Visual perspective-taking and image-like representations: We don't see it.

Steven Samuel

Klara Hagspiel

Madeline J. Eacott

Geoff G. Cole

(ACCEPTED 15/1/2021 at COGNITION)

Department of Psychology, University of Essex, U.K.

Word count: 10,465 (excluding abstract, acknowledgements, captions, references and tables, including footnotes)

Key words: perspective-taking, vision, theory of mind, perceptual simulation, naïve optics

Address for correspondence: Steven Samuel. University of Essex, Wivenhoe Park, Department of Psychology, CO4 3SQ. Email: ssamuea@essex.ac.uk

Abstract

The ability to represent another agent's visual perspective has recently been attributed to a process called "perceptual simulation", whereby we generate an image-like or "quasi-perceptual" representation of another agent's vision. In an extensive series of experiments we tested this notion. Adult observers were presented with pictures of an agent looking at two horizontal lines, one of which was closer to the agent and hence appeared longer from his/her visual perspective. In each case approximately as many participants judged the closer line to appear *shorter* as longer (to the agent), i.e., failures to take the agent's perspective. This occurred when clear depth cues were added to emphasise the agent's location relative to the stimuli, when the agent was moved closer to the lines, when the lines were oriented vertically, when judgments could be made while viewing the image, and when participants imagined themselves in the agent's place. It also persisted when we asked participants to imagine what a photo taken from the same location as the agent would show, ruling out a misinterpretation of the instructions. Overall, our data suggest that adults attempt to solve visual perspective-taking problems by drawing upon naïve and often erroneous ideas about how vision works rather than through attempting to simulate perception.

1. Introduction

Visual perspective-taking (VPT) is crucial to our ability to understand, communicate, and deal with other agents (Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Clark & Brennan, 1991; Linde & Labov, 1975), yet there is currently no formal theory of VPT (Cole & Millett, 2019; Cole, Millett, Samuel, & Eacott, 2020). A crucial challenge for any such theory is the representation problem; what exactly can we represent about another person's perspective when we assume a viewpoint? In this paper we report a series of studies that examine the claim that we might represent the contents of another person's viewpoint as something akin to an image, such that we effectively 'see what they see'.

The problem of representation concerns both the nature and the amount of content that a perspective-taker reconstructs from another agent or location, and scholars are increasingly beginning to address the issue explicitly. For example, Samson, Apperly, Braithwaite, Andrews, and Bodley Scott (2010) refer to VPT as concerning another's "visual experience"[1], Elekes, Varga, and Király (2016) reference the ability to "mentally map how a scene looks", and Moll and Kadipasaoglu (2013) refer to 'snapshot perspectives in a literal, i.e., optical sense of the term." Such notions support a theory of VPT by which we generate a representation of another agent's perception. Moreover, some scholars place VPT explicitly within the remit of a theory of mind (Elekes et al., 2016; Ferguson, Apperly, & Cane, 2017), which concerns the understanding of others' mental states more broadly, such as knowledge, desires, and beliefs (Premack & Woodruff, 1978). VPT has also been put forward as a candidate precursor to these more 'social' inferences (Kessler & Rutherford, 2010).

More detailed hypotheses concerning representation are now beginning to be generated. In a recent study, Ward, Ganis, and Bach (2019) asked adult participants to judge whether an alphanumeric character, sometimes rotated so that it was not viewed upright by the participant, was presented in either canonical or mirror-reversed form. Ward et al. replicated the usual positive

[1] Though Samson and colleagues do not argue that this visual experience will include detail necessary to solve Level 2 VPT problems.

correlation between response times and closeness of the character to its normal orientation, an effect attributed to the time it takes to mentally rotate the target (Shepard & Metzler, 1971). Crucially, they also found an effect of inserting a person into the scene; participants were faster to make judgments about left-rotated characters when the agent viewed them from the left, and faster to make judgments about right-rotated characters when the agent viewed them from the right. This influence of the other agent occurred despite that agent's perspective being task-irrelevant, i.e., participants were not instructed to take that agent's perspective during the task. The effect was also present in a control experiment in which participants were asked explicitly to take the agent's perspective. However, substituting the agent for an articulated lamp reduced the effect. Presented with an apparently vicariously-experienced mental rotation effect centred on another agent, the authors concluded that "the content of another's perspective is therefore spontaneously derived, takes a quasi-perceptual form, and can stand in for one's own sensory input during perceptual decision making." Ward et al. (2019) also argued that this representation is integrated with other processes like working memory in the same way as actual (direct) perceptual input. They termed this ability *perceptual simulation*.

Perceptual simulation is convenient as an account for VPT because, in theory, there is no type of VPT problem that a reconstructed copy of an agent's vision could not solve. Since it proposes that this reconstruction interacts with other cognitive systems in the same way as our own visual input, what we already know about the latter we can 'port' to the former. However, crucial to the perceptual simulation account is the idea that this representation is 'quasi-perceptual'. The meaning of this term in this context is not formally defined, but there can be little doubt that it refers to a depictive, image-like representation, one that relies on a reasonably faithful mapping to what is actually perceived. This is clear when Ward and colleagues state that this ability generates representations of others' vision and "'inserts' them onto one's own perceptual processes, as if they were one's own perceptual input" (Ward, Ganis, McDonough, & Bach, 2020); and when they describe the process as "'painting' a *mental image* of the content of another person's viewpoint onto one's own perceptual system" (italics added), Ward et al. (2019). It is also clear when they state that perceptual simulation allows people to "recognize items that would be more difficult to recognize from their own perspective," (Ward et al., 2020). As Cole and colleagues (Cole & Millett, 2019; Cole et al. 2020) have pointed out, 'quasi-perceptual' representations of a scene owes much to Kosslyn and colleagues' notion of mental imagery as being 'quasi-pictorial'. By this, Kosslyn et al. meant that there is a one-to-one mapping of a representation to the real percept, thus explaining the classic mental imagery scanning results whereby the time it takes to 'move' from one point on an image to another increases with distance in a linear manner (Kosslyn, Pinker, Smith, & Shwartz, 1979). This effect occurs, Kosslyn et al. argued, because the images are represented in a medium that goes beyond a symbolic code. In essence, the perceptual simulation account takes a step beyond traditional theories of perspective-taking as inference or knowledge attribution (e.g., Nuku & Bekkering, 2008) and suggests something new. The theory effectively argues that observers can represent a scene in the manner that Kosslyn suggested with respect to mental imagery. That is, there is a faithful one-to-one mapping of the agent's viewpoint. The principal test of perceptual simulation, then, is to check whether VPT implies the generation of something akin to a mental image representation.

For all these reasons, perceptual simulation has now become an important frontier in the understanding of VPT. Indeed, by being a spontaneous ability (one that works rapidly and outside of one's awareness), it could also provide an account for evidence of spontaneous VPT with other stimuli and tasks (e.g., Samson et al., 2010). However, a difficulty arises when we attempt to assess the account using past data. Many VPT experiments employ tasks that can be solved using heuristics and strategies that do not require perceptual or quasi-perceptual representations at all, such as drawing a mental line from an agent's eyes to an object and concluding the object is seen if the line is unbroken (Michelon & Zacks, 2006), shifting attention towards what an agent or object is facing irrespective of whether either can see (Cole, Atkinson, Le, & Smith, 2016; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014), or applying heuristics or 'shortcuts' such as utilising one's knowledge of the properties of stimuli. For example, understanding the digit 6/9 is reversible means that this knowledge and not an agent's "perspective" can be enough to solve a VPT problem with this stimulus. Similarly, reversing spatial mappings for an agent facing us (Yu & Zacks, 2017) suffices to solve problems of relative locations of visible objects. An additional issue is that many VPT studies rely on indirect measures of perspective-taking such as those used in attentional manipulation experiments in which RT is the dependent measure. While useful as indices of processing, RTs reveal little about the whether an image was employed to solve the task, nor how accurate that image might be.

A more direct way to test perceptual simulation is therefore to present participants with a VPT task in which stimuli vary by perspective *exclusively perceptually*, thereby maximising the requirement for imagery and minimising the potential for non-pictorial alternatives. Additionally, instead of measuring RTs, the accuracy of such representations can be measured directly. Figure 1 (left panel) illustrates such a scenario. The participant sees an agent located to the front and left of the participant looking at two lines on the wall. The lines are of equal physical length, but any representation that faithfully conforms to the agent's perception should code the line on the left as longer than the line on the right, as indicated by the right panel image which is a photograph taken from the same location as the agent's head. Crucially, the lines on the wall do not change in category (i.e., they are still labelled 'lines') or in their left/right relationships relative to the participant and the agent, and therefore alternative strategies to image-like-generation are minimised. After seeing this stimulus, participants can be asked to judge the relative lengths of the lines as they appeared visually from the agent's perspective. These judgments can be made on a continuous scale which indexes the accuracy of any representation generated. With its reliance on perception and image-like generation, rather than knowledge or symbolic reasoning, a perceptual simulation account of VPT predicts a one-to-one mapping of the agent's perspective to the representation (i.e., Kosslyn-esque). In this quasi-perceptual account the closer line to the agent should be longer than the further line. We ran an extensive series of experiments to test this hypothesis.

