2022-08-22

It's not you, it's me: A review of individual differences in visuo-spatial perspective taking.

Samuel, Steven

10.1177/17456916221094545
Perspectives on Psychological Science
SAGE Publications

It's not you, it's me: A review of individual differences in visuo-spatial perspective taking.

Steven Samuel[1,2]

Geoff G. Cole[2]

Madeline J. Eacott[2]

1 Department of Psychology, University of Plymouth, U.K.

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

ACCEPTED VERSION PERSPECTIVES IN PSYCHOLOGICAL SCIENCE

13/03/2022

Word count: 7653 (excluding abstract, acknowledgements, captions, references and tables, including footnotes)

Key words: perspective-taking, vision, theory of mind

Address for correspondence: Steven Samuel. University of Plymouth, Department of Psychology, Drake Circus, Plymouth, PL4 8AA. Email: steven.samuel@plymouth.ac.uk

Abstract

Visuo-spatial perspective taking (VSPT) concerns the ability to understand something about the visual relationship between an Agent or observation point on the one hand, and a target or scene on the other. Despite its importance to a wide variety of other abilities, from communication to navigation, and decades of research, there is as yet no theory of VSPT. Indeed, the heterogeneity of results from different (and sometimes the same) VSPT tasks point to a complex picture suggestive of multiple VSPT strategies, individual differences in performance, and context-specific factors that together have a bearing on both the efficiency and accuracy of outcomes. In this paper, we review the evidence in search of patterns in the data. We find a number of predictors of VSPT performance but also a number of gaps in our understanding that suggest useful pathways for future research and, possibly, a theory (or theories) of VSPT. Overall, this review makes the case for understanding VSPT by better understanding the perspective taker rather than the target agent or their perception.

Successful visuo-spatial perspective taking (VSPT) concerns the ability of an individual to understand something about the visual relationship between an *Agent* or observation point on the one hand, and a target or scene on the other. For example, a *Perspective Taker* might judge whether someone can see the salt shaker before they ask them to pass it (e.g., it is not hidden behind a menu), gauge what the view of the stage might be before booking theatre seats online, or rapidly assess the blind spots of another driver. VSPT is considered vital to humans' ability to interact and socialise with others (Brown-Schmidt et al., 2008; Clark & Brennan, 1991), and has been the subject of research and theorizing since at least the time of Piaget (Piaget & Inhelder, 1956). Nevertheless, there is currently no formal theory or model of VSPT (Cole & Millett, 2019; Cole et al., 2020). This is possibly due to the evidence for heterogeneity both in terms of *how* people take perspectives and *how well* they do so (Bukowski & Samson, 2017). An important step forwards would thus be to extract any patterns from these individual differences.

Here, we review the evidence for individual differences in strategy selection and performance in explicit (i.e., conscious) VSPT. The article is divided into three sections. In the first we examine whether the means by which a Perspective Taker opts to solve a VSPT problem (i.e. their *strategy choice*) is predictable by factors such as gender, culture, schizotypy, and autistic traits. This section also examines some external influences on strategy choice, namely direct instructions and the nature of the task. In the second section we describe and discuss results of studies showing that Perspective Takers tend to be egocentric, and that this affects not only processing efficiency but also accuracy. In the third section, we summarise the evidence for other individual differences in VSPT performance that occur cumulatively with, or irrespective of, strategy choice and egocentric biases.

Our central argument that follows from this review is that the Perspective Taker shapes the Agent's perspective, not the other way around. This is evidenced by how

Perspective Takers' unique experiences affect the nature and efficiency of the attributions they make to others. This approach emphasises the understanding of individual differences and how these flexibly interact with the variety of VSPT tasks, at the same time that it *de*-emphasises the role of the visual experiences that one is attempting to understand. What we gain from this understanding is the ability to begin to predict both *how* and *how well* people tackle different VSPT problems.

Factors influencing strategy selection

VSPT problems are usually about *visibility* (e.g., is the target object visible to the Agent?) or *appearance* (e.g., does the target object appear the same to the Agent as it does to you?). These problem types correspond to the classic Level 1/Level 2 distinction (Flavell et al., 1981; Masangkay et al., 1974). In experiments that examine Level 1 problems, participants are often given tasks which require them to interpret instructions according to an Agent's more restricted viewpoint. For example, in the Director Task the instruction to select the "top cup" in an array might require the participant to select the middle cup from their own perspective if the real top cup is not in the Agent's field of view (e.g., Apperly et al., 2010; Keysar et al., 2003). Tasks that examine the ability to understand relative appearance, on the other hand, typically require participants to judge what an object looks like to another person when it is mutually visible to both the Perspective Taker and the Agent, such as whether a digit looks like a 6 or a 9 (e.g., Samuel, Cole, et al., 2020; Surtees et al., 2013b; Surtees et al., 2012), how objects which are closer to an Agent loom larger (Samuel, Hagspiel, Eacott, et al., 2021), and how colours appear different to Agents if they view objects through colour filters (Samuel, Frohnwieser, et al., 2020).

It is a characteristic of VSPT that there are multiple and distinct cognitive strategies that an individual has at their disposal in order to solve a problem (Gardner et al., 2013;

Pearson et al., 2013). These strategies include, but are unlikely to be limited to, line-of-sight drawing (Michelon & Zacks, 2006), imagining oneself physically relocated in other spaces (Kessler & Rutherford, 2010), mentally rotating scenes until they align with one's own perspective (Wraga et al., 2000), understanding the spatial relationships between target objects and occluders (Santiesteban et al., 2015), and reversing left/right mappings for Agents opposite (Yu & Zacks, 2017). In essence, by VSPT strategy we therefore mean the procedure that the Perspective Taker goes through to formulate a response. Since the efficiency and accuracy of responses are in part predicated upon this choice, if we wish to understand and make predictions about how VSPT works then we need to understand the factors that lead an individual to select one strategy over another. At the same time, we need to understand what the consequences of different strategies would be for different people.

For relatively simple visibility questions (i.e., Level 1 VSPT) it can suffice to “draw a line” from the Agent’s eyes and conclude a target is seen if the line is uninterrupted. This strategy was elegantly demonstrated by Michelon and Zacks (2006), who found that adults took longer to judge whether a doll “saw” a target the further away the object was from the doll (attributed to the time taken to “draw” this line). Note that for this type of problem it is not important to know what the object looks like to the agent, only that it is visible at all. More complex Level 1 VSPT questions, such as arrays with multiple objects and occlusions, and Level 2 (appearance) problems would require different strategies. For example, in the Director Task the participant is faced with a complex arrangement of multiple objects and barriers to process. It has been argued that under such circumstances participants might prioritise an understanding of the relationship between the objects and the barriers themselves—object-centred spatial coding— rather than attempt to understand what the Agent actually 'sees' (Heyes, 2014; Santiesteban et al., 2015). Note that object-centred spatial coding, like line-of-sight, still obviates the need for any holistic representation of the Agent’s visual experience. By a representation psychologists and philosophers typically mean something in one person’s mind (the Perspective Taker in this case) which “stands for” something in the world (the Agent’s perspective, in this case). Even some Level 2 problems, usually associated with a need to generate precisely such a holistic representation (Lurz, 2009), do not always require more than a skeletal notion of another’s perspective. For example, for left/right judgements from the perspective of an Agent directly opposite (facing), Yu and Zacks (2017) have pointed out that the simple heuristic of reversing left/right mappings could be applied.

