The closing of the theory of mind: A critique of perspective-taking

Geoff G. Cole¹ · Abbie C. Millett¹

Published online: 12 September 2019
© The Psychonomic Society, Inc. 2019

Abstract

Theory of mind (ToM) is defined as the ability to attribute mental states to oneself and others and is often said to be one of the cornerstones of efficient social interaction. In recent years, a number of authors have suggested that one particular ToM process occurs spontaneously in that it is rapid and outside of conscious control. This work has argued that humans efficiently compute the *visual perspective* of other individuals. In this article, we present a critique of this notion both on empirical and theoretical grounds. We argue that the experiments and paradigms that purportedly demonstrate spontaneous perspective-taking have not as yet convincingly demonstrated the existence of such a phenomenon. We also suggest that it is not possible to represent the percept of another person, spontaneous or otherwise. Indeed, the perspective-taking field has suggested that humans can represent the *visual experience* of others. That is, going beyond assuming that we can represent another's viewpoint in anything other than symbolic form. In this sense, the field suffers from the same problem that afflicted the “pictorial” theory in the mental imagery debate. In the last section we present a number of experiments designed to provide a more thorough assessment of whether humans can indeed represent the visual experience of others.

Keywords Automaticity · Gaze cueing · Mental imagery · Mental rotation · Perspective-taking · Social cognition · Theory of mind · Visual cognition

In 1978, Premack and Woodruff published a seminal paper in which they emphasized the importance of knowing what other individuals know. They referred to this as “theory of mind” (ToM) and defined it as the ability to impute mental states to oneself and others (Premack & Woodruff, 1978). Although their paper was specifically concerned with the chimpanzee, the idea quickly became applied to humans, in particular, to the developing child (e.g., Wimmer & Perner, 1983). The ToM notion has now been incorporated within many aspects of cognition (e.g., decision-making; Torralva et al., 2007). One of the most recent applications has been within visual cognition. Samson, Apperly, Braithwaite, Andrews, and Bodley Scott (2010) argued that one particular aspect of ToM is computed spontaneously in that it occurs without conscious control. Specifically, the authors suggested that humans effortlessly represent the viewpoint of individuals we are currently interacting with, so-called altercentric intrusion. By

viewpoint, Samson et al. meant *visual perspective*, as opposed to what a person may think or an attitude they may have.

The present article presents a critique of the altercentric intrusion notion and the idea of perspective-taking more broadly. We divide our argument into four sections. In the first, we review the work that has been undertaken in relation to the spontaneous perspective-taking claim. We restrict the review to those studies that directly address the question of whether humans represent others' perspective when not explicitly asked to do so. Thus, it is not intended as a review of all that is known about perspective-taking.¹ In the second section, we will address an issue that has rarely been raised in the field. What can it actually mean to “take another person's perspective”? In other words, what processes occur when we consciously attempt to represent another's perspective? What is the representation concerned with? We will argue that the literature has been dominated by issues concerning function rather than representation, and that theory is now required. We will also suggest that the language

✉ Geoff G. Cole
ggcole@essex.ac.uk

¹ Centre for Brain Science, University of Essex, Wivenhoe Park, Colchester CO4 3 SQ, UK

¹ Although we will challenge the perspective-taking notion, we will use the term “perspective-taking” when referring to the phenomenon that is thought to occur by many authors. We do, however, occasionally place the term in speech marks.

employed in the field has been misleading and that it is not possible to represent another person's visual experience, as has been claimed. In the third section, we show how the same problems that persisted for one side of "the great debate" (i.e., mental imagery) that took place over more than 30 years in cognitive science, also persist for the notion that we can take another person's visual perspective. In the final section, we set out a number of experiments that we argue need to be undertaken in order to adequately assess whether we can represent what others see when not instructed to do so (i.e., spontaneously). In this section, we also posit a brief model of what it actually means to represent another person's visual perspective, spontaneously or not.