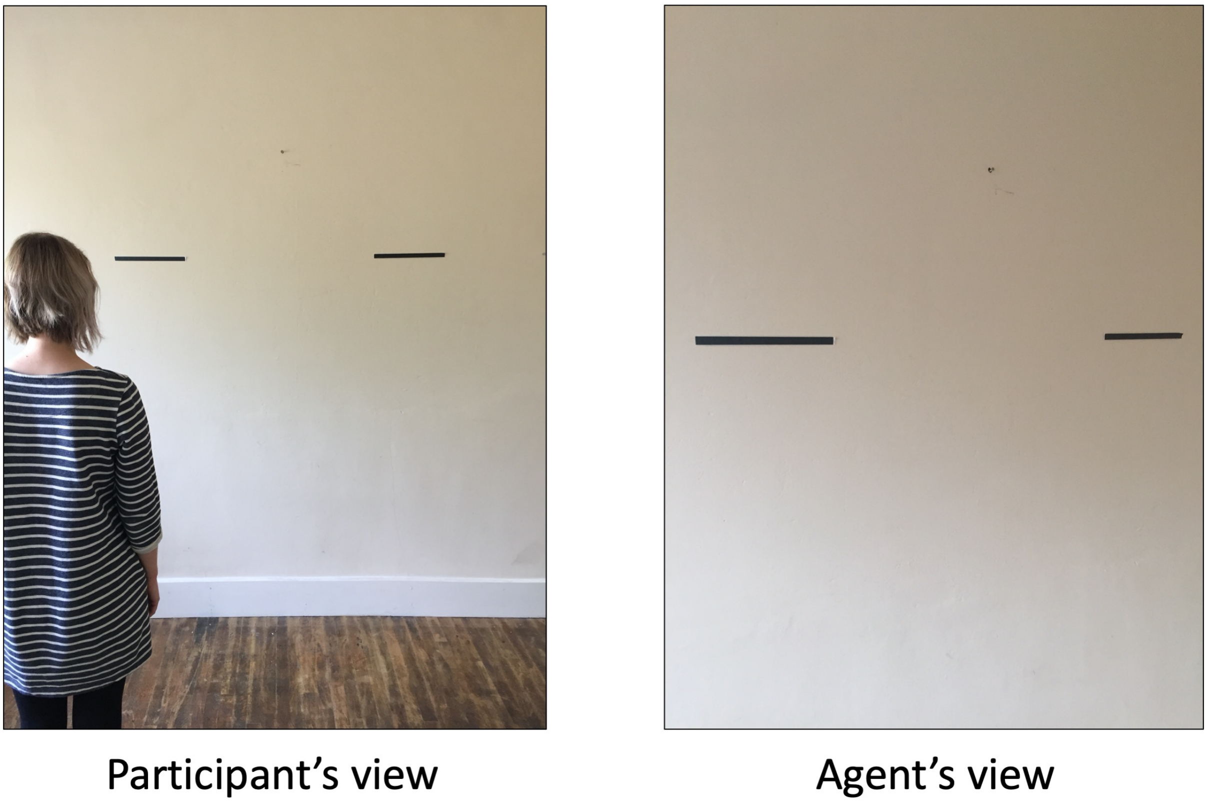


Fig 1. According to the participant's visual perspective (left) the two lines on the wall appear the same length, but given the agent's location in the room the closer line appears longer to her (right).

2. Experiments 1-11

2.1.1 General Method

We conducted a total of 11 online studies. The methods for each experiment overlapped considerably and we have combined the method section as a result.

2.1.2 Participants

Our dependent variables for each experiment were the two length judgments—one for the line closer to the agent, one for the line further from the agent. The alternative hypothesis (H_1) was that judgments for the line closest to the agent should be longer than judgments for the line furthest, as tested by a two-tailed, paired-sample t-test with an alpha of .05 and a power of 95%. We were interested in a medium effect size ($d = 0.5$) for this difference. An *a priori* calculation of the sample size required was conducted using G*Power software, with the resulting output suggesting 54 participants were required. If the data were not normally distributed we planned to perform Wilcoxon signed-rank tests, which required an N of 57. We therefore recruited approximately 65 in each study as we expected to need to make some later exclusions. Specifically, we excluded participants who gave a zero-length judgment for one or both lines, as this suggested either that they believed the line in question was invisible to the agent or that they had not followed instructions. Inclusion criteria were age (18-35), vision (normal or corrected-to-normal), and first language (English). Participants were recruited using Prolific Academic (www.prolific.ac.uk), were paid for their time, and were debriefed at the end of the experiment[2]. The exception was Experiment 11, which recruited from the University of Essex participant pool and offered course credit. Anyone who had participated in any version of this experiment was not permitted to participate in another. Ethical approval was received from the University of Essex Psychology Ethics Committee.

2.1.3 Materials and procedure for all Experiments.

The experiments were conducted using Qualtrics survey software. Participants were informed they were to take part in an experiment investigating the ability to recall the details of an image. They

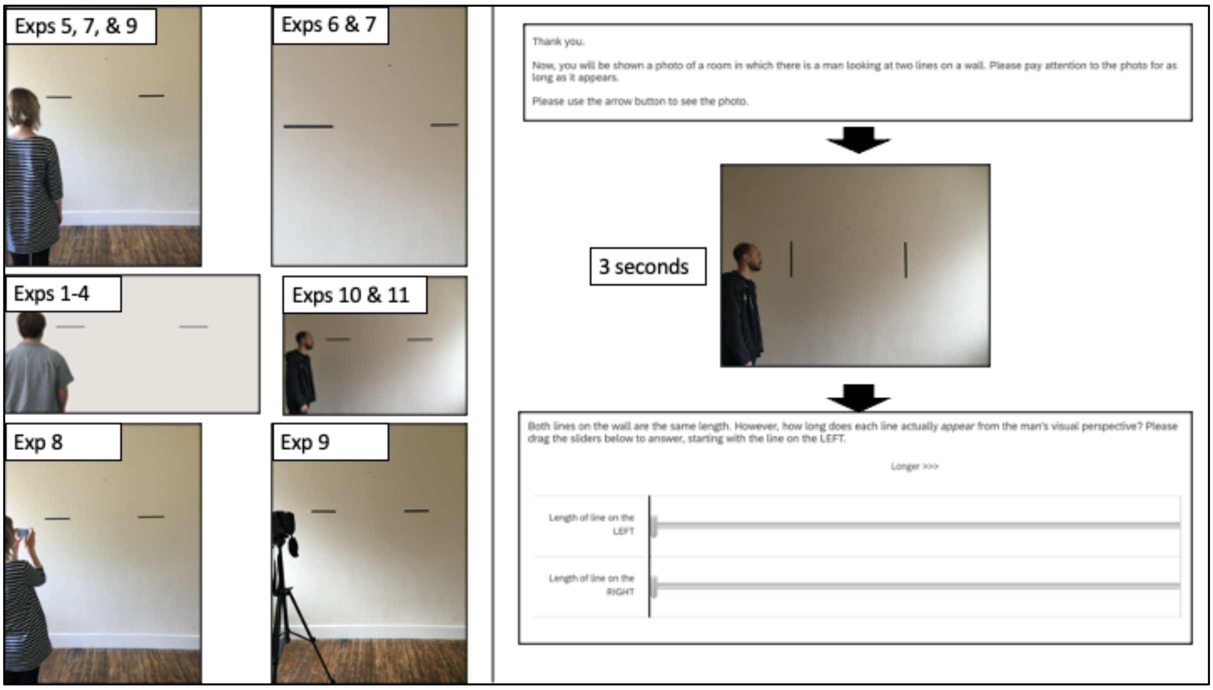
[2] We took a number of steps to ensure that the data we used in our analyses came from ‘true’ participation rather than bots or individuals who might click through a task at maximum speed without attending to the information. Our consent questions all began with the ‘no’ response option, meaning that clicking through first options would simply abort the experiment before it had begun. We employed free text boxes rather than menu responses for personal information questions such as age, gender, and first language. Our sliding scale measure not only made it impossible to click through this main portion of the task (movement of the slider, which began at zero, was required to leave the page).

were asked to pay attention to a picture that would be displayed for a few seconds as they would be asked two questions about it afterwards. The image was timed to appear for 3 seconds and then disappear. Participants were then asked to indicate on two separate sliders the length of each line in the picture they had just seen. In Experiment 1 these sliders appeared on separate screens, but thereafter they appeared together. In all experiments half of the participants made their judgments of the left line first, half the right. No grid lines, numbers or other markers were visible on the slider, but the slider produced a score from 0 (shortest) to 100 (longest) for each line.