These are just some of the strategies involved in VSPT. To illustrate how our knowledge of individual differences in strategy choice is currently insufficient, versions of the Director task in which the director is removed and participants are instead instructed to ignore items in front of occluders (making the task rule-based and non-social) generate both poorer *and* better performance relative to the original, director-led version (Apperly et al., 2010; Dumontheil, Apperly, et al., 2010; Dumontheil, Küster, et al., 2010). A potential explanation for this variability is individual differences in strategy preferences, with some participants performing the director-absent version as if it were a perspective taking task, and others performing the director-present version using object-centred spatial coding. Similarly, we might surmise that a Perspective Taker who knows that a 6 and a 9 look like each other when upside down can apply this knowledge to VSPT problems based solely on the Agent’s location relative to the digit. In contrast, an individual approaching this problem for the first time might select a more cognitively demanding strategy, such as mentally rotating the digit.

There are some areas where our understanding of strategy selection is better. This is the case for the influence of instructions. It is a frequent finding that simply asking a participant to use one particular strategy can produce evidence of that strategy’s use, even if the strategy in question is suboptimal and the participant would have no reason to suspect that non-compliance would be detectable. This is most evident in research assessing embodied VSPT and mental rotation strategies. We discuss this in more detail below as it provides a good test case not only for the influence of external instruction but also for a number of influences upon strategy selection more generally.

Case study: Embodiment vs. Mental Rotation.

Embodied perspective taking and array rotation are two well-known and cognitively distinct strategies for Level 2 VSPT. The former concerns imagining oneself in the location that provides the desired view and then making a judgement from this quasi-egocentric perspective. This process is known as perspective transformation, viewer rotation, or simply ‘embodiment’ (e.g., Kessler & Thomson, 2010; Wraga et al., 2000; Zacks & Tversky, 2005). Evidence for this process comes from impaired performance when participants’ bodies are rotated or restricted such that the shortest path to an imagined location is more difficult to attain (e.g., Kessler & Thomson, 2010; Surtees et al., 2013b; Yu & Zacks, 2017), and from erroneous manual responses consistent with perspectives just imagined rather than one’s actual location (Samuel, Legg, et al., 2019). A striking example comes from a study in which participants performed a visual perspective task in virtual reality while sitting in a chair that could be rotated by the experimenters. Deroualle et al. (2015) found that the participants were faster to adopt an alternative perspective if the chair rotated in the same direction as the movement required to ‘reach’ that new viewpoint. Since participants were wearing a virtual reality headset the feeling of motion alone (vestibular sensations) in the absence of visual feedback was enough to influence performance.

The other strategy is to mentally rotate the target object or scene instead, variably known as array rotation, object rotation, and mental rotation (Wraga et al., 2000; Wraga et al., 2003; Zacks & Tversky, 2005). This is usually considered analogous to the ability to mentally rotate objects more generally (Shepard & Metzler, 1971).

Although both embodied and mental rotation processes should culminate in identical outcomes, the distinction has been demonstrated both neurologically (Schurz et al., 2013; Zacks et al., 2003) and behaviourally. For example, Kessler and Thomson (2010) asked participants to make judgements about other visual perspectives or mentally rotate objects, observing an influence of congruent or incongruent body posture only with the former. In the aforementioned vestibular sensations study by Deroualle et al. (2015) there was no effect of the rotating chair if the task was not to imagine other viewpoints but to mentally rotate objects instead.

The selection of one strategy or another is determined in part by the nature of the VSPT task itself. For example, Zacks and Michelon (2005) reported that adults prefer to mentally rotate objects for same/different judgments but prefer embodiment for left/right judgments. Judgments about *objects* can also elicit a different strategy from judgments about *bodies* (e.g., Amorim et al., 2006; Muto & Nagai, 2020). Strategy selection is also influenced by overt instructions, to the extent that a more sub-optimal approach can be elicited simply by request (e.g., Presson, 1982; Wraga et al., 2000; Zacks & Tversky, 2005).

Instructions and target types are clearly external pressures on VSPT strategy rather than individual differences, but there is also evidence that strategy selection can be predicted by what we know about Perspective Takers. In healthy adults, responses to many VSPT tasks are usually achieved more quickly using embodiment than array rotation (Amorim & Stucchi, 1997; May, 2004; Zacks & Tversky, 2005), suggesting this might be the preferred way in which people solve VSPT tasks. Interestingly, there is evidence that this preference might be stronger in women. Kessler and Wang (2012) found that the effect of manipulating adults' body posture while making left/right judgements from alternative perspectives was stronger for women than for men, an effect the researchers attributed to the tendency for women to be greater "empathisers" and men greater "systemisers" (Baron-Cohen & Wheelwright, 2004), with empathisers attempting a "deeper" embodiment (see also Gronholm et al., 2012).

Another example of both preference and flexibility comes from a study that compared adults with high and low schizotypy. Schizotypy refers to a multidimensional construct that is thought to underlie schizophrenia, and is expressed in the personality of those with and without clinical diagnosis (Kwapil & Barrantes-Vidal, 2015; Wong & Raine, 2020). Schizophrenia is typically associated with atypical theory of mind (ToM), the ability to understand others' unobservable mental states such as their beliefs (Frith, 2004; Harrington et al., 2005; Lee et al., 2004), and perspective-taking (Langdon & Coltheart, 1999). In a VSPT task, Langdon and Coltheart (2001) presented forty adults with four coloured blocks arranged in a square on a stand and asked them questions about the appearance of the array from different angles. The questions were either *item* questions (i.e. "What colour block would be to the FRONT and RIGHT?") or *appearance* questions, which involved judging whether a picture matched what the array should look like after the transformation was imagined. Importantly, participants were instructed to imagine either that the stand that the blocks were on was rotated (array rotation) or that they themselves were "rotated" around the table (viewer rotation/embodiment). The participants were also divided into a "high"- or "low-schizotypy" group based on a median split of responses to a questionnaire designed to measure schizotypal traits. Results showed that both groups performed similarly on item questions, but on appearance questions the high-schizotypal group performed more slowly than the low-schizotypal group in the viewer rotation condition. In addition, the high-schizotypal group alone found viewer rotation more difficult than array rotation, and performed *faster* than the low-schizotypal group on array rotation trials. The researchers explained the performance of the high-schizotypy participants in terms of a reduced ability to simulate and select between multiple ‘first-person’ viewpoints of a fixed reality, compensated for with an enhanced ability to imagine alternative realities relative to the self. That this effect was found in healthy adults is particularly striking (the impairment in the embodied condition has also been found in patients with schizophrenia: Langdon et al., 2001).

