figure 6** Bar plot depicting dwell time percentages for all conditions. Error bars = 95% confidence intervals.



**Figure 7** Bar plot depicting area of interest transitions for all conditions. Error bars = 95% confidence intervals.

The current study had three main aims. First, to supplement behavioural and neurophysiological distinctions between low and high angular judgements with attention data. Second, to further examine the effects of avatar characteristic in L2-VPT using eye-tracking to measure the number of fixations, the number of transitions between agents and target objects, and the proportion of time spent fixating on the agents and objects. Third, to explore whether stronger thematic relationships between agent and object impact L2-VPT similarly to how they affect gaze judgements (Gong et al., 2021). Replicating previous work, results suggest that compared to 45°, perspective taking at 135° slowed response times and increased error-rates. The current results further our understanding by demonstrating that accompanying the longer response times and reduced accuracy, there is also a greater number of fixations, an increased number of transitions between agents and objects, and a greater consideration of the agent's position in space. Furthermore, extending the current understanding of agent characteristic effects in L2-VPT, results showed that age affected the proportion of time spent dwelling on the avatar compared to the object (with more time spent on the object when the avatar was an adult). Moreover, age also interacted with object relevance to modulate response times. Also, age interacted with object relevance and angular disparity with object irrelevance increasing fixations for child avatars at low angles but increasing fixations for adults at high angles. Finally, incongruent agent-object relationships resulted in a greater number of transitions between areas of interest.

A corpus of work supports the distinction between processing level-2 visual perspectives at low and high angular disparity (Janczyk, 2013) and our findings of slower response times and lower accuracy at higher angles support a body of existing research relating increasing angular disparity to greater cognitive effort (Michelon & Zacks, 2006; Kessler & Thomson, 2010; Surtees et al., 2013a). Previous neurophysiological findings report additional temporoparietal and executive network activity at higher angular disparity, responsible for inhibiting the egocentric view, selecting the alternate position, and manipulating the visuospatial information (Seymour et al., 2018). Seymour et al. contrast the bottom-up, externally driven processing at low angles when self and other perspectives are aligned to the need for rerouting visual information at higher angles to form an internal representation of the world. Our proportionate dwell time findings reveal that increased attention is paid to the avatar's position that occurs when participants inhibit the egocentric view and assume the alternate position. Furthermore, the increased fixations and area of interest transitions that we observed appear to reflect the greater processing of visual information (e.g., Blythe et al., 2009; Young & Hulleman, 2013; Molitor et al., 2014) that is required to form an internal representation at high angular disparities, supporting similar findings in L1-VPT (Wu & Keysar, 2007; Ferguson et al., 2017). In contrast, at low angles, accurate judgments can be achieved using egocentric visuospatial representations (Briscoe, 2009; Foley et al., 2015) which is reflected in reduced proportionate dwell time to the avatar and less agent-object transitions that are unnecessary for efficient perspective taking at low angles.

Evidence is increasing that agent attributes such as humanness, group-status and age affect the computation of visual perspectives (Todd et al., 2011; Ferguson et al., 2018; Ye et al., 2020). Accordingly, the current research demonstrated a significant effect of age on proportionate dwell time, with similarly aged avatars attended to less than dissimilarly aged avatars. This could reflect similarity facilitation reported in L2-VPT (Ye et al., 2020) where an easier 'step' into the ingroup other's position is accompanied by less attentional engagement. Initially, this appears contrary to the ingroup interference observed in L1-VPT (Savitsky et al., 2011; Simpson & Todd, 2017; Ferguson et al., 2018) and spatial navigation (Todd et al., 2011) where dwell times could have been expected to increase because of egocentric interference. However, the stimuli in Ye et al. (2020) are most similar to the task used in the current study, which highlights important differences between tasks. Egocentric interference impacts L1-VPT and an interpersonal communication task used by Todd et al. (2011) but the current task and Ye et al. (2020) employ non-communicative L2-VPT tasks and similarity appears to facilitate the effort required to respond correctly. Thus, the specific differences between L1- and L2-VPT tasks which result in different avatar-characteristic effects are worth further examination.