Different instructions were used depending on the experiment (see Table 1), and different images (see Figure 2). Participants were always randomly assigned to just one of the images. In Experiments 1-4 the image was presented in landscape format (640 pixels wide, 345 pixels tall), and showed an agent (male) looking at a wall on which two horizontal lines of equal length were drawn. The lines were at eye level for the agent, who stood either to the left or right of the them (counterbalanced), such that one line would be perceived to be longer (i.e., it would take up more of the agent's field of view) due to its relative proximity. This was confirmed by the authors who took a photograph of the agent's view and by means of pilot testing in which people held up a piece of paper under the lines and drew the lengths of the lines on it. In Experiments 1-4 the lines were added to the wall after the photo was taken and the wall was edited to be of uniform colour to ensure that the only asymmetry in the image concerned the location of the agent.

For Experiments 5, 7, 8 and 9 the agent was *female*, and the room and the lines on the walls were not digitally added but pinned to the wall itself. The images in these experiments were in portrait format (450 pixels high, 337 pixels wide), and included a portion of the floor for a better depiction of depth (see Experiment 5 for details). Whereas the image with the male agent was digitally flipped horizontally in order to create the left/right versions, for the experiments with the female agent separate photos were taken at each location. In Experiment 6 the image was again in portrait format and depicted the view of the lines as seen by the agent—the photo was taken from the same location as the agent's head. Again, separate photos were taken for the left and right viewpoints. In Experiment 8 the agent was the same female as in Experiments 5, 7 and 9, the format portrait, but this time the agent was shown holding a camera (smartphone) to take a picture of the wall. In Experiment 9 we used two images in which a camera (once on the left, once on the right) was seen facing the wall from the same location as the female agent.

In Experiment 10 we used one image with a male agent looking at horizontal lines and one where the lines were vertical. Each was digitally flipped horizontally to create the image with the agent on the other side. The agent was positioned closer to the wall, still to one side of the lines, and with his body oriented inwards and head facing the lines. These images were 640 pixels wide and 480 high. In Experiment 11 the image with the horizontal lines was used again, and participants made their responses on the sliders while the photograph was still visible (i.e., the three-second presentation limit was removed). As a result of this change, we removed any part of the instructions that described the photo as appearing only momentarily, or that mentioned recalling details of the image.



Thank you.

Now, you will be shown a photo of a room in which there is a man looking at two lines on a wall. Please pay attention to the photo for as long as it appears.

Please use the arrow button to see the photo.

Both lines on the wall are the same length. However, how long does each line actually *appear* from the man's visual perspective? Please drag the sliders below to answer, starting with the line on the LEFT.

Longer >>>

Length of line on the LEFT

Length of line on the RIGHT

Fig. 2: Images used in the experiments, and an example of the online trial procedure taken from Experiment 10 (vertical lines condition). Note that some images were used in more than one experiment, and the ratio of height to width of these images is correct but size is scaled differently for each image for ease of presentation within the frame of this figure.

Table 1. Main instructions to participants across all experiments. Italics and bold font as in originals.

Exp.	Instruction
1	Please indicate the length of the line on the ***left*** side of the wall *as it appeared to the man*. [Second line judgment on separate screen].
2	How long did each line look to the man in the photo? [Both line judgments made on same screen from this experiment forward]
3	The man in the photo knows that both lines on the wall are the same length. However, how long did each line actually *appear* from his visual perspective?
4	Both of the lines on the wall are the same length. However, imagine you were standing where the man was standing. *From this location*, how long would each line appear?
5	The woman in the photo knows that both lines on the wall are the same length. However, how long did each line actually *appear* from her visual perspective?
6	Both lines on the wall are the same length. However, how long did each line actually *appear* from your visual perspective?
7 (Agent)	As 5.
7 (View)	Both lines on the wall are the same length. However, how long did each line actually *appear* from your visual perspective?
8	The woman with the camera knows that both lines on the wall are the same length. However, how long will each line actually *appear* in her photo?
9 (Agent)	Both lines on the wall are the same length. However, how long does each line actually *appear* from the woman's visual perspective?
9 (Cam.)	Both lines on the wall are the same length. However, how long will each line actually *appear* in a photograph taken by the camera?
10	[You will be shown a photo of a room in which there is a man looking at two lines on a wall.] Both lines on the wall are the same length. However, how long does each line actually *appear* from the man's visual perspective?
11	As 10.

2.1.4 Analyses

The data for all eleven experiments are available online (https://osf.io/chw4d). We were interested in the *relative* lengths of the two line judgments as absolute lengths were uninformative. We performed paired-sample t-tests except where data were not normally distributed (Shapiro-Wilks tests, p <.05), when non-parametric Wilcoxon Signed-Rank tests are reported. Following Dienes (2014), we chose to interpret null results as "meaningful" if Bayesian analyses revealed the data were at least three times more probable under the null (H_0) than the alternative hypothesis that the length judgments differed between the two lines (H_1).

As an additional test, we also examined contrast ratios of line length judgements. These were calculated by dividing the judgement for the longer line by the judgment for the shorter line, applying a positive valency where the *closer* line was judged longer and negative valency where the further line was judged longer. This meant that a judgement of 40 for the close line and 20 for the far line would result in a ratio of +2 (40/20, positive because the closer line is judged longer), indicating that the closer line was judged to be twice as long as the further line. Note that if these judgements were 80 and 40 or 100 and 50 this ratio would be the same. In contrast, a judgement of 80 for the close line and 100 for the far line a *negative* ratio of -1.25 because the *further* line was judged longer. These ratios therefore indicate the length of the closer line relative to the further line irrespective of *absolute* length judgements. For the purpose of the analyses, ratios of 1 (i.e., no difference between line length judgements) were coded as zero to ensure that they did not have a positive or negative polarity. This has no effect on the tests of these data, which were always against a null hypothesis that the median ratio was zero (no difference between the lines). Wilcoxon tests were applied to these data owing to non-normal data distributions.

Mean (*M*) line judgments for the first 9 experiments are presented in Figure 3, along with the relevant *p* values for the Wilcoxon/paired-sample t-tests, and Bayes Factors (BF_{10}) for the same comparison. The BF analyses were always based on a paired-sample t-tests as there is currently no agreed-upon method for Bayesian testing with the non-parametric equivalent, and these BFs should be interpreted only as a guide. Median (*Mdn*) judgments are presented in the text with Wilcoxon test results as these are the figures these tests are based on. Finally, Figure 4 displays the full distribution of responses across the first nine experiments.

2.2.1. Experiment 1

2.2.2 Results

Of the 65 participants recruited for this experiment, three had their data excluded for providing at least one zero length judgment, leaving a final N of 62 (M_{age} = 26 yrs, 16 males). There was no evidence that line length judgements differed as a function of the agent's location: Mdn_{Close} = 22; Mdn_{Far} = 24.5; $W(62) = 824.5$, $Z = 0.241$, $p = .810$, $r = .022$, and BF analyses found that the data were 6.4 times more probable under the null hypothesis. A one-sample Wilcoxon test found that the observed median of the ratios of line length judgements was not significantly different from zero: *Mdn* = 0; range = -4.5 to +4.5, $W(62) = 869$, $Z = 0.105$, $p = .917$, $r = .013$, with precisely the same percentage of participants (47%) judging the closer line to be longer and the further line to be longer.

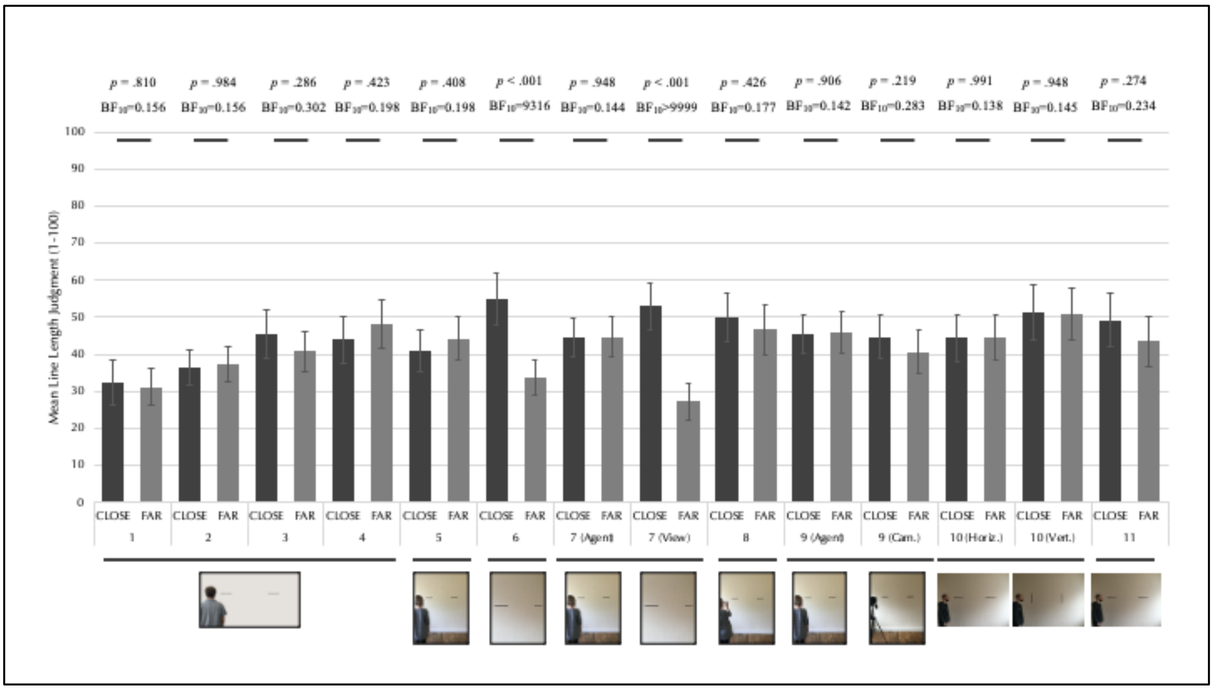


Figure 3. Mean line judgments and associated 95% confidence intervals for Experiments 1-11. An example image from the experiment is shown below the relevant bars, and the results of the Wilcoxon Signed-Rank/paired sample comparisons and associate Bayesian analyses are presented above.