Spontaneous perspective-taking

Although the notion that humans often infer what others can see has been around in various forms for many years (e.g.,

Piaget & Inhelder, 1956), a variant of this idea gained particular strength with the publication of Samson et al. (2010). This was mainly because the authors argued that (1) the process occurs spontaneously and, (2) it is concerned with *visual* perspective. The notion was initially based on a series of three experiments Samson et al. reported in which variants of the same paradigm were employed. In what some later authors have called the "dot perspective task," participants were presented with an image of a virtual room and a number of dots could appear on either the left-hand wall, the right-hand wall, or both walls (see Fig. 1). A human agent (an "avatar") was also located in the room, positioned in the center, and faced towards either the left-hand or right-hand wall. The participant's task was to discriminate the number of dots shown from either their own perspective (on some trials) or that of the agent's (on other trials). The critical manipulation was the number of dots that would be seen by the agent and the number seen by the participant. Specifically, whether the agent saw the same number of dots that the participant saw or a different

Fig. 1 The dot perspective paradigm. Participants are required to determine how many dots are present in the display. RTs are typically shorter when the agent sees the same number of dots as the participant (upper panel), as opposed to when he or she sees a different number (lower panel)

number. For example, if two dots (only) were presented, both on the left-hand wall, and the agent looked towards the left, then both the participant and agent would see the same number of dots (i.e., two). If, however, one dot appeared on the left wall and one on the right, and the agent faced to the left, the participant and agent would see a different number of dots. The agent would only see one, and the participant, by virtue of seeing both walls, would see both. The central results showed that response time (RT) to make the dot number discrimination was shorter when the participant's view was consistent with the agent's, as opposed to when their views were inconsistent. Most significantly, this occurred both when participants made the judgment from their own perspective ("altercentric intrusion") and when they did this from the agent's perspective ("egocentric intrusion"). Because this "consistency effect" was found even when the agent's perspective was not relevant to the participant's task, Samson et al. argued that this demonstrated that the agent's perspective was computed—that the dots were "seen by the other person."

The notion that humans spontaneously compute the perspective of others is often placed within the broader theory that humans have two systems for reasoning about the mental state of others (Apperly & Butterfill, 2009). One is said to require effort and develops later in life whilst the other is assumed to be implicit and occurs in infancy. Demonstration of spontaneous perspective-taking would provide good evidence for the latter. It is perhaps for this reason that the spontaneous perspective-taking idea has generated much interest and many empirical questions. For example, how it might relate to other ToM processes, such as empathy (Nielsen, Slade, Levy, & Holmes, 2015) and nonvisual (i.e., social) perspective-taking (Moll & Kadipasaoglu, 2013); the degree to which it is associated with motor as opposed to verbal processes (Kessler & Rutherford, 2010); and whether it is "embodied" (Erle & Tropolinski, 2017). Some authors have even asked whether the perspective of *multiple* individuals can be computed (Capozzi, Cavallo, Furlanetto, & Becchio, 2014).

The argument that humans spontaneously represent the perspective of others had previously been used by a number of authors to explain results from a related paradigm. In the so-called gaze-following or gaze-cueing effect, processing of stimuli located at a position gazed at by another person is facilitated (Friesen & Kingstone, 1998). In the typical paradigm, an agent (i.e., face), located in the center of a display, looks towards the left or right and is followed by the onset of the target. The RT benefit accrued by targets appearing at a looked-at location is assumed to be as a result of an attentional shift induced by the gazing agent. In a variant of this paradigm, Teufel, Alexis, Clayton, and Davis (2010) informed participants that goggles worn by the gazing agent were either opaque or transparent. Thus, the critical manipulation was the participant's belief as to whether the agent could see the target or not. Results showed that the basic gaze-following effect was significantly larger

when the participant believed that the agent could see the inducing stimuli. This followed Nuku and Bekkering (2008), who found that attention was not shifted *at all* if the gaze cue could not see the targets. As with Samson et al. (2010), these data suggest that observers compute the visual perspective of others. However, as we will see, and unlike many later authors, Teufel et al. and Nuku and Bekkering were clear in arguing that their results were due to an *attribution* process in which participants *inferred* that the gazing agent could or could not see the inducing stimuli (i.e., targets).