Like schizophrenia and schizotypy, autism has also been associated with atypicality or deficits in the ability to reason about others’ mental states (e.g., Baron-Cohen, 1995; Frith, 2001; Leslie & Thaiss, 1992), though recent scholarship and research suggests the matter is more complex than previously thought, and may itself reflect a particular case of individual differences and intersubjective perspective taking (e.g., Brewer et al., 2016; Milton, 2012; Williams, 2021). Here, we concern ourselves with evidence for differences in VSPT strategies specifically. In a Level 2 VSPT study comparing autistic children and typically developing controls, Pearson et al. (2016) found that both groups were as good at judging what a three dimensional figure looked like from different angles when they were asked to imagine themselves in a different location around the figure as when they were asked to imagine what someone else would see from similar locations. However, they also found that the autistic children's performance was related to a mental rotation task, but the typically-developing controls’ performance was related to a body matching task, suggesting that the two groups nevertheless adopted different strategies (see also Hamilton et al., 2009; Kessler et al., 2014; Kessler & Wang, 2012).

The case of embodiment vs. mental rotation demonstrates the roles of both internal and external factors in VSPT strategy selection: individuals tend to come to a VSPT problem with a preference for a particular strategy, a preference that is partly predictable by individual differences such as gender, culture, and clinical or sub-clinical traits. Nevertheless, these preferences are readily undermined by formal instructions. These findings favour a flexible, plural, and context-specific approach to VSPT which we can begin to understand better by learning more about the Perspective Takers at their centre.

Folk optics as wildcard strategies.

Embodiment and rotation are not the only means by which people choose to tackle VSPT problems. Since VSPT problems are usually accompanied by an instruction or desire to understand aspects of an Agent's visual experience, some apply their understanding of vision to a problem. However, there is evidence that many people hold quite erroneous theories about how vision really works. For example, in the Venus Effect an observer sees an Agent looking into a mirror and also sees the agent's reflection. Observers tend to believe that the Agent sees their reflection in the mirror as they do, despite the Agent and mirror not being along the participant's line of sight (Bertamini et al., 2003; Bertamini & Soranzo, 2018). This effect occurs in approximately 75% of adults (Bertamini et al., 2003), and it is an illusion exploited by art and film to avoid revealing the camera or artist in scenes where performers look into mirrors. Such effects have been attributed to *naïve* or *folk optics*, namely folk beliefs about how vision works. These beliefs vary from person to person and can be inconsistent with accepted science and even people's own declarative knowledge (Croucher et al., 2002; Samuel, Hagspiel, Cole, et al., 2021). Different theories and heuristics related to visual reasoning, such as imagining top-down geometric viewpoints or applying past experience of moving through similar scenes, lead to different responses to the same problems depending on whom is asked (Bertamini & Soranzo, 2018; Croucher et al., 2002).

When applied to VSPT, folk optics could lead to unexpected and inaccurate responses. However, to date there is little research linking folk optics to VSPT. In a recent study by Samuel, Hagspiel, Eacott, et al. (2021), adults were presented with an image of an Agent looking at two lines on a wall (Figure 1A) and were asked to indicate how long each line appeared to that person. The lines were identical but the closer line to the Agent appeared *visually* longer, as the photo taken from the Agent's location in Figure 1C shows. It was made clear to participants that the Agent knew that the two lines were of the same length in reality and that the aim was to judge *visual appearance*. Results showed no evidence that participants could successfully take the Agent's perspective of the lines. This result persisted when the Agent was replaced by a camera and participants asked how long the lines would appear in a photo, eliminating the possibility that they were correcting for the Agent's knowledge that the lines were identical (Figure 1B). The effect also persisted when the disparity with the participant's view was made more salient, and the lines were turned 90 degrees to ensure that participants' difficulty was not based on problems extending length into depth (Figures 1D and 1E). However, participants had little trouble judging the closer line to appear longer when they saw the aforementioned photo (Figure 1C). Surprisingly, the data from these experiments showed that participants were not simply judging the lines to appear the same length; about as many participants erroneously judged the *further* line to appear longer (those data points below the zero midline in the scatterplots to the right of each image). One explanation could be that these participants applied folk optics; in this case, the erroneous belief that visual processing compensates for stretch into depth by enlarging more distant objects. Different beliefs about how vision works helps explain why presenting multiple participants with the same stimuli will generate a range of responses, even *opposite* responses, that cannot be attributed to or predicted by the Agent's real perspective. Until we know more about naïve optics and how these relate to VSPT it will be difficult to predict the outcome of VSPT tasks which rely heavily on them.

[INSERT FIGURE 1 ABOUT HERE]

Figure 1. Example stimuli from Samuel, Hagspiel, Eacott, et al. (2021) along with spread of results. Participants responded by judging the length of each line on a slider. Here each dot in the scatterplot represents the judgement of a single participant, with points above the zero mark indicating the closer line was judged longer (closer line judgement minus further line judgement). Note the spread of responses suggesting approximately as many participants erroneously judged the further *line to appear longer except when they saw a photo taken from the Agent's location (C).*

Summary and predictions

We have seen how there are various means by which VSPT tasks can be approached. Strategy selection will depend in part upon the nature of the problem; if it is a visibility judgement then the chances that a line of sight process will be selected are increased, but if a left/right judgement is required then an embodied process or reversal of spatial mappings should be favoured. Individual differences also predict VSPT strategy. Expertise should influence strategy selection because the Perspective Taker should feel more confident using a more practised strategy. For example, a tour guide who is used to reversing left/right mapping for listeners should be more likely to do adopt this heuristic for left/right judgments than people without such experience, who might instead go the "longer" way around by using embodiment or mental rotation. We have seen the evidence that women appear biased *towards* embodiment, but that autism and schizophrenia (and their non-clinical counterparts) tend to bias people *away* from it. We have seen how strategy selection can influence the accuracy of outcomes negatively if an individual is encouraged to use a personally suboptimal strategy. Folk optics are also strategies, and represent a "wildcard" influence on strategy selection, as they might lead an individual to bypass entirely the “classic” strategies of embodiment and rotation. They can also lead to wildly inaccurate responses. Figure 1 summarises these predictions.

[INSERT FIGURE 2 ABOUT HERE]

Figure 2. Suggested preferred strategy selection pathways. SZ = Schizophrenia.