Whilst acknowledging that salient characteristics modulate attention in previous research (e.g., Wu et al., 2012; Kawakami et al., 2014; Denkinger & Kinn, 2018; Strickland-Hughes et al., 2020), we advise caution interpreting our main effects of age as evidence of outgroup classification effects. First, the main effects of age are arguably better explained by the significant interactions. For instance, although a significant main effect of age was found for number of fixations inside AOIs, the three-way interaction demonstrates the age relationship was precise and nuanced. Specifically, at low angles irrelevant objects increased the number of fixations for child avatars, whereas at high angles irrelevant objects led to more fixations for adult avatars. Similarly, for the main effect of age on response times, the two-way interaction with object-relevance showed that child avatars looking at relevant objects were processed more quickly than those looking at irrelevant objects, with object relevance having no impact for adult agents. Second, and more fundamentally, participant attention was focused significantly more on the object regardless of avatar-age. Therefore, whether avatar-age was, or even could have been, processed independently of the object and their relationship remains unclear. Thus, considering the small effects of age in previous work (Ford et al. 2023), we hesitate to suggest that age had a significant *independent* impact in the current experiment. Further clarification is required using a task isolating avatar-age and where attention is not explicitly directed towards the object, perhaps by varying stimuli configuration (e.g., non-centrally appearing agents and objects).

As hypothesised, agent-object incongruence increased the number of AOI transitions, extending the known influence of thematic relationships on language comprehension (Estes et al., 2011) to include attentional allocation in sociocognitive processes. One explanation could be that more efficient attention allocation results simply from the efficient processing of semantically related information in a general representation mechanism (Xu et al., 2018). Here, incongruent pairings increase the demand on attentional mechanisms because their saliency is highlighted which may disrupt processing automaticity (Chaigneau et al., 2009; Estes et al., 2011; Kalénine et al., 2012). A second explanation of agent-object effects could be intentionality processing. Despite ‘full-blown’ ToM not being required in terms of explicit reasoning or action description (Butterfill & Apperly, 2013), the combination of thematic relationships and gaze-cues may represent minimal requirements for attributions of intentionality (Castiello, 2003; Becchio et al., 2008; Chaigneau et al., 2009), or at-least increase the likelihood of a mentalising-based route to perspective judgments. Thus, considering the close relationship between ToM and VPT (Dumontheil et al., 2010; Hamilton et al., 2009), the representation of an intentional mind may be one specific mechanism that affects attention allocation in L2-VPT. If so, the current findings represent the first step towards understanding how implicit inferences of intentionality arising from the interaction between individual attributes and social-contextual cues affect attentional processes involved in perspective representation. This also has consequences for predicting behaviour and how unlikely relationships between agents and objects are processed by inferential and interpretative processing networks (Brass et al., 2007; Richardson & Saxe, 2020).

Object relevance was also found to interact with avatar-age. Whilst object relevance did not affect response times for adult avatars, for child avatars, age-relevant objects reduced response times and age-irrelevant objects increased response times. This suggests that while less interference is experienced from incongruent cues when paired with an in-group member, additional social cues facilitate or disrupt perspective representation of outgroup others. Importantly, real-world experience is important for developing thematic relationships (Estes et al., 2011) as well as our bias to perceive outgroups as more homogenous (Boldry et al., 2007; Konovalova & Le Mens, 2020). As such, the current findings suggest that experience may influence our conceptual flexibility. Specifically, the experience we have with in-group others increases conceptual flexibility of their interactions with objects in the environment. Conversely, the lesser experience we have with out-groups constrains our conceptions of what objects they are likely to interact with, and the more likely agent-object pairings are to influence concurrent processing. Given the lack of previous research examining outgroup schemas and agent-object interaction, our findings offer initial support that congruent pairings, which confirm rigid ideas about outgroup-object relationships, may reduce processing demand relative to incongruent pairings. Conversely, ingroup heterogeneity may extend to object interaction and agent-object incongruencies do not significantly impact processing difficulty. Thus, although most VPT research isolates specific task cues, the current findings highlight the importance of integrating multiple cues, especially if we are to develop predictive theories about VPT in everyday social interactions.