In addition to the dot-perspective and gaze-following paradigms, further evidence for the perspective-taking notion has come from a procedure in which observers are shown a display that includes a single ambiguous number. For example, the number located on the table in Figure 2 is 9 from our viewpoint and 6 from the viewpoint of the agent. Surtees, Samson, and Apperly (2016; see also Kuhn, Vacaityte, D'Souza, Millett, & Cole, 2018) found that RTs to discriminate an ambiguous number were relatively long when its identity was different for the agent and observer, as opposed to when it was the same. Although the authors did not suggest that this particular effect was spontaneous, the result was again placed within the context of perspective-taking. A variant of the ambiguous number method was also reported by Zhao, Cusimano, and Malle (2015). The authors presented observers with images that included a person sitting behind a table looking at a number (as reproduced in Fig. 2). Observers were asked one question "What number is on the table?" Results showed that 42% of responses were from the viewpoint of the agent (i.e., 6), again suggestive of perspective-taking. Furthermore, these results come from a method in which observers are asked to perform a very different task to that of other perspective-taking paradigms (i.e., a single question and response).

The interpretation of the "perspective-taking; data has not, however, gone unchallenged. One limiting aspect of some studies is that they tend to provide *confirmatory* evidence. Cole, Atkinson, Le, and Smith (2016) argued that alternative

Fig. 2 The ambiguous number paradigm. The number is ambiguous because it appears as 9 to the participant and 6 to the agent

paradigms are needed that attempt to falsify the perspective-taking theory. In a number of experiments, Cole and colleagues modified both the dot-perspective (Cole et al., 2016) and gaze-following paradigms (Cole, Smith, & Atkinson, 2015) to include a physical barrier placed between the agent and inducing stimuli. If the data pattern observed in those paradigms is because the agents can see the targets, no such effect should occur when the targets are obscured from their view. Results, however, showed that the same data pattern was observed irrespective of whether or not the agents could see the targets. This does not support the view that an agent's vision is spontaneously taken. Wilson, Soranzo, and Bertamini (2017) also adopted this “seeing/nonseeing” manipulation and similarly found perspective-taking-like data (i.e., a consistency effect) when the agent was blindfolded and when the agent was replaced by a camera. This has also been the case when visibility was manipulated in the ambiguous number paradigm (Kuhn et al., 2018). However, one has to note that Baker, Levin, and Saylor (2016), and Morgan, Freeth, and Smith (2018) have reported the absence of a perspective-taking effect when the barrier method was employed, thus supporting the perspective-taking theory. Furthermore, Furlanetto, Becchio, Samson, and Apperly (2016) also found evidence for the perspective-taking notion when they manipulated observer *belief* about whether an agent could see through colored goggles or not (as in Teufel et al., 2010; see above). Specifically, a consistency effect was found when participants believed that the agents could see through the lenses, but no such effect occurred when they believed vision was obscured. Conway, Lee, Ojaghi, Catmur, and Bird (2017), however, were not able to replicate this belief-manipulation effect.

Why does the consistency effect occur?

Failure to elucidate on the mechanisms and processes that generate an effect is common when a phenomenon is first reported. This is understandable as it would be unreasonable to always expect authors to immediately undertake and report all the work necessary to understand the mechanisms responsible for a phenomenon. Visual cognition researchers will often examine an effect's various parameters and “boundary conditions,” initially asking questions such as how long a phenomenon lasts, is it automatic, is it perceptual, attentional, or as a result of a decision process (e.g., inhibition of return: Posner & Cohen, 1984; attentional blink: Raymond, Shapiro, & Arnell, 1992). However, it is also incumbent upon researchers to posit explanatory mechanisms and processes that can be tested against alternative explanations. This has been very much lacking within spontaneous perspective-taking research. Instead, the field has been dominated by a long list of empirical studies that, aside from their inherent interest, have

not generated many explanations. The perspective-taking notion is now in need of a theory.

Samson et al.'s (2010) paper did suggest an account based on a variant of the processes associated with the classic gaze-cueing effect (described above). Specifically, the authors argued that a shift of attention induced by the avatar's gaze would enable observers to rapidly compute its line-of-sight. The significance of this is that such a computation is thought to be central to one particular perspective-taking process. With so-called Level 1 perspective-taking (Flavell, Everett, Croft, & Flavell, 1981), an observer knows that another person can see an object, a process that can occur via a line-of-sight computation (Michelon & Zacks, 2006). This contrasts “Level 2 perspective-taking” in which an observer knows how an object looks to the person. The problem with the gaze-cueing explanation of the perspective-taking phenomenon, however, is that, for many authors, gaze following is itself due to a ToM process. Recall that Nuku and Bekkering (2008) and Teufel et al. (2010) reported reduced or absent gaze following when the gazing agents could not see the targets. If gaze following is indeed modulated by ToM, then the gaze-following effect cannot be a mechanism with which the consistency phenomenon occurs. This would mix up the phenomenon with the explanation. If, however, the gaze following effect is *not* due to ToM, then any attentional shift induced by the agent's gaze could be considered a confound (i.e., an effect based on attentional orienting rather than ToM). The essential problem is that the consistency manipulation of the dot-perspective paradigm maps directly on to the “validity” manipulation used in the basic gaze-following experiments.