It’s not you, it’s me: Egocentric bias

Egocentric bias is the tendency for one’s own privileged knowledge to interfere when attempting to be objective about other people’s experiences (e.g., Birch & Bloom, 2004; Keysar et al., 2000; Ross et al., 1977). For example, people overextend to others their opinions (Mullen et al., 1985; Ross et al., 1977), their personality and behaviour (Dunning & Hayes, 1996; Van Boven et al., 2005), and their valuations of objects (Van Boven et al., 2000). People even show egocentric bias when judging others’ sensations of pleasantness or unpleasantness (Silani et al., 2013) and drive states like thirst and discomfort (Nordgren et al., 2011; Van Boven & Loewenstein, 2003). As a minimum, egocentric biases in perspective taking demonstrate the difficulty that Perspective Takers have in making truly objective judgments about other Agents’ experiences.

Egocentric bias has also been demonstrated in VSPT. For example, in the Director Task described earlier, participants are required to select objects from an array according to instructions from a director whose view is limited by occluders. Participants are often slower to select an object from an Agent’s perspective if there is a competitor object which matches the instruction but nevertheless only they can see (e.g., Epley et al., 2004; Keysar et al., 2003; Wu & Keysar, 2007). In some versions of the Dot Perspective task, in which participants are instructed to verify how many dots an avatar sees, adults are slower on trials where the avatar sees a different number from themselves (e.g., Qureshi et al., 2010; Samson et al., 2010; Santiesteban et al., 2014).

The continuing relevance of the self-perspective is also evident in tasks with more complex stimuli. Surtees et al. (2012) found that children and adults were slower to verify which number an Agent saw if their own perspective made it look different, such as an upside-down 6 appearing to be a 9, than if the number was identical regardless of perspective, such as an ‘8’ (see also Surtees et al., 2016). A recent study from our own lab (Samuel, Cole, et al., 2020) found not one but two different types of interference from the self-perspective. Participants needed to locate a target digit (a ‘6’ or a ‘4’) in a grid from the perspective of an avatar. The target was always upright from the avatar’s perspective and the entire scene was always presented in top-down view, such that if the avatar was at the top of the grid the digits would be upside-down for the participant. We found the classic ‘pull’ of an egocentrically-correct distractor; participants were slower to select a ‘6’ from the avatar’s perspective if there was an alternative ‘6’ in the array from their own viewpoint (e.g., Keysar et al., 2003). In addition, we also found that responses were slower when participants had to select an upside-down 6 than an upside-down 4. This suggested that participants were reluctant to select a target that is not merely an ambiguous shape but *another number*. Participants’ knowledge thus influences performance in not one but two ways; the ability to ignore what is egocentrically correct, and the ability to select a target with an alternative identity to the instruction. Critically, each of these factors concern things only the Perspective Taker sees, yet they impact responses that are meant to reflect the perspective of another Agent.

The results described above come from response time paradigms, which index processing efficiency but tell us little when it comes to thinking about how we actually *represent* other perspectives (Cole & Millett, 2019). However, Perspective Takers' egocentric bias also seems to extend to their understanding of *what* the Agent sees. For example, in the Director task participants not only experience interference from objects only they can see but also select them (Apperly et al., 2010; Epley et al., 2004; Keysar et al., 2003), and even when under no time pressure (Legg et al., 2017). This seems to suggest that they attribute their own knowledge to others. However, there is an alternative explanation. Perspective Takers may fail on such tasks not because they imagine the Agent sees what they see, but because they fail to engage in perspective taking at all (effectively performing the task as if no other agent existed). The case for misattribution would be stronger if we could be more certain that the Perspective Taker was trying to pay attention to the Agent's perspective when responding. This is the case in a series of studies by Wardlow (2013) and others, in which participants gave rather than received the instructions. She found that descriptions of a target were often qualified with a contrasting adjective like "small" or "large" when the participant could see a different-sized competitor but the listener could only see one item matching the description (see also Damen et al., 2019; Wardlow Lane & Ferreira, 2008; Wardlow Lane & Liersch, 2012). The important point about these results is that Perspective Takers are less likely to be ignoring the Agent when they are describing objects *for* them, and when they are a real person directly in front of them rather than a computer avatar. These errors are thus more likely to be perspective misattributions (i.e., believing that the other agent sees what we see when this is untrue) rather than neglecting to consider the agent's point of view at all. Just as response times in VSPT tasks can be intentionally 'dialled up' by presenting Perspective Takers with egocentric distractors and misleading targets (Samuel, Cole, et al., 2020), the tendency to egocentrically misrepresent other perspectives can be enhanced by presenting conflicts such as contrasting pairs that are visible only to the Perspective Taker (Wardlow, 2013).

In sum, egocentric bias is a *degrader* both of efficiency and objectivity in VSPT. Egocentric bias also makes sense, because Perspective Takers attribute perspectives and cannot ‘take’ them – it follows that these attributions will be coloured by the one’s own knowledge and perceptions. We can therefore hypothesise that a Perspective Taker’s understanding of other visual perspectives is predictable by the knowledge and perceptions of the Perspective Taker; the greater the correspondence between these and the perspective of the Agent, the more accurate their perspective estimates are likely to be. It also follows that a better-informed Perspective Taker should produce more accurate responses than a less well-informed Perspective Taker. For example, someone knowledgeable in colour mixing should be better able to judge that a yellow object perceived by an Agent through a blue filter will appear green than someone without this knowledge.

Individual differences in performance

We have seen how strategy selection can have an impact on VSPT performance, and how egocentric bias can degrade it. There is also evidence for individual differences in VSPT accuracy more broadly. However, our understanding of these variables is currently limited, and what we do know often comes from developmental studies and belief reasoning (ToM) rather than adults or VSPT. Nevertheless, what evidence there is converges on the distinct likelihood that such factors also influence VSPT performance.

Culture

Evidence for cultural variation in ToM performance comes largely from developmental studies. This work suggests that the relationship is a complex one concerning not only accuracy but the stages at which different components of ToM emerge (e.g., Liu et al., 2008; Wellman, 2017; Wellman et al., 2001). Influences of culture demonstrate flexibility and cognitive penetrability and are problematic for nativistic theories of social cognition (Lillard, 1998; Wellman, 2018).

It is unclear how much evidence from ToM/belief reasoning can tell us about VSPT specifically. We are aware of only three studies that explicitly relate culture to VSPT. Wu and Keysar reported that Chinese nationals were more sensitive to other agents' perspectives than Americans on the Director Task, both in terms of their fixation patterns and their behavioural responses. They attributed this finding to the Chinese participants' interdependent, collectivistic culture which promotes consideration of non-egocentric views (see also Wu et al., 2013). In other work assessing both visibility judgements and left/right judgements from an avatar's perspective, East Asian (Chinese) adults were again better (i.e., faster) than their Western counterparts (Kessler et al., 2014). Like Wu and colleagues, the authors speculated that the reason for this was the more other-centred culture of the East Asian group. However, using a computerised Director Task, Wang et al. (2019) found no meaningful differences between a group of Taiwanese (interdependent) and British (independent) adults. There are a number of methodological differences between these studies that might explain this discrepancy, but only with further research on culture in adult VSPT is a clearer picture likely to emerge.