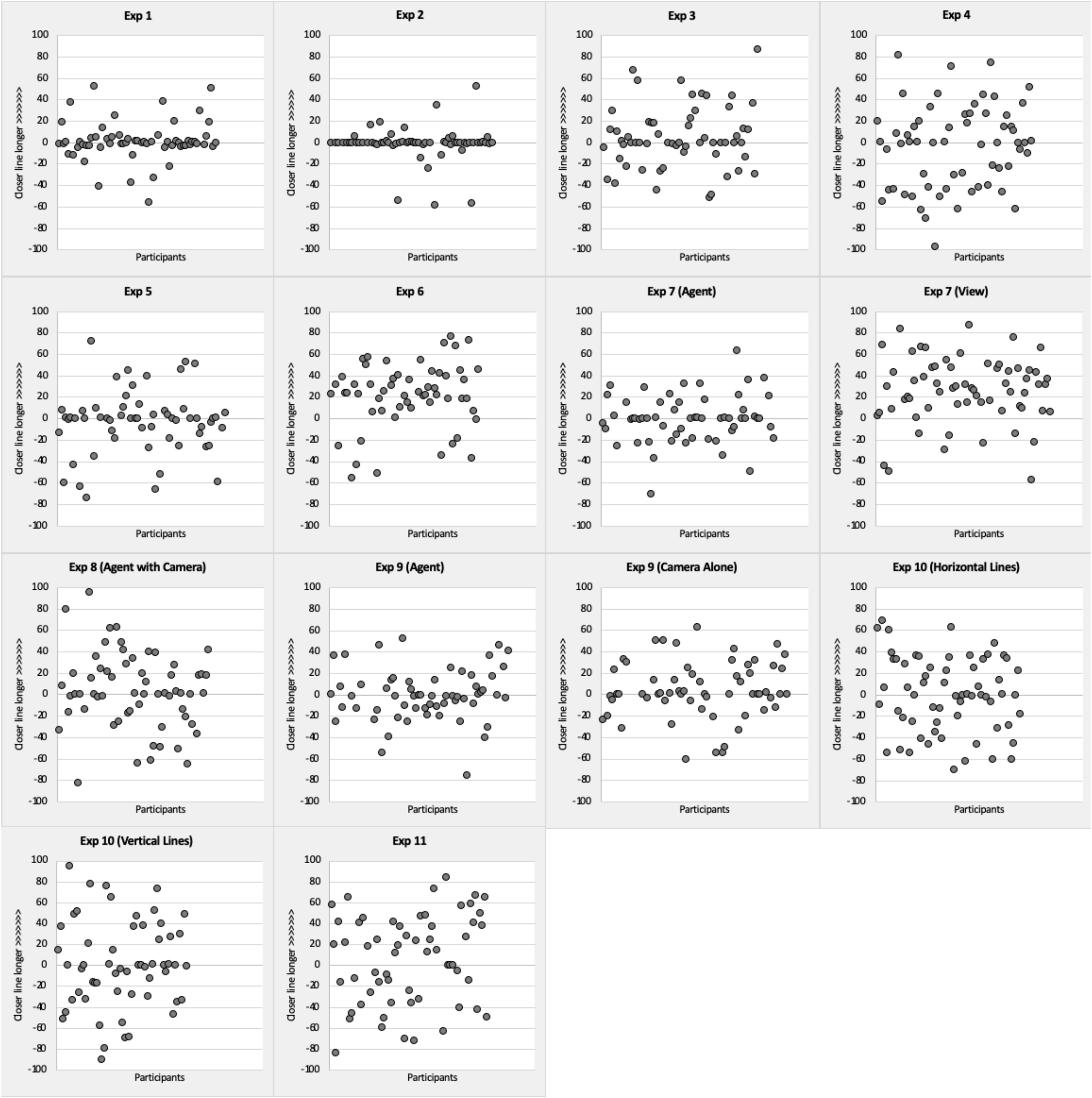


Figure 4. Mean line length judgments for the closer line to the agent minus mean judgments for the further line. Higher values thus represent longer judgments for the closer line. Each data point represents one participant.

2.3.1 Experiment 2

One possible reason for the absence of an effect in Experiment 1 was that the requirement to make judgments for each line on separate screens made it harder for participants to accurately compare their responses. In Experiment 2, we showed participants both line judgment sliders on the same screen. We edited the text of the instruction to accommodate this change, and instructed participants to begin with the top slider (half the time this was the left line, half the right).

2.3.2 Results

Of the 65 participants recruited for this experiment, one had their data excluded for providing at least one zero length judgment, leaving a final N of 64 (M_{age} = 27 yrs, 28 males). Again there was again no evidence that line length judgments differed as a function of the agent's location: Mdn_{Close} = 33.5; Mdn_{Far} = 36; $W(64) = 247$, $Z = 0.020$, $p = .984$, $r = .002$. BF analyses found that the data were again 6.4 times more probable under the null hypothesis. A one-sample Wilcoxon test found that the observed median of the ratios of line length judgements was not significantly different from zero: *Mdn* = 0; range = -4.4 to +2.7, $W(64) = 246.5$, $Z = 0.029$, $p = .977$, $r = .004$, with 25% of participants judging the closer line to be longer and 23% the further line.

2.4.1. Experiment 3

In Experiment 3, we made even clearer to participants that we wanted them to judge the *appearance* of the length of the lines rather than their actual length (or knowledge of their lengths). We therefore edited the instruction to read: "The man in the photo knows that both lines on the wall are the same length. However, how long did they actually *appear* from his perspective?" If participants were making their judgments based on the agent's knowledge rather than the agent's perception, this change should emphasise that the question is not about knowledge but visual appearance.

2.4.2 Results

Of the 66 participants recruited for this experiment, five had their data excluded for providing at least one zero length judgment, and two for stating that they were not native English speakers, leaving a final N of 59 (M_{age} = 27 yrs, 17 males). Consistent with the first two experiments, there was again no evidence that line length judgments differed as a function of the agent's location: Mdn_{Close} = 46; Mdn_{Far} = 39; $W(59) = 484$, $Z = 1.067$, $p = .286$, $r = .098$. BF analyses found that the data were 3.3 times more probable under the null hypothesis. A one-sample Wilcoxon test found that the observed median of the ratios of line length judgements was not significantly different from zero: *Mdn* = 0; range = -14.5 to +22.8, $W(59) = 661.5$, $Z = 0.754$, $p = .451$, $r = .098$, with 44% of participants judging the closer line to be longer and 37% the further line.

2.5.1. Experiment 4

Across three experiments and despite explicit and even leading questions we had failed to find evidence that participants could make judgments consistent with an image-like representation of another agent's vision. In Experiment 4, we tested the possibility that participants would be able to take their *own* perspective of the lines of the wall, as it were, by asking them to imagine they stood where the agent stood. If the difficulty in accurately assessing relative line length comes from the demands of taking another agent's perspective, then switching the task to imagining what the participant herself might cause a difference to emerge.

2.5.2 Results

Of the 66 participants recruited for this experiment, four had their data excluded for providing at least one zero length judgment, leaving a final N of 62 (M_{age} = 25 yrs, 35 males, 1 non-binary). Again, there was no evidence that line length judgments differed as a function of perspective—this time the participant's own, imagined perspective: Mdn_{Close} = 44; Mdn_{Far} = 45; $W(62) = 959$, $Z = 0.801$, $p = .423$, $r = .072$. BF analyses found that the data were 5 times more probable under the null hypothesis. A one-sample Wilcoxon test found that the observed median of the ratios of line length judgements was not significantly different from zero: *Mdn* = 0; range = -33.3 to +11.7, $W(62) = 773.5$, $Z = 0.635$, $p = .525$, $r = .081$, with 48% of participants judging the closer line to be longer and 45% the further line.