A subtle variant of the attentional shift hypothesis was proposed and examined by Santiesteban, Catmur, Hopkins, Bird, and Heyes (2014). These authors argued that the avatar's gaze and body orientation are likely to shift attention to the gazed-at side of the display. This would facilitate responses when the only dots in the display are located on the wall looked at by the agent (i.e., on consistent trials). Likewise, RTs would be longer when a dot appears on the wall not gazed at (i.e., on inconsistent trials) because attention would have to reorient to the other side of the display. In support of this account, Santiesteban et al. found that an arrow, which does not have a mental state, but still shifts attention, resulted in a consistency/cueing effect as well as an agent. The authors concluded that results obtained in the dot-perspective paradigm are thus due to domain-general processes, rather than perspective-taking. Wilson et al. (2017) also found consistency effects when they employed an arrow, camera, and agent, as did Nielsen et al. (2015) when using dual-colored squares in place of an agent. However, generating a consistency effect with other stimuli in addition to an agent does not falsify the argument that spontaneous perspective-taking also generates the consistency effect. As Cole, Atkinson, D'Souza, and Smith (2017) pointed out, showing that arrows can shift attention does not challenge the perspective-taking

idea. Mimicking the typical results obtained in a paradigm under condition x does not mean that those typical results are due to x . This would be equivalent to arguing, for instance, that in the natural environment plants grow solely as a result of ultraviolet light because UV light can be used to mimic such growth in a laboratory. The essential point here is that the consistency and attentional cueing effects could be driven by different processes, even though they result in the same behavioral data. Thus, despite suggestions to the contrary, showing a consistency effect with arrows cannot tell us whether results from the dot-perspective task are due to processes associated with ToM.

A number of studies have examined whether the male and female avatars employed by Samson et al. (2010) are actually able to shift attention laterally. Cole et al. (2017) found no orienting effect when the agent and target appeared simultaneously, nor when the target followed the agent by 100 ms. Similarly, Gardner, Hull, Taylor, and Edmonds (2018) found no effect at 100 ms (agent–target interval) but did so at 300 ms. Bukowski, Hietanen, and Samson (2015) did, however, find an orienting effect at an interval of zero ms when attention was artificially drawn to the avatar location. Overall, these data suggest that any agent-induced attentional shift requires some time to occur—certainly too long for this to be considered rapid/spontaneous.

Michael et al. (2018) reasoned that if spatial cueing underlies the consistency effect then the presentation of two avatars *simultaneously*, as opposed to one or none, should facilitate performance on the dot perspective task. This rationale was based on the fact that two different locations can be simultaneously facilitated (in the precuing paradigm) when two cues are presented, relative to a baseline in which no cues occur (e.g., Maylor, 1985). The authors further reasoned that the perspective-taking theory predicts the opposite because two avatars, as opposed to one, will always have a different perspective when looking in different directions (and thus dots; but see Capozzi et al., 2014, for evidence that the perspective of two avatars is not computed). Although Michael et al. found that performance was facilitated with two avatars, this was only observed when an interval of 800 ms occurred between avatar/s and dots. When these were presented simultaneously this effect was reversed such that performance was better with one as opposed to two avatars. The authors suggested that perspective-taking mechanisms may operate at earlier intervals and spatial cueing later. Such an effect has also been reported by Gardner, Bileviciute, and Edmonds (2018). Gardner et al. were primarily interested in whether the dot-perspective and gaze-cueing effects are modulated by (what might be called) the “rubbernecking effect”; the gaze-following phenomenon has been shown to be greater when the agent’s head is oriented at a different angle to its body (Hietanen, 2002). It has been suggested that this is because such a stance indicates explicit intentional behavior (Moors, Germeyns, Pomianowska, & Verfaillie, 2015). Gardner et al.

employed this rubbernecking manipulation in both the gaze-cueing and dot perspective-paradigms. The authors reasoned that whereas gaze following should be modulated by avatar stance (i.e., Hietanen, 2002), the dot-perspective effect should not. This is because the avatar’s perspective of the stimuli in the latter is the same irrespective of its stance. As predicted, these effects were observed.