Bilingualism

An alternative or perhaps complementary explanation for the better performance of the Asian samples in the studies by Wu and Keysar (2007) and Kessler et al. (2014) is that all or part of the sample were recruited from English-language universities and therefore bilingual. Bilingual children have sometimes been found to outperform monolinguals on ToM tasks (Kovács, 2009), including VSPT (for reviews see Rubio-Fernández, 2017; Schroeder, 2018). For example, Goetz (2003) found that English-Mandarin bilingual preschoolers were better able to judge whether a picture of a turtle showed it on its feet or on its back from the perspective of someone sitting opposite them than their monolingual peers at one time of testing, though a week later the monolinguals had already caught up. Evidence from adult populations is both more limited and more equivocal. Better performance by bilinguals has been reported on a false belief task (Rubio-Fernández & Glucksberg, 2012), but not in a referential communication task (Ryskin et al., 2014) or Director Task (Samuel et al., 2016). More research is clearly needed, and also an understanding of why bilingualism may offer a potential advantage in VSPT. One explanation is that the management of two languages enhances Executive Function (e.g., Bialystok, 2017), but this claim has proven controversial and a number of large-scale studies, reviews, and meta-analyses have failed to find supporting evidence (e.g., De Bruin et al., 2015; Paap et al., 2015; Samuel et al., 2018). An alternative account is that bilingualism teaches an individual that concepts and their labels are not inextricably linked (i.e., greater metalinguistic awareness; see Leopold, 1971; Vygotsky, 1962), and by extension views of the world are more person-specific that might be assumed by monolinguals (see Doherty & Perner, 2020 for a discussion of metalinguistic awareness and its relation to perspective taking; see also Rubio-Fernández, 2017). The two accounts differ in emphasis, with the former explaining enhanced performance through practice with domain-general processes but the latter enhanced metarepresentation and an appreciation of the subjectivity of experiences. Overall, it appears that an influence of bilingualism on VSPT performance is a possibility that requires further research before firmer conclusions can be drawn.

Social Class

The evidence for the potential for socioeconomic status to influence VSPT performance is preliminary. Dietze and Knowles (2020) recently reported that adults who self-reported belonging to a "lower" category on a social class scale (ranging from “poor, working class, middle class, upper-middle class, or upper class”) performed better on a Director Task than those who judged themselves higher, a finding they speculate could be due to individuals lower on the scale believing that others are more relevant to their experience.

Executive Function (EF)

With the exception of spontaneous perspective-taking (e.g., Qureshi et al., 2010; Samson et al., 2010; Ward et al., 2019), most scholars concur that VSPT is an effortful process (e.g., Apperly & Butterfill, 2009; Apperly et al., 2010; Lin et al., 2010). Executive Functions are domain-general abilities such as inhibition which regulate performance on a broad range of tasks (e.g., Miyake & Friedman, 2012). A substantial body of research in ToM, particularly but not exclusively in development, has demonstrated that more efficient inhibitory control correlates with better performance on perspective-taking tasks (e.g., Brown-Schmidt, 2009; Carlson et al., 2015; Sabbagh et al., 2006; Samson et al., 2005). These results are usually attributed to the ability to successfully suppress pre-potent and perspective-inappropriate (i.e., egocentric) information when considering others’ points of view (Brown-Schmidt, 2009; Carlson et al., 2004). Inhibition has also been related to the ability to flexibly switch between one’s own beliefs and another’s (Bradford et al., 2015).

Similar relationships have often been found between EF and VSPT specifically. For example, in tasks in which adults describe objects for another person in the presence of a competitor in privileged ground, a greater use of redundant adjectives is associated with lower working memory capacity (Wardlow, 2013) and inhibitory control (Long et al., 2018; Wardlow, 2013). In a Director Task, Lin et al. (2010) found that adults with better working memory spent less time between first fixating and then selecting targets and less time considering a competitor in privileged ground. They also found these participants made fewer errors overall. In a further variant of the Director Task, Samuel, Roehr-Brackin, et al. (2019) found that adults were faster on trials from an Agent who shared their perspective (self-perspective trials) than trials from an Agent who did not (other-perspective trials), but that this egocentric advantage disappeared on self-perspective trials that immediately followed an other-perspective trial. The researchers attributed this result to the likelihood that participants applied inhibition to the egocentric perspective in order to perform other-perspective trials, and that this inhibition had not decayed sufficiently prior to the following self-perspective trial (see also Chiu et al., 2019). In a similar task, children's ability to ignore competitors in privileged ground was related to a measure of their inhibitory control (Nilsen & Graham, 2009).

The influence of EF on VSPT performance allows us to make predictions about participant *x*'s performance on VSPT task *y* based on their EF performance (*z*). It also demonstrates again the utility of exploring VSPT beginning from the Perspective Taker. However, EF is a complex and multifaceted suite of abilities, and how each component might relate to VSPT (and indeed different VSPT problems) is not well understood. For example, sometimes individual differences in EF do *not* translate into performance on VSPT, or sometimes do so only in limited ways. In a Level 2 VSPT task Samuel, Cole, et al. (2020) found no evidence that adults' performance on a test of inhibition (the Simon Task: Simon & Rudell, 1967) correlated with either the ability to ignore egocentrically correct competitors or the ability to select a visually misleading but nevertheless correct target. Qureshi et al. (2019) found that adults' performance on two different VSPT tasks was predicted by *different* measures of EF, and moreover that performance on both tasks did not correlate with each other, though this discrepancy might have been due to strong methodological differences such as the presence of both self- and other-perspective trials in one task (Dot Perspective

Task) but only other-perspective trials in the other (Director task). Overall, the balance of the evidence suggests some role for individual differences in EF as predictors of VSPT performance, but precisely how and how much remains to be learned.

Summary and predictions

Overall, the evidence for individual differences in VSPT performance is preliminary but suggests a number of potential relationships. There is some support for enhanced performance as a function of a more other-centred rather than individualistic culture, as well as bilingualism relative to monolingualism, though in some cases these different predictors may have been conflated. To date, a single study has suggested that lower social class predicts better performance on a Director task. Research assessing a potential relationship between EF and VSPT is more plentiful but continues to paint a complex and sometimes erratic picture. An interesting question for any of these variables is precisely how and at what stage in processing they might influence VSPT performance. For instance, would coming from a more other-centred culture mean that the balance of "self" and "other" perspective information is already tilted in favour of the latter, or does it mean that egocentricity is overcome more easily but at a later stage, just prior to response?

Conclusions and limitations

This review makes the case for understanding VSPT by better understanding the Perspective Taker, rather than the Agent or their perception. There is an analogy here with theorizing that has occurred in a quite different field. In his book on language evolution, *The Talking Ape: How Language Evolved*, Robbins Burling (2005) pointed out that we ask people if they can *speak* a language, not if they can *understand* it. He argued that this emphasises the wrong end of the equation; any innovation in language production would have died with its creator if the ability to comprehend the new language had not evolved earlier. Language evolution is thus the story of the evolution of comprehension. In the same sense, perspective taking is an ability whose scope and limitations are set not by the Agent but by the Perspective Taker.