2.6.1. Experiment 5

In Experiment 5, we explored the possibility that the reason for these null results was that there were not enough depth cues in the image to elicit perspective effects. The image we had been using relied on the agent's location relative to the lines, but it could have been clearer how near or far the agent was from each line. In Experiment 5 we used a new image. This time, a female agent stood in a room with two lines of equal length attached to a wall. The picture was switched from landscape to portrait format to include more of the floor and hence provide a clearer indication of the agent's location relative to the lines. Additionally, we did not alter the image in any way—it was a photo of a real room, a real agent, a real wall, and real lines. If the reason for the absence of visual perspective-taking in Experiments 1-4 was because the image lacked a clear sense of depth or realism more generally, then we should find that the new image elicits longer line judgments for the line closest to the agent.

2.6.2 Results

Of the 66 participants recruited for this experiment, five had their data excluded for providing at least one zero length judgment. Five participants contacted the research team to say the image had failed to load, and two of these had gone on to offer responses regardless. The data from these two were also excluded, leaving a final N of 59 (M_{age} = 26 yrs, 23 males, 1 non-binary). Again there was no evidence that line length judgments differed as a function of the agent's location: M_{Close} = 41; M_{Far} = 44; $t(58) = 0.833$, $p = .408$, $d = .108$. BF analyses again found that the data were 5 times more probable under the null hypothesis. A one-sample Wilcoxon test found that the observed median of the ratios of line length judgements was not significantly different from zero: *Mdn* = 0; range = -12 to +7.4, $W(59) = 517.5$, $Z = 0.945$, $p = .345$, $r = .123$, with 41% judging the closer line to be longer and 42% the closer line.

2.7.1. Experiment 6

Experiment 5 had again demonstrated that participants were not representing the agent's visual perspective. Moreover the results of previous experiments were unlikely to be due to any insufficiencies in depth cueing or realism. In Experiment 6 we conducted a crucial control experiment. We showed participants an image of the lines as taken by a camera in the same location as the female agent's head. No agent was shown or referred to; participants were simply asked to judge the lengths of the lines in the image. Additionally, given the reports we obtained of technical errors in the last experiment, in this experiment we explicitly asked participants at the end of the task whether the image had loaded and they had seen it.

2.7.2 Results

Of the 61 participants recruited for this experiment, three had their data excluded for providing at least one zero length judgment, and one stated they were in fact not a native speaker of English, leaving a final N of 57 (M_{age} = 27 yrs, 22 males, 1 non-binary). A technical error meant that the first five participants didn't receive the post-test question, but all the other 52 participants said that they had seen the image, so we analysed the data from all participants. The data did not deviate from normality, so we proceeded with parametric paired-sample t-tests. For the first time participants *did* judge the closer line to be longer: M_{Close} = 55; M_{Far} = 34; $t(56) = 5.336$, , $p < .001$, $d = .707$, and BF analyses found that the data were 9,316 times more probable under the alternative hypothesis than the null. A one-sample Wilcoxon test found that the observed median of the ratios of line length judgements was for the first time significantly different from zero, with the closer line judged to appear approximately twice as long as the further line: *Mdn* = 2; range = -15.3 to +20.3, $W(57) = 1344.5$, $Z = 4.116$, $p < .001$, $r = .545$, with 82% of participants judging the closer line to be longer and 18% the further line.

2.8.1. Experiment 7

Experiment 6 demonstrated that participants have no problem judging an agent's view of the lines when this view was shown to them 'pre-represented' in image form. Note also that the effect size for the comparison (*d* = .707) is particularly large. We ran a further study in which we recruited double the number of participants so that we could randomly assign them to either the perspective-taking task in Experiment 5 or the control task in Experiment 6. We were interested in whether there would be an interaction indicating that closer lines were judged to be longer in what we termed the 'Agent's View' condition (as per Experiment 6) relative to the 'Agent' condition (as per Experiment 5).

2.8.2 Results

Of the 131 participants recruited for this experiment, six had their data excluded for providing at least one zero length judgment, one for not being a native speaker of English, and one for responding to the post-test question with the answer that the image didn't load. This left a final N of 123, with 58 in the Agent condition (M_{age} = 26 yrs, 20 males) and 65 in the Agent's View condition (M_{age} = 26 yrs, 29 males, 1 non-binary, 1 gender-fluid).

We conducted a 2: (Group: Agent vs. Agent's View) x 2: (Distance: Close vs. Far) mixed-design ANOVA on the data. The data distributions for the two cells pertaining to the Agent group both deviated from normality. The number of cells that were non-normal increased to three after log transformation. There were also issues relating to homogeneity of variance and covariance. For the greatest ease of comparison with previous experiments and to test for evidence of an interaction we proceeded with the planned ANOVA with raw scores but also ran parallel non-parametric tests at the within-group level to check for corroborating evidence. The pattern of results was entirely consistent across all statistical approaches.

The ANOVA found a main effect of Distance, $F(1, 121) = 27.354$, $p < .001$, $\eta_p^2 = .184$, and crucially an interaction, $F(1, 121) = 28.474$, $p < .001$, $\eta_p^2 = .19$. Follow-up Wilcoxon Signed-Rank tests found no significant difference between line length judgments in the Agent condition: Mdn_{Close} =

44.5; Mdn_{Far} = 47; W(58) = 546.500, Z = 0.066, p = .948, r = .006. Bayesian analyses, based on a paired-sample t-test, found that the data were 7 times more likely under the null than the alternative that there is a difference in this group. In contrast, length judgments of the closer line were longer than those of the more distant line in the Agent's View condition: Mdn_{Close} = 52; Mdn_{Far} = 22; W(65) = 268, Z = 5.258, p < .001, r = .461. Bayesian analyses found the data in this group were over 2 million times more likely under the alternative hypothesis that there is a difference than the null. There was no effect of Group, F(1, 121) = 2.220, p = .139, η_p^2 = .018. These findings were also supported by one-sample Wilcoxon tests, showing that the observed median of the ratios of line length judgements was not significantly different from zero in the Agent group: Mdn = 0; range = -6 to +3.4, W(58) = 534.5, Z = 0.066, p = .948, r = .009, with the same percentage of participants judging the closer line to be longer and the further line to be longer (40%). However the closer line was judged to be over twice as long in the Agent's View group: Mdn = 2.39; range = -4.7 to +26.3, W(65) = 1933.5, Z = 5.627, p < .001, r = .698, with 86% of participants judging the closer line to be longer and 14% the further line. Supporting the difference in the parametric test on raw judgement scores, a Mann Whitney U test also found a significant difference in the distribution of ratios between the two groups, U (123) = 3003.5, p = <.001.

2.9.1 Experiment 8

Experiment 7 demonstrated again that participants have no problem judging an agent's visual perspective of lines if they are shown it, they only have a difficulty if they are asked to imagine it. In an eighth experiment, we tested two potential alternative explanations for our data. Firstly, it may be that despite the phrasing of our questions participants continue to incorporate the agent's *knowledge* that the two lines are the same length when making their judgments. According to this view, participants who judge that the agent sees the two lines the same length could be considered as being *adept* perspective-takers because they represent that agent's knowledge successfully; they use their theory of mind (Premack & Woodruff, 1978). Secondly, it could be that participants interpret the question to be about how long the lines actually are rather than how long they appear from different vantage points, even despite our attempts to get them to focus on the latter. Note that for either of these hypotheses to be correct, it would require that participants had been consistently misinterpreting the clear instruction *not* to make judgments based on the agent's knowledge *or* their own in previous experiments.

A classic control for mental state attributions to an agent is a functionally similar judgment based instead on a photograph (e.g., Zaitchik, 1990). The contents of a photograph do not carry knowledge; for instance, a photo cannot be *false* in the way agent's beliefs can be—they are merely accurate snapshots of what is in front of them at a given time (Perner & Leekam, 2008). Photos are also only visual in content; there would be no other interpretation of a question to judge the lengths of the lines in the photograph. In Experiment 8, we therefore presented participants with an image of the same (female) agent as in Experiments 5 and 7, but this time she was holding up a smartphone to take a picture of the wall. Instead of instructing participants to take the agent's perspective, we now asked them how long the lines would appear in the photo that she took. They were never shown the contents of the photo. If participants continue to fail to judge the closer line to look longer then their failure cannot be due to the application of knowledge or a misunderstanding of the question.

2.9.2 Results

Of the 63 participants recruited for this experiment, two had their data excluded for providing at least one zero length judgment, leaving a final N of 61 (M_{age} = 27 yrs, 28 males). All participants reported they had seen the image. Again, there was no evidence that line length judgments differed as a function of the location from which the photograph was taken: Mdn_{Close} = 50; Mdn_{Far} = 44; $W(61)$ = 675, Z = 0.796, p = .426, r = .072. BF analyses found that the data were 5.6 times more probable under the null hypothesis. A one-sample Wilcoxon test found that the observed median of the ratios of line length judgements was not significantly different from zero: Mdn = 1; range = -22.3 to +25, $W(61)$ = 866.5, Z = 0.809, p = .419, r = .104, with 51% judging the closer line to be longer and 39% the further line.