A related avenue of research has attempted to tease apart the possible differences in the mechanisms responsible for the results observed in the dot-perspective and gaze-cueing effects. Michael and D’Ausilio (2015) argued that the dot-perspective paradigm may engage both domain-general (e.g., attention orienting) and domain-specific (e.g., mentalizing) processes. Although faces, as well as arrows, provide a clear directional cue, the authors made reference to the fact that the former induce a number of effects not observed with the latter. For instance, gaze cues, as opposed to arrow cues, induce orienting effects even when participants are informed that targets appear at the gazed-at location on only a minority of trials (e.g., 20%; Driver, Davis, Riciardelli, Kidd, Maxwell, & Baron-Cohen, 1999). Michael and D’Ausilio thus argued that whilst social and nonsocial cues are represented by different functional systems, they also engage the same attentional processes.

A further concern has been the issue of whether perspective-taking can be considered as a spontaneous/automatic process. As Moors and De Houwer (2006) pointed out in their extensive review, many different authors have associated automaticity with a large number of processes and mechanisms (e.g., efficient, effortless, fast, unconscious, goal independent, difficult to suppress). Cole et al. (2016) argued that “*spontaneous* computation of others’ perspective should not require observers to occasionally assume this perspective” (p. 166). In the dot-perspective procedure, observers are typically required to perform the dot-number judgement from both their own perspective and, on other trials, from the perspective of the agent. This often occurs within-block such that participants are informed at the beginning of each trial which perspective they should adopt. One of the potential consequences of this is that participants may represent the agent’s perspective even when they are not explicitly instructed to do so; they know that perspective is an important part of the experiment. The subtlety with which top-down knowledge is able to influence attention to features of a display has been well established since the findings of Folk, Remington, and Johnston (1992). This “attentional control settings” work has shown how a feature of a display item that is nominally task irrelevant can in fact form part of an observer’s attentional set. Importantly, this type of top-down influence has been shown to occur in perspective-taking paradigms (e.g., Stephenson & Wicklund, 1983). That is, merely instructing observers to consider their own perspective seems to induce consideration of an alternative perspective. Cole et al. (2017; Cole et al., 2016; Cole et al., 2015; see also Conway et al.,

2017) did not therefore include the manipulation in which either the agent's perspective was taken or the participant's. Results showed perspective-taking-like data under this condition (but recall also did so when the agent could not see). Bukowski et al. (2015) argued that when participants are required to take their own perspective in the dot-perspective paradigm, this induces what they called a social mind-set that increases the salience of the avatar. The authors tested this by overlaying instruction prompts (e.g., dot number) onto the avatars with the intention of artificially increasing their salience. Bukowski et al. (2015) found that this manipulation induced a gaze-congruency effect even when the avatar and targets were presented simultaneously. As above, this suggests that some form a top-down set for perspective operates in the dot perspective and related tasks.

In summary, studies have now reported a robust effect showing that responses are facilitated when an agent looks towards a stimulus that an observer is required to respond to. It is however uncertain as to whether this, and related effects, is due to humans spontaneously representing the visual perspective of other individuals. In terms of article numbers and experiments, it is certainly true that the large majority support the perspective-taking notion. However, apart from the fact that positive evidence is far more likely to find itself in a journal, there are too many reports of perspective-taking when it should not occur (i.e., when the agent cannot see any targets).

What can it mean to take someone's visual perspective?

In this section we will ask a fundamental question that, given the amount of work undertaken on the subject, has rarely been raised in the perspective-taking literature. What does it mean when an individual says that they are taking another person's perspective? What is being represented or computed? Note that the question is concerned with the processes that occur when another person is explicitly asked to take an alternative viewpoint. That is, when all perspective-taking researchers would agree that a different viewpoint is attempted to be taken, such as when a participant is specifically asked to consider what another individual can see. This is critically different from the question of what is causing the consistency effect in the dot perspective and related paradigms (as discussed in the previous section). We will argue that there are problems with the notion of perspective-taking. Our central contention is that it is not actually possible to represent another person's visual experience, as claimed. If this is indeed the case, it follows that perspective-taking cannot occur spontaneously.