The interaction between avatar-age and object relevance visible in eye-tracking data was modulated by angular disparity. At low angles, where perspective taking can rely on an egocentric frame of reference (Wang et al., 2016), object incongruence increased fixations for child agent trials. Here, dissimilar others who share our perspective present a conflict of group categorisation and coupled with agent-object incongruency, which may violate outgroup homogeneity assumptions, these inconsistencies increase attentional engagement (Henderson et al., 1999; Vö & Henderson, 2009). Conversely, similarly aged others offer little interference from categorisation processes or from agent-object incongruency perhaps due to in-group heterogeneity biases. Conversely, at high angular disparity, object incongruence increased fixations for adult agent trials. At high angles, perspective taking is effortful and involves a deliberate mental self-rotation (Kessler & Thomson, 2010) with increased focus on the agent because their location indicates when self-rotation is complete (Ward et al., 2020). In another L2-VPT task, Ye et al. (2020) concluded that dissimilar avatars produce less interference as a function of outgroup dehumanisation. Similarly, we propose that with the increased focus on mental self-rotation and spatial processing, social agency is more easily disregarded for outgroup members and the agent-object incongruency less salient. Yet, with ingroup agents, participants may suffer greater egocentric interference because of the combination of ingroup status, mental self-rotation, and perspective conflict. This is in line with other findings from L1-and-L2-VPT where ingroup status interferes with processing (Savitsky et al., 2011; Simpson & Todd, 2017; Todd et al., 2011; Ferguson et al., 2018) as well as reports where conflicting perspectives increase fixation counts in L1-VPT (Wu & Keysar, 2007).

## LIMITATIONS

On all trials, objects and agents were located in predictable positions. Real-world social cues are often dynamic with regular gaze shifts, varying distances and changing agent-object relationships. To better resemble real-world social processing, future eye-tracking research could vary room layouts or present stimuli as to encourage naturalistic viewing behaviour (Kim et al. 2020). Manipulating age also has potential limitations. First, participant's experience with children may modulate responses to agent-object pairings (Estes et al., 2011) with the undergraduate sample recruited potentially having less experience with young children than the general population. However, agent age was not manipulated for specific age-related hypotheses. Instead, age was a pragmatic choice to represent a general outgroup of everyday relevance, discernible from salient characteristics (Ferguson et al., 2018). Relatedly, we are unable to ascertain how participants categorised agents or conceptualised agent-object relationships. For instance, Ferguson et al. (2018) suggest that adults assume a reduced mental capacity for children which reduces perspective saliency. Thus, perhaps reduced readiness for mentalising is a crucial factor and future research is required to understand precisely what moderates differential attentional processing towards groups and agent-object relationships.

## FUTURE DIRECTIONS

Our angular disparity findings raise an interesting question regarding attention allocation required for different types of L2-VPT. Previous work distinguishes between visual and spatial forms of L2-VPT and finds that judging whether a '6' or '9' appears on the left or right is more effortful than whether the agent 'sees' a '6' or '9' (Surtees et al., 2013a; Surtees et al., 2013b). Based on our findings, increased effort could be a function of increased information gathering when it is necessary to assess object-agent relationships on *spatial* perspective taking. Conversely, for *visual* perspective tasks where targets are easily represented from any view, participants need only consider the approximate agent position. Future eye-tracking research could elucidate attentional processes underlying this visual-spatial distinction. Furthermore, our hypothesis that implicit mentalising about intentionality affects L2-VPT should be examined. For example, replicating the current method but grouping participants by autistic traits would be pertinent, as has been conducted in gaze-cueing research (Morgan et al. 2021). Also, understanding how intentionality judgments impact perspective-taking across the lifespan are warranted considering the development and decline of perspective taking and theory-of-mind in younger and older age groups (Klindt et al., 2017; De Lillo & Ferguson, 2023).

## CONCLUSION

Evidence that agent-object relationships significantly impact cognitive effort and attention allocation in L2-VPT demonstrate the role of perspective representation in understanding intentionality and predicting behaviour. Our results are the first to demonstrate the effects angular disparity on attention allocation in L2-VPT, extending the current cognitive and neurophysiological evidence (Kessler & Thomson, 2010; Seymour et al., 2018). Our findings are also the first evidence of social cues affecting attention orienting in L2-VPT and build on previous work in L1-VPT (Wu & Keysar, 2007; Ferguson et al., 2017; Wei et al., 2020) and gaze-cueing (Kawakami et al., 2018). Specifically, we extend the findings reported for isolated agent-characteristics (Savitsky et al., 2011; Todd et al., 2011; Simpson & Todd, 2017; Ferguson et al., 2018; Ye et al., 2020) to agent-object relationships. More generally, despite the fundamental spatial processing at the heart of L2-VPT tasks (Kessler & Thomson, 2010; Ward et al., 2020), our findings suggest that human attributes of agents are being considered along with other socio-contextual cues. Understanding what specific contexts and which specific cues affect visuospatial processing will support the development of integrative accounts of perspective taking. These accounts can incorporate underlying processes, person-attributes, task demands and relevant sociocontextual information, allowing researchers to make theory-led predictions about when representing other's experiences is facilitated or disrupted.