The influence of individual differences in VSPT appears to run counter to recent claims that privilege the Agent's perspective over the Perspective Taker's (e.g., Quesque et al., 2018), for individual differences should have little impact if this were the case. In recent years evidence has emerged that people might take other's visual perspectives spontaneously, i.e., without conscious effort or awareness. Initially, it was speculated that only a limited amount of information could be gleaned in this manner, such as what an Agent sees, but not necessarily how that appeared (Samson et al., 2010). More recently it has been argued that spontaneous VSPT extends to level 2 VSPT and can also provide the Perspective Taker with information complementary to their own perspective, such that objects can be recognised more efficiently if others are better placed to identify them (Ward et al., 2020). This review is concerned with explicit VSPT, and thus it does not speak directly to the debate around spontaneous VSPT (e.g., Apperly & Butterfill, 2009; Cole & Millett, 2019; Santiesteban et al., 2014; Ward et al., 2019). However, we take it as uncontroversial that in *any* form of VSPT a Perspective Taker attributes a perspective to an Agent, and not the other way around (Cole & Millett, 2019; Cole et al., 2020). As we have seen throughout this review, judgments about Agents' perspectives can vary widely among Perspective Takers even when the Agent's perspective is held constant throughout. Thus, the Agent's perspective is unlikely to be a major factor in VSPT except as a yardstick by which to measure the accuracy of Perspective Takers' attributions.

The plurality of VSPT strategies also rules out, to our minds, the possibility that people can represent other visual perspectives veridically, i.e. in a quasi-perceptual or image-like form. This is the claim made by Ward and colleagues (Ward et al., 2019; Ward et al., 2020), who also argue that such representations are generated spontaneously. If this claim was correct then there would need to be an explanation as to why people do not use these presumably highly accurate representations all the time rather than these explicit and sometimes suboptimal strategies, and why people are often inaccurate in their perspective attributions (Samuel, Hagspiel, Eacott, et al., 2021).

There is often a meaningful distinction between the way in which we solve VSPT problems and how we deal with non-perceptual questions about others' beliefs, knowledge, emotions, and so on. For example, people cannot usually make judgements about others' emotions using line of sight, though what someone is attending to might be helpful in order to infer emotional states (and others' emotional states might boost attention to others' lines of sight, such as in fearful faces; e.g., Tipples, 2006). For this reason we do not extend our conclusions to ToM. Of course, this is despite the fact that much research in ToM actually employs tasks which pivot on other agents' visual perspectives but only in the limited sense that 'what an agent sees is what an agent knows about'. For example, the classic false-belief task and its variants rely on agents not seeing the target object move, rendering their belief false (e.g., Keysar et al., 2003; Krupenye et al., 2016; Wimmer & Perner, 1983). Even assessments of Level 2 problems typically pivot on manipulating knowledge and how this then interacts with appearance judgements, such as knowing that what looks like a rock is in fact a sponge (Flavell et al., 1981; Masangkay et al., 1974). Most strategies involved in VSPT do not relate to such questions, or at least not directly. However, one exception could be embodied VSPT, where there is evidence that it can influence the Perspective Taker's sensitivity also to non-perceptual perspectives. Work by Erle and colleagues has found that even making simple left/right judgements from another agent's visual perspective can enhance feelings of similarity with that person, boost sensitivity to their mental states (Erle & Topolinski, 2017), and even facilitate liking of and trust in that individual (Erle et al., 2018).

This would be consistent with the speculation by some scholars that VSPT may have been an evolutionary pre-requisite for other forms of psychological perspective taking (Kessler & Thomson, 2010).

Throughout this paper we have talked about VSPT rather than only visual perspective taking (VPT) or spatial perspective taking (SPT). This is in large part because we feel a broader definition is presently better-placed to capture information that could be crucial, at least until such time that any differences become clear. In VSPT research, 'visual' has traditionally been used to describe tasks in which participants are believed to process what another agent perceives, and 'spatial or visuospatial' used to describe tasks in which judgments are made about where objects are in relation to another agent (Erle & Topolinski, 2017; Kessler & Thomson, 2010; Surtees et al., 2013a). However, the line is frequently blurred. For example, Kessler and Thomson (2010) use the term 'spatial' perspective taking (SPT) but nevertheless argue explicitly for a role of perception, such that "SPT essentially comprises an emulation of the sensory consequences (visual and proprioceptive) of a mental rotation of the self" (p.84). It has also been shown that visual and spatial perspective tasks rely on overlapping mechanisms (Surtees et al., 2013a). In our view, issues around terminology and definitions are understandable given of the absence of a theory of visual perspective taking, and in particular a theory concerning how people *represent* others' perspectives, if that is what occurs (Cole & Millett, 2019; Cole et al., 2020; Samuel, Hagspiel, Eacott, et al., 2021).

The question of representation is also important for our understanding of what constitutes success in VSPT, and by extension what to measure individual differences *against*. Traditionally, success is measured in terms of consistency with the Agent's viewpoint. However, it is sometimes possible to achieve success on VSPT tasks, even Level 2 tasks, without considering perspective at all. For example, we recently conducted a study in which participants took the perspective of an agent opposite them, such that when the participant saw a 6 they should respond that the agent saw a 9 (and vice-versa). However, this led some participants (12-21%) to the error of concluding that, when they saw 69, the agent must have seen 96. This likely arose from a strategy of "number flipping", whereby the numbers 6 and 9 were inverted and the result of this taken to equate to the agent's perspective (Samuel et al., in press). This approach, which we term stimulus-centered rather than agent-centered, could also have led to the *correct* responses on the single-digit trials.

There are many other questions for which there are presently no firm answers, and this review sheds light on some of these factors but also on the gaps where knowledge is insufficient. Our review has not been exhaustive; we have not covered the emerging evidence for individual differences in VSPT performance as a function of arguably more temporary psychological states, such as sub-clinical depression (Erle et al., 2019), and we have focussed primarily on a group of the most common VSPT tasks, such as the Director Task and Dot Perspective Task. One question that we have not tackled here is whether individuals are aware (or can be made aware) of the strategy they adopt; for example, we know that external factors such as instructions influence strategy selection, which suggests that participants have some conscious control over how they approach VSPT problems, but might some individuals have more or less control over selection, and if so, what might be the explanation? Most pressingly in our view is the need to understand better who selects which VSPT strategies, and why. Currently, our information is limited to a scattering of studies contrasting vast demographic variables such as "Asian" versus "Western" culture. Overall, the research reviewed here points to VSPT as a flexible and context-specific suite of abilities rather than a one-size-fits all process, or an innate or modular system.

References

Amorim, M.-A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations:" Body analogy" for the mental rotation of objects. *Journal of Experimental Psychology: General, 135*(3), 327-347.