2.10.1 Experiment 9

The results of Experiment 8 suggested that the difficulty participants had with perspective taking was not due to a difficulty in correcting for the agent's knowledge or a misinterpretation of the question to make judgments according to visual appearance. Taken together, the results of these eight experiments suggest that we have difficulty making accurate judgments about visual appearance from other agents' perspectives and other vantage points. However, a potential weakness in Experiment 8 was that there may have been some residual effect of the presence of the agent in the image that introduced her knowledge into participants' processing. In Experiment 9 we showed participants either the scene with the agent (as in Experiments 5 and 7) or a new image with a camera on a tripod in the same location as the agent. Participants were asked to judge the lengths of the lines from either the agent's perspective or according to what the photo taken by the camera would depict. We again hypothesised that if participants correct appearance judgments for knowledge then this correction should be evident in the form of an interaction showing the closer line to be longer than the further line in the camera condition relative to the agent condition.

2.10.2 Results

Of the 132 participants recruited for this experiment, eight had their data excluded for providing at least one zero length judgment, two for not being native speakers of English, and one for responding to the post-test question with the answer that the image didn't load. This left a final N of 121, with 60 in the Agent condition (M_{age} = 28 yrs, 26 males) and 61 in the Camera condition (M_{age} = 28 yrs, 20 males).

The data distributions for one of the four cells (Camera, Far) deviated from normality and log transforming the data made all cells non-normal. We thus retained parametric testing on the raw data but also ran a parallel non-parametric test at the within-group level for the group that saw the Camera image. Note that the pattern of results was consistent across all statistical approaches.

We conducted a 2: (Group: Agent vs. Camera) x 2: (Distance: Close vs. Far) mixed-design ANOVA on the data. There was no effect of Distance, $F(1, 119) = 0.672$, $p = .414$, $\eta_p^2 = .006$, no effect of Group, $F(1, 119) = 0.863$, $p = .355$, $\eta_p^2 = .007$, and crucially no interaction, $F(1, 119) = 0.959$ $p = .329$, $\eta_p^2 = .008$. However, separate tests by group were conducted to ensure an appropriate non-parametric test was applied to those data that were not normally distributed in the Camera group. These found no significant difference between line length judgments in the Agent condition: $M_{Close} = 45.6$; $M_{Far} = 46.0$; $t(59) = 0.118$, $p = .906$, $d = .015$, with 42% of participants judging the closer line to be longer and 53% the further line, or the Camera condition, $Mdn_{Close} = 47$; $Mdn_{Far} = 36$; $W(61) = 489$, $Z = 1.229$, $p = .219$, $r = .111$, with 46% of participants judging the closer line to be longer and 34% the further line. Bayesian analyses found that the data were 7 times more likely under the null in the Agent group and 3 times more likely under the null in the Camera group. Bayesian analyses also found the data were three times more likely under the null hypothesis that there was no difference in relative distance judgments between the agent and camera conditions ($BF_{10} = 0.303$ on the interaction). These findings were also supported by one-sample Wilcoxon tests, showing that the observed median of the ratios of line length judgements was not significantly different from zero in either the Agent group: $Mdn = -1$; range = -11.8 to +9.2, $W(60) = 785.5$, $Z = 0.326$, $p = .745$, $r = .042$, or the Camera group: $Mdn = 0$; range = -3.4 to +8.2, $W(61) = 741.5$, $Z = 1.283$, $p = .199$, $r = .164$. Supporting the absence of any between-group difference in the parametric test on raw judgement scores, a Mann Whitney *U* test also found no significant difference in the distribution of ratios between the two groups, $U(121) = 20079$, $p = .196$.

2.11.1 Experiment 10

Participants had so far consistently failed to accurately judge that the closer line to the agent appeared visually longer. In Experiment 10 we made two important changes. Firstly, the agent (now male) now looked at the lines with his body at a right angle to both the lines and the participant. This meant that the agent's posture and location was much more clearly different from the participant's, and by extension his perspective of the lines on the wall. This should increase the likelihood that participants understand that the agent has a different visual perspective from their own. Secondly, in one condition we changed the lines on the wall so that they were now vertical. This addressed the possibility that it is more difficult to accurately represent the visual effect of how horizontal stimuli stretch into depth. Vertical lines do not have this property, and therefore any additional difficulty with representing the perspective of horizontal lines should be eliminated with vertical lines. Note that it is questionable whether perceptual simulation as currently conceived would predict that some percepts are more difficult to represent than others, but it would nevertheless be useful to establish whether this is the case. In a second condition, we retained the horizontal lines stimuli. We changed the instructions to make it clear that the agent could see the lines from where he stood.

2.11.2 Results

Of the 129 participants recruited for this experiment, eight had their data excluded for providing at least one zero length judgment and one for not being a native speaker of English. This left a final N of 120, with 63 in the Horizontal Lines condition (M_{age} = 28 yrs, 18 males, 1 non-binary, 1 gender-fluid) and 57 in the Vertical Lines condition (M_{age} = 27 yrs, 20 males). One of the four data cells showed evidence of non-normal data distribution ('far' line judgments in the group which saw vertical lines, $p = .046$, all other *p*s > .06). Log transformation of the data did not normalize the data but actually made all cells non-normal. We thus proceeded with the raw data for the purposes of a 2: Group (Horizontal vs. Vertical) x 2: Distance (Close vs. Far) mixed-design ANOVA with repeated measures over the second factor, but we also report tests for the two groups separately, which allows us to report an appropriate non-parametric test for the non-normal data.

The ANOVA found a main effect of Group, $F(1, 118) = 4.573$, $p = .035$, $\eta_p^2 = .037$, with average line length judgments scores being 7 points smaller in the Horizontal group ($M = 44$) than the Vertical group ($M = 51$). However, this difference does not speak to the contrast between the two lines, for which we would expect a main effect of Distance. Crucially, there was no main effect of Distance, $F(1, 118) = 0.005$, $p = .946$, $\eta_p^2 = 0$ or any interaction, $F(1, 118) = 0.007$, $p = .935$, $\eta_p^2 = 0$.

The analyses by group found no evidence that horizontal line length judgments differed as a function of the perspective, $M_{Close} = 44.4$; $M_{Far} = 44.4$; $t(62) = 0.011$, $p = .991$, $d = .001$. BF analyses found that the data were 7.3 times more probable under the null hypothesis. The results from the

Vertical Lines condition also showed no evidence that line length judgments differed as a function of the perspective, Mdn_{Close} = 53; Mdn_{Far} = 47; *W*(57) = 670, *Z* = 0.066, *p* = .948, *r* = .001. BF analyses found that the data were 6.9 times more probable under the null hypothesis. These findings were also supported by one-sample Wilcoxon tests, showing that the observed median of the ratios of line length judgements was not significantly different from zero in either the Horizontal Lines group: *Mdn* = 0; range = -16 to +30, *W*(63) = 956, *Z* = 0.536, *p* = .592, *r* = .068, with 48% of participants judging the closer line to be longer and 46% the further line, or the Vertical Lines group: *Mdn* = 0; range = -27.3 to +20, *W*(57) = 624.5, *Z* = 0.361, *p* = .718, *r* = .048, with 40% of participants judging the closer line to be longer and 49% the further line. Supporting the absence of any between-group difference in the parametric test on raw judgement scores, a Mann Whitney *U* test also found no significant difference in the distribution of ratios between the two groups, *U* (120) = 1653, *p* = .590.

2.12.1 Experiment 11

In Experiment 11 we repeated Experiment 10 (horizontal lines condition), but this time we presented the sliders and the photograph on the same page and removed the 3-second appearance time limit for the image. By doing so, we aimed to rule out a number of potential explanations for the absence of evidence of perspective-taking so far. Firstly, if participants were failing to represent the agent's visual perspective because they did not have time to process the image, the present experiment removes this restriction. Secondly, if they failed because they did not know that they would be asked the length of the lines while they looked at the photograph, this design also eliminates this possibility, as both the photograph and response method are presented simultaneously. Thirdly, if the time lag between the disappearance of the photo and the response caused the decay of a successful representation of the agent's perspective, then this delay was now also eliminated. Relatedly, although in all the previous experiments we instructed participants to take the other agent's point of view, a design factor that situates our studies within the remit of explicit perspective-taking, the present experiment eliminates any reliance there might have been on spontaneous perspective-taking owing to the limited duration of the image before responding.