## DATA ACCESSIBILITY STATEMENT

Raw data available at: <https://doi.org/10.17605/OSF.IO/8GDM6>.

## ETHICS AND CONSENT

This research project was carried out in accordance with the Declaration of Helsinki. Ethical approval was obtained from Edge Hill Psychology's Science Research Ethics Committee (ETH2122-0080).

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR AFFILIATIONS

**Ben Ford**  [orcid.org/0000-0001-7502-7505](https://orcid.org/0000-0001-7502-7505)

University of Gloucestershire, UK; Edge Hill University, UK

**Prof. Rebecca Monk**  [orcid.org/0000-0002-3554-9007](https://orcid.org/0000-0002-3554-9007)

Edge Hill University, UK

**Dr Damien Litchfield**  [orcid.org/0000-0001-7760-293X](https://orcid.org/0000-0001-7760-293X)

Edge Hill University, UK

**Dr Adam Qureshi**  [orcid.org/0000-0002-7698-2691](https://orcid.org/0000-0002-7698-2691)

Edge Hill University, UK

## REFERENCES

- Barr, D. J.** (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, 4, 328. <https://doi.org/10.3389/fpsyg.2013.00328>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J.** (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- 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>
- Becchio, C., Bertone, C., & Castiello, U.** (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, 12(7), 254–258. <https://doi.org/10.1016/j.tics.2008.04.005>
- Blythe, H. I., Liversedge, S. P., Joseph, H. S., White, S. J., & Rayner, K.** (2009). Visual information capture during fixations in reading for children and adults. *Vision Research*, 49(12), 1583–1591. <https://doi.org/10.1016/j.visres.2009.03.015>