Amorim, M.-A., & Stucchi, N. (1997). Viewer-and object-centered mental explorations of an imagined environment are not equivalent. *Cognitive Brain Research, 5*(3), 229-239.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological review, 116*(4), 953.

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology, 63*(6), 1201-1217.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT press.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders, 34*(2), 163-175.

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.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological bulletin*, *143*(3), 233-262.

Birch, S. A., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in cognitive sciences*, *8*(6), 255-260.

Bradford, E. E., Jentzsch, I., & Gomez, J.-C. (2015). From self to social cognition: Theory of mind mechanisms and their relation to executive functioning. *Cognition, 138*, 21-34.

Brewer, R., Biotti, F., Catmur, C., Press, C., Happé, F., Cook, R., & Bird, G. (2016). Can neurotypical individuals read autistic facial expressions? Atypical production of emotional facial expressions in autism spectrum disorders. *Autism Research, 9*(2), 262-271.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic bulletin & review, 16*(5), 893-900.

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.

Bukowski, H., & Samson, D. (2017). New insights into the inter-individual variability in perspective taking. *Vision, 1*(1), 8.

Burling, R. (2005). *The talking ape: How language evolved*. Oxford University Press.

Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development, 16*(1), 186-197.

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of experimental child psychology, 87*(4), 299-319.

Chiu, C.-D., Ng, H. C., Kwok, W. K., & Tollenaar, M. S. (2019). Feeling Empathically Toward Other People and the Self: The Role of Perspective Shifting in Emotion Sharing and Self-Reassurance. *Clinical Psychological Science*, 2167702619863058.

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

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, 4*(2), 30.

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.

Damen, D., van der Wijst, P., van Amelsvoort, M., & Krahmer, E. (2019). Perspective-Taking in Referential Communication: Does Stimulated Attention to Addressees' Perspective Influence Speakers' Reference Production? *Journal of Psycholinguistic Research, 48*(2), 257-288.

De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science, 26*(1), 99-107.

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

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

Doherty, M., & Perner, J. (2020). Mental files: Developmental integration of dual naming and theory of mind. *Developmental Review, 56*, 100909.

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.

Dumontheil, I., Küster, O., Apperly, I. A., & Blakemore, S.-J. (2010). Taking perspective into account in a communicative task. *Neuroimage, 52*(4), 1574-1583.

Dunning, D., & Hayes, A. F. (1996). Evidence for egocentric comparison in social judgment. *Journal of personality and social psychology, 71*(2), 213-229.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of experimental social psychology, 40*(6), 760-768.

Erle, T. M., Barth, N., & Topolinski, S. (2019). Egocentrism in sub-clinical depression. *Cognition and Emotion, 33*(6), 1239-1248.

Erle, T. M., Ruessmann, J. K., & Topolinski, S. (2018). The effects of visuo-spatial perspective-taking on trust. *Journal of experimental social psychology, 79*, 34-41.

Erle, T. M., & Topolinski, S. (2017). The grounded nature of psychological perspective-taking. *Journal of personality and social psychology, 112*(5), 683-695.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology, 17*(1), 99-103.

Frith, C. D. (2004). Schizophrenia and theory of mind. *Psychological medicine, 34*(3), 385-389.

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron, 32*(6), 969-979.

Gardner, M., Brazier, M., Edmonds, C. J., & Gronholm, P. (2013). Strategy modulates spatial perspective-taking: evidence for dissociable disembodied and embodied routes. *Frontiers in human neuroscience, 7*, 457.

Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition, 6*(1), 1-15.

Gronholm, P. C., Flynn, M., Edmonds, C. J., & Gardner, M. R. (2012). Empathic and non-empathic routes to visuospatial perspective-taking. *Consciousness and cognition, 21*(1), 494-500.

Hamilton, A. F. d. C., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition, 113*(1), 37-44.

Harrington, L., Siegert, R., & McClure, J. (2005). Theory of mind in schizophrenia: a critical review. *Cognitive neuropsychiatry, 10*(4), 249-286.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science, 9*(2), 131-143.

Kessler, K., Cao, L., O'Shea, K. J., & Wang, H. (2014). A cross-culture, cross-gender comparison of perspective taking mechanisms. *Proceedings of the Royal Society B: Biological Sciences, 281*(1785), 20140388.

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.

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

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*(1), 25-41.

Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental science, 12*(1), 48-54.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science, 354*(6308), 110-114.

Kwapil, T. R., & Barrantes-Vidal, N. (2015). Schizotypy: looking back and moving forward. *Schizophrenia bulletin, 41*(suppl_2), S366-S373.

Langdon, R., & Coltheart, M. (1999). Mentalising, schizotypy, and schizophrenia. *Cognition, 71*(1), 43-71.

Langdon, R., & Coltheart, M. (2001). Visual perspective-taking and schizotypy: evidence for a simulation-based account of mentalizing in normal adults. *Cognition, 82*(1), 1-26.

Langdon, R., Coltheart, M., Ward, P. B., & Catts, S. V. (2001). Visual and cognitive perspective-taking impairments in schizophrenia: A failure of allocentric simulation? *Cognitive neuropsychiatry, 6*(4), 241-269.

Lee, K.-H., Farrow, T., Spence, S., & Woodruff, P. (2004). Social cognition, brain networks and schizophrenia. *Psychological medicine, 34*(3), 391-400.

Legg, E. W., Olivier, L., Samuel, S., Lurz, R., & Clayton, N. S. (2017). Error rate on the director's task is influenced by the need to take another's perspective but not the type of perspective. *Royal Society open science, 4*(8), 170284.

Leopold, W. (1971). The Study of Child Language and Infant Bilingualism. In A. Bar-Adon & W. Leopold (Eds.), *Child Language. A Book of Reading* (pp. 1-13). Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*(3), 225-251.

Lillard, A. (1998). Ethnopsychologies: cultural variations in theories of mind. *Psychological bulletin, 123*(1), 3.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of experimental social psychology, 46*(3), 551-556.

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: a meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology, 44*(2), 523.

Long, M. R., Horton, W. S., Rohde, H., & Sorace, A. (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition, 170*, 25-30.

Lurz, R. (2009). If chimpanzees are mindreaders, could behavioral science tell? Toward a solution of the logical problem. *Philosophical Psychology, 22*(3), 305-328.

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development*, 357-366.

May, M. (2004). Imaginal perspective switches in remembered environments: Transformation versus interference accounts. *Cognitive psychology, 48*(2), 163-206.

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

Milton, D. E. (2012). On the ontological status of autism: the 'double empathy problem'. *Disability & Society, 27*(6), 883-887.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8-14.

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of experimental social psychology, 21*(3), 262-283.

Muto, H., & Nagai, M. (2020). Mental rotation of cubes with a snake face: The role of the human-body analogy revisited. *Visual Cognition, 28*(2), 106-111.

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive psychology, 58*(2), 220-249.