2.12.2 Results

Of the 64 participants recruited for this experiment, five had their data excluded for providing at least one zero length judgment. This left a final N of 59 (M_{age} = 19 yrs, 23 males). Again, there was no evidence that line length judgments differed as a function of the perspective: Mdn_{Close} = 48; Mdn_{Far} = 35; *W*(59) = 664, *Z* = 1.093, *p* = .274, *r* = .1. BF analyses found that the data were 4.3 times more probable under the null hypothesis. A one-sample Wilcoxon test found that the observed median of the ratios of line length judgements was not significantly different from zero: *Mdn* = 1.48; range = -10.3 to +6.3, *W*(59) = 927, *Z* = 1.052, *p* = .293, *r* = .137, with 53% of participants judging the closer line to be longer and 42% the further line.

4. General Discussion

Across eleven experiments we found no evidence that participants could make judgments consistent with any visual perspective other than their own. Participants consistently failed to judge that a line closer to another agent would appear longer than a line further away. This failure occurred whether participants made their line length judgments on the same screen (Experiment 1) or separate screens (Experiment 2). It persisted when the need to make their judgments based on appearance rather than reality was emphasized (Experiment 3 onwards) and when they were asked to imagine themselves in the agent's place (Experiment 4). It also persisted when clear depth cues were added (Experiments 5, 7 and 9), when the lines were oriented vertically instead of horizontally (Experiment 10), when the other agent's location and posture was more clearly different from the participant's own and the ratio of line length difference increased (Experiments 10 and 11), and when length judgments could be made at leisure with the image visible throughout (Experiment 11). When we asked participants to imagine what a photo taken from the same location as the agent would show, they still did not respond that the closer line would appear longer (Experiments 8 and 9). These results in particular rule out an account by which participants failed in the agent conditions because they based their responses on the agent's knowledge, as photos have no knowledge of what they depict. They also rule out the possibility that participants failed because they interpreted the questions as about something other than visual appearance. These results suggest that participants did not or could not generate images of perspectives that were not their own, at least certainly not in large enough numbers to generate a statistically significant effect. However, when instead of asking participants to take the agent's perspective we simply *showed* them that perspective, a powerful effect emerged; the closest line was judged to appear longer than the furthest line.

4.1 Does VPT involve perceptual simulation?

Overall, our results suggest that adults generally fail to imagine how things appear visually from non-egocentric vantage points. These data thus contradict the theory that we simulate other agents' perception, or indeed what would be perceived from another uninhabited location. However, there is ample evidence that adults *can* solve appearance questions, such as comprehending that a 6 can look like a 9 depending on where it is viewed, or that what is on one person's left can be on another's right, and so on. One reason for the relative success participants display in such tasks might be that these problems are solved not by representing the appearance of the stimuli or scene in any image-like way but by other means, such as the reconfiguration of one's motor representations to facilitate judgements based on relative spatial locations (e.g., Deroualle, Borel, Devèze, & Lopez, 2015; Kessler & Thomson, 2010; Samuel, Legg, Manchester, Lurz, & Clayton, 2019), or by means of heuristics such as those outlined in the Introduction. The reason for the difference between the results of our study and others might be found in the nature of our stimuli. As we pointed out in our Introduction, an important aspect of the lines in our tasks is that they change according to viewing angle but in a continuous, analogue fashion rather than a discrete, digital way. This change is captured best by vision rather than categories or language. A simple test for this distinction is to check whether the target stimulus in a task is labelled differently according to perspective. A six is not a six unless it is viewed a certain way, ditto left and right. There is a great deal of evidence that we process what we see differently depending on whether stimuli vary categorically or on a continuum, usually with more efficient processing for the former (Gilbert, Regier, Kay, & Ivry, 2006, 2008; Holmes, Moty, &

Regier, 2017; Roberson & Hanley, 2010; Roberson, Pak, & Hanley, 2008; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009; Winawer et al., 2007; Yun & Choi, 2018). It might therefore be easier to solve VPT problems about stimuli that vary categorically by perspective because they are compatible with the ways we can process symbolic information more generally, making non-imagistic strategies available. Our stimuli exposed this in a way other stimuli do not.

A related possibility is that categorical stimuli are generally easier to represent because we can more readily generate mental imagery of them. These images would then be manipulated to solve a VPT problem. For instance, as a thought experiment we could easily call up images of a 6 or a 9 and 'flip' that image because we have visual archetypes for them. Imagining a line and then a slightly *longer* line requires more creativity, and the resulting images might be harder to use. VPT as a process employing visual archetypes is more compatible with the notion of VPT as about imagery, but it does not require any sense of simulating another agent's vision. Indeed, whether VPT is about category manipulation, heuristics, or visual archetypes, each implies a subtle but important shift in the emphasis *away* from the other agent's perceptual experience and back to the perspective taker; rather than use the other agent's experience to inform ours, we use our experience to inform *theirs*.

Could it be that despite our attempts to ensure that participants understood the questions were about appearance, they nevertheless fell back on a response about knowledge? If so, the perceptual simulation account would not be contradicted, because we would have failed to elicit representations of vision in the first place. There are two principal arguments against this, one empirical and one theoretical. The empirical evidence comes directly from our camera conditions. These showed that even when appearance was the only option participants still failed to judge that the closer line would appear shorter in the photograph. The theoretical argument concerns instructions. We conducted multiple experiments using this paradigm because we wanted to ensure that the absence of an effect was not the result of participants understanding the question to be about something other than appearance, such as knowledge. As the experiments were completed and the results analysed, we made the wording of these questions more and more explicit, to the extent that if we *had* found evidence that participants made judgments based on appearance, it could very easily be argued that we had biased participants too strongly in this direction. The crucial point here is that if these instructions failed to elicit perceptual simulation then studies with less explicit instructions are unlikely to have done so either, and alternatives to perceptual simulation in these studies should be considered instead.

One aspect of the perceptual simulation account is that it should occur spontaneously (Ward et al., 2019; Ward et al., 2020), but a version of perceptual simulation that does not require it to be spontaneous would still have significant explanatory power. In the present experiments our questions to the participant were always explicit, whether they came immediately after the image had disappeared (Experiments 1-10) or while the image was being viewed and there was no time limit on responses (Experiment 11). Our data thus suggest that perceptual simulation does not occur in enough participants to generate a reliable result either explicitly or, by extension, spontaneously.

Is it possible to integrate our findings with perceptual simulation? There are two ways in which this might be possible, but to our minds neither is satisfactory. Firstly, we cannot rule out from our data that a minority of participants generated images that copied the agent's visual perspective, in the manner described in the perceptual simulation account. This could explain those cases where closer lines were judged to be longer. However, a minority do not support a theory of VPT as conducted by perceptual simulation. Different means of VPT would have to be considered at least as common, if not more common. This need not be fatal to the account, but it would make it optional, perhaps preferred by some individuals but not others. Secondly, it might be argued that perceptual simulation involves representing *processed* visual input which factor in size constancy, depth perception, and so on. According to this view, the reason that we found no evidence that the closer line was judged longer is because the agent's visual systems corrected for the stretch-into-depth of the lines, just as we do (usually very successfully) in everyday life. If so, participants were thus *accurate* in their judgements that they were the same length. However, this possibility also rejects the perceptual simulation account. Recall that mental images preserve spatial properties of their percepts, such that it takes longer to move from one point in an image to another just as it would if the image were an object (Kosslyn, Ball, & Reiser, 1978). Crucially then, if we allow corrections for depth perception what we have is not an image but an abstraction; we cannot “paint” depth perception and size constancy into an image. In any case, correction for depth perception does not provide a plausible account for our data, because correcting for depth should make participants judge the lines to be the same length, but only a minority displayed this behaviour, with the group averages being created by a broad range of responses. The data from those conditions where we asked participants to judge the lengths of the lines in a photo also argue against corrections leading to null results, for cameras do not possess human depth perception but results patterned the same as with human agents.

Our results tally with some theoretical criticism of the plausibility of mental imagery both in solving VPT problems and in cognition more broadly. Cole and Millett (2019) make the argument that it is not possible to represent the visual experience of other agents, not only because an observer does not have access another person’s perceptual systems, but also because a pictorial representation runs afoul of the homunculus problem, requiring an ‘inner eye’ to inspect. These issues are cumulative with evidence that some VPT effects that were originally attributed to another agent’s vision have been shown to persist not only when the agent cannot see the stimuli in question owing to barriers (Cole et al., 2016; though see Furlanetto, Becchio, Samson, & Apperly, 2016, for a counter claim) but also when the agent is replaced with arrows (Santiesteban et al., 2014) and chairs (Cavallo, Ansuini, Capozzi, Tversky, & Becchio, 2017; Millett, D’Souza, & Cole, 2019), which cannot see at all. As Cole and Millet make clear, much of these criticisms are a logical extension into the field of VPT of the ‘great debate’ over the plausibility of mental imagery that retains properties of its subject (Kosslyn, Ganis, & Thompson, 2001; Pylyshyn, 2003). Perceptual simulation, with its emphasis on imagery, is broadly in alignment with Kosslyn’s view, and is therefore subject to the same advantages and criticisms that the original debate raised (see also Cole et al., 2020).