- Boldry, J. G., Gaertner, L., & Quinn, J.** (2007). Measuring the measures: A meta-analytic investigation of the measures of outgroup homogeneity. *Group Processes & Intergroup Relations*, 10(2), 157–178. <https://doi.org/10.1177/1368430207075153>
- Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G.** (2007). Investigating action understanding: inferential processes versus action simulation. *Current Biology*, 17(24), 2117–2121. <https://doi.org/10.1016/j.cub.2007.11.057>
- Briscoe, R.** (2009). Egocentric spatial representation in action and perception. *Philosophy and Phenomenological Research*, 79(2), 423–460. <https://doi.org/10.1111/j.1933-1592.2009.00284.x>
- Brüne, M.** (2005). “Theory of mind” in schizophrenia: a review of the literature. *Schizophrenia Bulletin*, 31(1), 21–42. <https://doi.org/10.1093/schbul/sbi002>
- Butterfill, S. A., & Apperly, I. A.** (2013). How to construct a minimal theory of mind. *Mind & Language*, 28(5), 606–637. <https://doi.org/10.1111/mila.12036>
- Cane, J. E., Ferguson, H. J., & Apperly, I. A.** (2017). Using perspective to resolve reference: The impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(4), 591–610. <https://doi.org/10.1037/xlm0000345>
- Castiello, U.** (2003). Understanding other people’s actions: intention and attention. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 416–430. <https://doi.org/10.1037/0096-1523.29.2.416>
- Chaigneau, S. E., Barsalou, L. W., & Zamani, M.** (2009). Situational information contributes to object categorization and inference. *Acta Psychologica*, 130(1), 81–94. <https://doi.org/10.1016/j.actpsy.2008.10.004>
- Cole, G. G., Smith, D. T., & Atkinson, M. A.** (2015). Mental state attribution and the gaze cueing effect. *Attention, Perception, & Psychophysics*, 77(4), 1105–1115. <https://doi.org/10.3758/s13414-014-0780-6>
- De Lillo, M., & Ferguson, H. J.** (2023). Perspective-taking and social inferences in adolescents, young adults, and older adults. *Journal of Experimental Psychology: General*. <https://doi.org/10.31234/osf.io/8z2tf>
- Denkinger, B., & Kinn, M.** (2018). Own-age bias and positivity effects in facial recognition. *Experimental Aging Research*, 44(5), 411–426. <https://doi.org/10.1080/0361073X.2018.1521493>
- Dumontheil, I., Apperly, I. A., & Blakemore, S. J.** (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13(2), 331–338. <https://doi.org/10.1111/j.1467-7687.2009.00888.x>
- Estes, Z., Golonka, S., & Jones, L. L.** (2011). Thematic thinking: The apprehension and consequences of thematic relations. In *Psychology of Learning and Motivation* (Vol. 54, pp. 249–294). Academic Press. <https://doi.org/10.1016/B978-0-12-385527-5.00008-5>
- Ferguson, H. J., Apperly, I., & Cane, J. E.** (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *The Quarterly Journal of Experimental Psychology*, 70(8), 1646–1660. <https://doi.org/10.1080/17470218.2016.1199716>
- Ferguson, H. J., Brunsdon, V. E. A., & Bradford, E. E. F.** (2018). Age of avatar modulates the altercentric bias in a visual perspective-taking task: ERP and behavioral evidence. *Cognitive, Affective, & Behavioral Neuroscience*, 18(6), 1298–1319. <https://doi.org/10.3758/s13415-018-0641-1>
- Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R.** (1981). Young children’s knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, 17, 99–103. <https://doi.org/10.1037/0012-1649.17.1.99>
- Foley, R. T., Whitwell, R. L., & Goodale, M. A.** (2015). The two-visual-systems hypothesis and the perspectival features of visual experience. *Consciousness and Cognition*, 35, 225–233. <https://doi.org/10.1016/j.concog.2015.03.005>
- Ford, B., Monk, R., Litchfield, D. & Qureshi, A.** (2023). Manipulating avatar age and gender in level-2 visual perspective taking. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-023-02249-7>
- Frith, C. D., & Frith, U.** (2006). How we predict what other people are going to do. *Brain Research*, 1079, 36–46. <https://doi.org/10.1016/j.brainres.2005.12.126>
- Gong, Y., Wang, Y., Chen, Q., Zhao, J., Zhao, N., Zou, M., .... & Wang, Y.** (2021). The influence of cooperative action intention on object affordance: evidence from the perspective-taking ability of individuals. *Psychological Research*, 86, 1–12. <https://doi.org/10.1007/s00426-021-01523-4>
- Hamilton, A. F. de C., Brindley, R., & Frith, U.** (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition*, 113(1), 37–44. <https://doi.org/10.1016/j.cognition.2009.07.007>
- Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A.** (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), 210. <https://doi.org/10.1037//0096-1523.25.1.210>
- Hughes, C., & Leekam, S.** (2004). What are the links between theory of mind and social relations? Review, reflections and new directions for studies of typical and atypical development. *Social Development*, 13(4), 590–619. <https://doi.org/10.1111/j.1467-9507.2004.00285.x>
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G.** (2005). Grasping the intentions of others with one’s own mirror neuron system. *PLoS Biology*, 3(3), e79. <https://doi.org/10.1371/journal.pbio.0030079>