Nordgren, L. F., McDonnell, M.-H. M., & Loewenstein, G. (2011). What constitutes torture? Psychological impediments to an objective evaluation of enhanced interrogation tactics. *Psychological Science, 22*(5), 689-694.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex, 69*, 265-278.

Pearson, A., Marsh, L., Ropar, D., & Hamilton, A. (2016). Cognitive Mechanisms underlying visual perspective taking in typical and ASC children. *Autism Research*, *9*(1), 121-130.

Pearson, A., Ropar, D., & Hamilton, A. F. d. (2013). A review of visual perspective taking in autism spectrum disorder. *Frontiers in human neuroscience, 7*, 652.

Piaget, J., & Inhelder, B. (1956). *The child's concept of space*. Routledge & Paul.

Presson, C. C. (1982). Strategies in spatial reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(3), 243-251.

Quesque, F., Chabanat, E., & Rossetti, Y. (2018). Taking the point of view of the blind: Spontaneous level-2 perspective-taking in irrelevant conditions. *Journal of experimental social psychology*, *79*, 356-364.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230-236.

Qureshi, A. W., Monk, R. L., Samson, D., & Apperly, I. A. (2019). Does interference between self and other perspectives in theory of mind tasks reflect a common underlying process? Evidence from individual differences in theory of mind and inhibitory control. *Psychonomic bulletin & review*, 1-13.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of experimental social psychology*, *13*(3), 279-301.

Rubio-Fernández, P. (2017). Why are bilinguals better than monolinguals at false-belief tasks? *Psychonomic bulletin & review*, *24*(3), 987-998.

Rubio-Fernández, P., & Glucksberg, S. (2012). Reasoning about other people's beliefs: Bilinguals have an advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 211.

Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T. (2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. *Journal of Memory and Language, 74*, 46-76.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and US preschoolers. *Psychological Science, 17*(1), 74-81.

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.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain, 128*(5), 1102-1111.

Samuel, S., Cole, G. G., & Eacott, M. J. (2020). Two independent sources of difficulty in perspective-taking/theory of mind tasks. *Psychonomic bulletin & review*, 1-7.

Samuel, S., Eacott, M., & Cole, G. G. (in press). Visual perspective taking without visual perspective taking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/https://doi.org/10.1037/xlm0001121

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*, Article 104607.

Samuel, S., Hagspiel, K., Cole, G. G., & Eacott, M. J. (2021). 'Seeing'proximal representations: Testing attitudes to the relationship between vision and images. *PloS one, 16*(8), e0256658.

Samuel, S., Hagspiel, K., Eacott, M. J., & Cole, G. G. (2021). Visual perspective-taking and image-like representations: We don't see it. *Cognition, 210,* Article 104607.

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.

Samuel, S., Roehr-Brackin, K., Jelbert, S., & Clayton, N. S. (2019). Flexible egocentricity: Asymmetric switch costs on a perspective-taking task. *Journal of experimental psychology. Learning, memory, and cognition, 42*(2), 213-218.

Samuel, S., Roehr-Brackin, K., & Roberson, D. (2016). 'She says, he says': Does the sex of an instructor interact with the grammatical gender of targets in a perspective-taking task? *International Journal of Bilingualism, 20*(1), 40-61.

Samuel, S., Roehr-Brackin, K., Pak, H., & Kim, H. (2018). Cultural Effects Rather Than a Bilingual Advantage in Cognition: A Review and an Empirical Study. *Cognitive Science, 42*(7), 2313-2341.

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.

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

Schroeder, S. R. (2018). Do bilinguals have an advantage in theory of mind? A meta-analysis. *Frontiers in Communication, 3*, 36.

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in human neuroscience, 7*, 712.

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

Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *Journal of Neuroscience, 33*(39), 15466-15476.

Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: the effect of an irrelevant cue on information processing. *Journal of applied psychology, 51*(3), 300-304.

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

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.

Surtees, A., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology, 30*(1), 75-86.

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition, 148*, 97-105.

Tipples, J. (2006). Fear and fearfulness potentiate automatic orienting to eye gaze. *Cognition & Emotion, 20*(2), 309-320.

Van Boven, L., Dunning, D., & Loewenstein, G. (2000). Egocentric empathy gaps between owners and buyers: misperceptions of the endowment effect. *Journal of personality and social psychology, 79*(1), 66-76.

Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and social psychology bulletin, 29*(9), 1159-1168.

Van Boven, L., Loewenstein, G., & Dunning, D. (2005). The illusion of courage in social predictions: Underestimating the impact of fear of embarrassment on other people. *Organizational Behavior and Human Decision Processes, 96*(2), 130-141.

Vygotsky, L. S. (1962). *Thought and language (E. Hanfmann & G. Vakar, trans.)*. MIT Press.

Wang, J. J., Tseng, P., Juan, C.-H., Frisson, S., & Apperly, I. A. (2019). Perspective-taking across cultures: shared biases in Taiwanese and British adults. *Royal Society open science*, *6*(11), 190540.

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.

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic bulletin & review, 20*(4), 766-772.

Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1466.

Wardlow Lane, L., & Liersch, M. J. (2012). Can you keep a secret? Increasing speakers' motivation to keep information confidential yields poorer outcomes. *Language and Cognitive Processes, 27*(3), 462-473.

Wellman, H. M. (2017). The development of theory of mind: Historical reflections. *Child Development Perspectives, 11*(3), 207-214.

Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology, 15*(6), 728-755.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655-684.

Williams, G. L. (2021). Theory of autistic mind: A renewed relevance theoretic perspective on so-called autistic pragmatic 'impairment'. *Journal of Pragmatics, 180*, 121-130.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.

Wong, K. K. Y., & Raine, A. (2020). Schizotypal personality disorder. *The Wiley Encyclopedia of Personality and Individual Differences: Clinical, Applied, and Cross‐Cultural Research*, 115-120.

Wraga, M., Creem, S. H., & Proffitt, D. R. (2000). Updating displays after imagined object and viewer rotations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 151-168.

Wraga, M., Thompson, W. L., Alpert, N. M., & Kosslyn, S. M. (2003). Implicit transfer of motor strategies in mental rotation. *Brain and cognition*, *52*(2), 135-143.

Wu, S., Barr, D. J., Gann, T. M., & Keysar, B. (2013). How culture influences perspective taking: differences in correction, not integration. *Frontiers in human neuroscience*, *7*, 822.

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.

Zacks, J. M., & Michelon, P. (2005). Transformations of visuospatial images. *Behavioral and Cognitive Neuroscience Reviews, 4*(2), 96-118.

Zacks, J. M., & Tversky, B. (2005). Multiple systems for spatial imagery: Transformations of objects and bodies. *Spatial Cognition and Computation, 5*(4), 271-306.

Zacks, J. M., Vettel, J. M., & Michelon, P. (2003). Imagined viewer and object rotations dissociated with event-related fMRI. *Journal of Cognitive Neuroscience, 15*(7), 1002-1018.