It is important to note that perceptual simulation was a hypothesis developed in response to data, and therefore if perceptual simulation does not explain Ward and colleagues’ results (Ward et al., 2019; Ward et al., 2020), what does? A plausible account to our minds is a cognitively low-level alternative, which is that given the otherwise arbitrary choice of rotating the character clockwise or anti-clockwise to make their judgement, participants rotated the character in the direction that produced the shortest path from the agent's viewing angle to the participant's perspective. In other words, when the agent appeared on the left participants rotated the character counter-clockwise, and when the agent was on the right they rotated the character clockwise. This approach would generate the same pattern of results without any need to represent the agent's perspective of the character at all. This is a plausible strategy because experience tells us that people usually sit somewhere where they can view things in their normal, upright position, and thus the shortest path to that position is likely to provide the correct orientation of the target more quickly. This is *especially* true of alphanumeric characters. This strategy also explains why replacing the agent with a lamp did not produce the same effect (Experiment 2, Ward et al., 2019), and why diverting the agent's gaze failed to eliminate the effect; their normal viewing axis is the important factor key, not their unusual (and thus presumably fleeting) gaze direction (Ward et al., 2020).

4.2 How then do we solve VPT problems of appearance?

What then do our results *support*? Despite the issues our results create for perceptual simulation, our findings could nevertheless be consistent with the possibility that *some* people simulate other agents' perception, just not enough to make a statistically reliable effect for a group of participants. Such an interpretation requires a greater understanding of individual differences in approaches to VPT problems. The spectrum of these differences is likely to be broad. A glance at the distribution of responses across our experiments shows that participants were often as likely to err in the direction of the closer line being *shorter* as longer, and often only a minority judged the lines to be of equal length. The former finding argues against the possibility that participants were simulating but just doing so inaccurately, because we would not expect so many participants to pattern in the *opposite* direction to what should be predicted. Support for inaccurate simulation would instead come from variation in *how much longer* the closer line appears than the further line.

Instead of perceptual simulation, our data thus seem to be better captured at least in spirit by naïve optics. Naïve optics is nicely exemplified by the Venus Effect, whereby an observer sees an agent and a mirror and believes that the agent sees their reflection in the mirror as the observer does, despite the agent and mirror not being along the observer's line of sight (Bertamini, Latto, & Spooner, 2003; Bertamini & Soranzo, 2018). The effect has been reported to occur in approximately 75% of adults (Bertamini et al., 2003). Effects like this have been attributed to folk beliefs about how optics and vision works, which can be inconsistent with accepted science and even people's own declarative knowledge (Croucher, Bertamini, & Hecht, 2002). Such beliefs can vary widely from person to person, and lead to very different and often inaccurate responses to the same problems (Bertamini & Soranzo, 2018; Croucher et al., 2002). Folk theories about how vision works would appear to be better-placed to explain the variety of responses we found in our experiments, presenting a potential alternative account of VPT: VPT is not about representation of others' perception but the selection of an ad-hoc heuristic or naïve theory of how vision could work, the choice of which (and accuracy of which) is influences by our subjective ideas. For example, one reason that participants judged the further line to be longer could be the erroneous calculation that vision 'makes up' for depth by extending what is further away until it is consistent with fact.

However, while naïve optics can account for variation it does not allow us to make predictions about *who* chooses *which* approach to a VPT problem, and *when*. Our results thus suggest avenues for future research to further our understanding of VPT. Firstly, a clearer understanding of how stimuli that vary categorically or continuously influence VPT could help to clarify questions around representation. Research using languages that differ in the number of categories they apply to a continuum, such as colour, is one example of how this question might be approached (e.g., Samuel, Frohnwieser, Lurz, & Clayton, 2020). Secondly, our results make a strong case for the need to better understand individual differences in how VPT problems are tackled. Naïve optics could serve as a loose framework for such investigations. For example, it might be predicted that accuracy on tasks that test optics (e.g. questions about visibility and reflections at different viewing angles) will correlate to some extent with accuracy in VPT tasks with agents. Relatedly, idiosyncrasies in individual experiences are also likely to play a role. This has already been shown in other perspective-taking tasks with broadly-defined variables such as social class (Dietze & Knowles, 2020), gender

(Kessler & Wang, 2012), social skills (Kessler & Wang, 2012), and culture (Wu & Keysar, 2007). We speculate that more finely-grained variables, such as experience with photography and visual arts, might make increase accuracy in perspective-taking tasks like the ones presented here.

In sum, although our results do not rule out perceptual simulation as one way in which some individuals may attempt to solve some VPT problems, it is unlikely to be the principal means by which people do so. Instead, our results point to the potential for a cluster of different processes adopted by different individuals according to context and experience.

Acknowledgements: None

Disclosure of interest: The authors report no conflicts of interest.

References

Bertamini, M., Latto, R., & Spooner, A. (2003). The Venus effect: people's understanding of mirror reflections in paintings. *Perception, 32*(5), 593-599.

Bertamini, M., & Soranzo, A. (2018). Reasoning about visibility in mirrors: A comparison between a human observer and a camera. *Perception, 47*(8), 821-832.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition, 107*(3), 1122-1134.

Cavallo, A., Ansuini, C., Capozzi, F., Tversky, B., & Becchio, C. (2017). When far becomes near: Perspective taking induces social remapping of spatial relations. *Psychological Science, 28*(1), 69-79.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition, 13*(1991), 127-149.

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

Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychonomic bulletin & review*, 1-16.

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

Croucher, C. J., Bertamini, M., & Hecht, H. (2002). Naive optics: Understanding the geometry of mirror reflections. *Journal of Experimental Psychology: Human Perception and Performance, 28*(3), 546.

Deroualle, D., Borel, L., Devèze, A., & Lopez, C. (2015). Changing perspective: The role of vestibular signals. *Neuropsychologia, 79*, 175-185.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in psychology, 5*, 781.

Dietze, P., & Knowles, E. D. (2020). Social Class Predicts Emotion Perception and Perspective-Taking Performance in Adults. *Personality and social psychology bulletin*, 0146167220914116.

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and cognition, 41*, 93-103.

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. *Quarterly Journal of Experimental Psychology, 70*(8), 1646-1660.

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

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences, 103*(2), 489-494.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and language, 105*(2), 91-98.

Holmes, K. J., Moty, K., & Regier, T. (2017). Revisiting the role of language in spatial cognition: Categorical perception of spatial relations in English and Korean speakers. *Psychonomic bulletin & review, 24*(6), 2031-2036.

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.

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

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.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4*(1), 47-60.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience, 2*(9), 635-642.

Kosslyn, S. M., Pinker, S., Smith, G. E., & Shwartz, S. P. (1979). On the demystification of mental imagery. *Behavioral and brain sciences, 2*(4), 535-548.

Linde, C., & Labov, W. (1975). Spatial networks as a site for the study of language and thought. *Language*, 924-939.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & psychophysics, 68*(2), 327-337.

Millett, A. C., D'Souza, A. D., & Cole, G. G. (2019). Attribution of vision and knowledge in 'spontaneous perspective taking'. *Psychological research*, 1-8.

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

Nuku, P., & Bekkering, H. (2008). Joint attention: Inferring what others perceive (and don't perceive). *Consciousness and cognition, 17*(1), 339-349.

Perner, J., & Leekam, S. (2008). The curious incident of the photo that was accused of being false: Issues of domain specificity in development, autism, and brain imaging. *The Quarterly Journal of Experimental Psychology, 61*(1), 76-89.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences, 1*(4), 515-526.

Pylyshyn, Z. (2003). Return of the mental image: are there really pictures in the brain? *Trends in cognitive sciences, 7*(3), 113-118.

Roberson, D., & Hanley, J. R. (2010). An Account of the Relationship between Language and Thought in the Color Domain. *Words and the Mind*, 183.

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition, 107*(2), 752-762.

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.

Samuel, S., Frohnwieser, A., Lurz, R., & Clayton, N. (2020). Reduced egocentric bias when perspective-taking compared to working from rules. *Quarterly Journal of Experimental Psychology*, 1747021820916707.

Samuel, S., Legg, E., Manchester, C., Lurz, R., & Clayton, N. (2019). Where was I? Taking alternative visual perspectives can make us (briefly) misplace our own. *Quarterly journal of experimental psychology (2006)*, 1747021819881097.

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 929.

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

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences, 106*(11), 4567-4570.

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology, 29*(5), 874-880. e874.

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.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, 104*(19), 7780-7785.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science, 18*(7), 600-606.

Yu, A. B., & Zacks, J. M. (2017). Transformations and representations supporting spatial perspective taking. *Spatial Cognition & Computation, 17*(4), 304-337.

Yun, H., & Choi, S. (2018). Spatial semantics, cognition, and their interaction: a comparative study of spatial categorization in English and Korean. *Cognitive Science, 42*(6), 1736-1776.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition, 35*(1), 41-68.