- Janczyk, M.** (2013). Level 2 perspective taking entails two processes: evidence from PRP experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(6), 1878. <https://doi.org/10.1037/a0033336>
- Kalénine, S., Mirman, D., Middleton, E. L., & Buxbaum, L. J.** (2012). Temporal dynamics of activation of thematic and functional knowledge during conceptual processing of manipulable artifacts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1274. <https://doi.org/10.1037/a0027626>
- Kawakami, K., Friesen, J., & Vingilis-Jaremko, L.** (2018). Visual attention to members of own and other groups: Preferences, determinants, and consequences. *Social and Personality Psychology Compass*, 12(4), e12380. <https://doi.org/10.1111/spc3.12380>
- Kawakami, K., Williams, A., Sidhu, D., Choma, B. L., Rodriguez-Bailón, R., Cañadas, E., .... & Hugenberg, K.** (2014). An eye for the I: Preferential attention to the eyes of ingroup members. *Journal of Personality and Social Psychology*, 107(1), 1–20. <https://doi.org/10.1037/a0036838>
- Kessler, K., & Rutherford, H.** (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology*, 1, 213. <https://doi.org/10.3389/fpsyg.2010.00213>
- 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>
- Kessler, K., & Wang, H.** (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: differences predicted by sex and social skills score. *Spatial Cognition & Computation*, 12(2–3), 133–158. <https://doi.org/10.1080/13875868.2011.634533>
- Kim, H. C., Jin, S., Jo, S., & Lee, J. H.** (2020). A naturalistic viewing paradigm using 360 panoramic video clips and real-time field-of-view changes with eye-gaze tracking. *NeuroImage*, 216, 116617. <https://doi.org/10.1016/j.neuroimage.2020.116617>
- Klindt, D., Devaine, M., & Daunizeau, J.** (2017). Does the way we read others' mind change over the lifespan? Insights from a massive web poll of cognitive skills from childhood to late adulthood. *Cortex*, 86, 205–215. <https://doi.org/10.1016/j.cortex.2016.09.009>
- Konvalova, E., & Le Mens, G.** (2020). An information sampling explanation for the in-group heterogeneity effect. *Psychological Review*, 127(1), 47–93. <https://doi.org/10.1037/rev0000160>
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M.** (2019). emmeans. R package version 1.8.8.
- Leslie, A. M.** (1987). Pretense and representation: The origins of “theory of mind.” *Psychological Review*, 94(4), 412–426. <https://doi.org/10.1037/0033-295X.94.4.412>
- Michelon, P., & Zacks, J. M.** (2006). Two kinds of visual perspective taking. *Perception & Psychophysics*, 68(2), 327–337. <https://doi.org/10.3758/BF03193680>
- Molitor, R. J., Ko, P. C., Hussey, E. P., & Ally, B. A.** (2014). Memory-related eye movements challenge behavioral measures of pattern completion and pattern separation. *Hippocampus*, 24(6), 666–672. <https://doi.org/10.1002/hipo.22256>
- Monk, R. L., Colbert, L., Darker, G., Cowling, J., Jones, B., & Qureshi, A. W.** (2020). Emotion and liking: How director emotional expression and knowledge of (dis)liking may impact adults' ability to follow the instructions of an ignorant speaker. *Psychological Research*, 85, 2755–2768. <https://doi.org/10.1007/s00426-020-01441-x>
- Morgan, E. J., Freeth, M., & Smith, D. T.** (2018). Mental state attributions mediate the gaze cueing effect. *Vision*, 2(1), 11. <https://doi.org/10.3390/vision2010011>
- Morgan, E. J., Smith, D. T., & Freeth, M.** (2021). Gaze cueing, mental States, and the effect of autistic traits. *Attention, Perception, & Psychophysics*, 1–9. <https://doi.org/10.3758/s13414-021-02368-0>
- Nummenmaa, L., & Calder, A. J.** (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, 13(3), 135–143. <https://doi.org/10.1016/j.tics.2008.12.006>
- R Core Team** (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL <http://www.R-project.org/>
- Richardson, H., & Saxe, R.** (2020). Development of predictive responses in theory of mind brain regions. *Developmental Science*, 23(1), e12863. <https://doi.org/10.1111/desc.12863>
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E.** (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1255–1266. <https://doi.org/10.1037/a0018729>
- Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A.** (2011). The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*, 47(1), 269–273. <https://doi.org/10.1016/j.jesp.2010.09.005>
- Seymour, R. A., Wang, H., Rippon, G., & Kessler, K.** (2018). Oscillatory networks of high-level mental alignment: a perspective-taking MEG study. *NeuroImage*, 177, 98–107. <https://doi.org/10.1016/j.neuroimage.2018.05.016>

- Simpson, A. J., & Todd, A. R.** (2017). Intergroup visual perspective-taking: Shared group membership impairs self-perspective inhibition but may facilitate perspective calculation. *Cognition*, 166, 371–381. <https://doi.org/10.1016/j.cognition.2017.06.003>
- SR Research Experiment Builder 1.10.165 [Computer software].** (2011). Mississauga, Ontario, Canada: SR Research Ltd.
- Strickland-Hughes, C. M., Dillon, K. E., West, R. L., & Ebner, N. C.** (2020). Own-age bias in face-name associations: Evidence from memory and visual attention in younger and older adults. *Cognition*, 200, 104253. <https://doi.org/10.1016/j.cognition.2020.104253>
- Surtees, A., Apperly, I., & Samson, D.** (2013a). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, 129(2), 426–438. <https://doi.org/10.1016/j.cognition.2013.06.008>
- Surtees, A., Apperly, I., & Samson, D.** (2013b). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in human neuroscience*, 7, 698. <https://doi.org/10.3389/fnhum.2013.00698>
- Teufel, C., Alexis, D. M., Clayton, N. S. and Davis, G.** (2010a). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception, & Psychophysics*, 72(3), pp. 695–705. <https://doi.org/10.3758/APP.72.3.695>
- Teufel, C., Fletcher, P. C., & Davis, G.** (2010b). Seeing other minds: attributed mental states influence perception. *Trends in Cognitive Sciences*, 14(8), pp. 376–382. <https://doi.org/10.1016/j.tics.2010.05.005>
- Todd, A. R., Hanks, K., Galinsky, A. D., & Mussweiler, T.** (2011). When focusing on differences leads to similar perspectives. *Psychological Science*, 22(1), 134–141. <https://doi.org/10.1177/0956797610392929>
- Tomasello, M.** (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences*, 115(34), 8491–8498. <https://doi.org/10.1073/pnas.1804761115>
- Unity Technologies.** (2020). Unity (Version 2021.1.0) [computer software]. <https://unity.com/releases/editor/archive>
- Võ, M. L. H., & Henderson, J. M.** (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, 9(3), 24–24. <https://doi.org/10.1167/9.3.24>
- Wang, H., Callaghan, E., Gooding-Williams, G., McAllister, C., & Kessler, K.** (2016). Rhythm makes the world go round: An MEG-TMS study on the role of right TPJ theta oscillations in embodied perspective taking. *Cortex*, 75, 68–81. <https://doi.org/10.1016/j.cortex.2015.11.011>
- Ward, E., Ganis, G., McDonough, K. L., & Bach, P.** (2020). Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze. *Cognition*, 199, 104241. <https://doi.org/10.1016/j.cognition.2020.104241>
- Wei, Y., Wan, Y., & Tanenhaus, M.** (2020). Effects of Coordination on Perspective-taking: Evidence from Eye-tracking. In *CogSci*.
- Weiss, M. J., & Harris, S. L.** (2001). Teaching social skills to people with autism. *Behavior Modification*, 25(5), 785–802. <https://doi.org/10.1177/0145445501255007>
- Westfall, J., Kenny, D. A., & Judd, C. M.** (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5), 2020–2045. <https://doi.org/10.1037/xge0000014>
- Wu, E. X. W., Laeng, B., & Magnussen, S.** (2012). Through the eyes of the own-race bias: Eye-tracking and pupillometry during face recognition. *Social Neuroscience*, 7(2), 202–216. <https://doi.org/10.1080/17470919.2011.596946>
- Wu, S., & Keysar, B.** (2007). The effect of culture on perspective taking. *Psychological Science*, 18(7), 600–606. <https://doi.org/10.1111/j.1467-9280.2007.01946.x>
- Xu, Y., Wang, X., Wang, X., Men, W., Gao, J. H., & Bi, Y.** (2018). Doctor, teacher, and stethoscope: neural representation of different types of semantic relations. *Journal of Neuroscience*, 38(13), 3303–3317. <https://doi.org/10.1523/JNEUROSCI.2562-17.2018>
- Ye, T., Furumi, F., Catarino da Silva, D., & Hamilton, A.** (2020). Taking the perspectives of many people: Humanization matters. *Psychonomic Bulletin & Review*, 28, 888–897. <https://doi.org/10.3758/s13423-020-01850-4>
- Young, A. H., & Hulleman, J.** (2013). Eye movements reveal how task difficulty moulds visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 168–190. <https://doi.org/10.1037/a0028679>

#### TO CITE THIS ARTICLE:

Ford, B., Monk, R., Litchfield, D., & Qureshi, A. (2024). Agent-Object Relationships in Level-2 Visual Perspective Taking: An Eye-Tracking Study. *Journal of Cognition*, 7(1): 72, pp. 1–17. <https://doi.org/10.5334/joc.398>

**Submitted:** 07 November 2023

**Accepted:** 21 August 2024

**Published:** 10 October 2024

#### COPYRIGHT:

© 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Journal of Cognition* is a peer-reviewed open access journal published by Ubiquity Press.