We will also reject the central distinction between Level 1 and Level 2 perspective-taking (e.g., Flavell et al., 1981). Recall that the former is defined as a person knowing that another person can see an object, and the latter is knowing

how it looks to them. Thus, the dot-perspective paradigm, and most of the work reviewed in the first section, was designed to examine Level 1 processes. Although Level 1 perspective-taking is of course based on a simple truism (i.e., we can know what another person can see), we argue that only Level 2 can really refer to a visual perspective—how something looks. Thus, it is difficult to conceive of how a person can have a visual perspective of an object without knowing how it looks. If one does not know how the object will look, it is not a perspective. This is supported by the fact that one only has to know if a straight line (of sight) can be drawn between an agent and object in order to know if they can see it (Michelon & Zacks, 2006). Of course, it is very useful in everyday parlance to refer to both as “perspective-taking”. However, as we will set out, problems arise if researchers take the perspective-taking notion to mean the computation of an actual perspective—that is, visual experience.

The perspective-taking homunculus

One thing that is clear from the perspective-taking literature is that, by definition, the agent's *perspective* is somehow being represented. A consequence of the lack of perspective-taking theory is that there is a real danger of researchers inadvertently (or not) advocating the idea that an *actual perspective or visual experience* is computed. Consider what has been said within the literature about the act of taking another's perspective. Capozzi et al. (2014), for instance, stated that “in simple perspective-taking tasks, one's own and others' *visual experience* influence each other” (p. 1, italics added). They later add that “observers may efficiently process what another person sees” (p. 2). The notion of a visual experience being computed was also stated by Samson et al. (2010) in their original description of the phenomenon. Similarly, whilst Moll and Kadipasaoglu (2013) refer to “learning about others’ as well as one's own ‘snapshot’ perspectives in a literal, i.e., optical sense of the term” (p. 1), Erle and Topolinski (2017) refer to “imagining how the visual world appears to that person” (p. 1). Kessler and Rutherford (2010) also make reference to picture-like generation (i.e., imagery) as well as referring to the representation of a “virtual perspective” and invoke (in quotation marks) the notion that “I see the world through your eyes.” Picture generation is also central to one of the few attempts to describe what it actually means to take another person's visual perspective. Although it was published some years before the Samson et al. work appeared, Amorim (2003) presented a model directly concerned with perspective-taking representations and processes. We do suggest, however, that the model is limited because of its explicit reliance on image generation and Kosslyn's depictive explanation of mental imagery (see below).

For many authors, then, perspective-taking equates to the phenomenology of seeing—the actual visual experience. This, of course, is precisely why the phenomenon is referred

to as *perspective* taking and not, for instance, “position taking.” This is also why imagery is often invoked. This is not to say that many authors within this field do not refer to and examine the computation of the relative positions of humans and objects. Indeed, there does seem to be something qualitatively different about representing a scene from a different position when that position is occupied by another human as opposed to an inanimate object (e.g., Bertamini & Soranzo, 2018; Tversky & Hard, 2009). We do, however, suggest, or perhaps remind authors, that it is not possible to take the perspective, the percept, of another person. Indeed, it is not possible to compute the perspective from *any* position that is not our own current perspective. Apart from its useful everyday usage, the perspective-taking notion has been misleading and references to “visual experience,” “literal snapshots,” and so forth, without explaining what these mean is unhelpful.

Assuming another person’s snapshot visual experience would require some medium or mechanism, a homunculus, that represents the sensory processes associated with the other person’s perspective. The central problem can be illustrated with a demonstration based on the fact that the color of a stimulus can appear very different dependent upon viewing distance. This is also a nice illustration of the fact that the world is not colored; color, as with many object properties, is not “out there,” it is a construction of the brain. A checkerboard is generated that is no larger than approximately 3–4 cm square (see Fig. 3). The small squares are alternatively colored, red and green. When this is viewed relatively close (i.e., about 10 cm), the red and green squares can be easily resolved and we see those two colors. Thus, the perception is as described above (i.e., a red/green checkerboard). To determine whether we really can take the perspective of another person, try the following with a friend. Hold up Fig. 3 so that you see it

Fig. 3 A test of perspective-taking. View the image from about 3–5 cms and ask a friend to look at it simultaneously from about five meters (i.e., peering over your shoulder). Tell them to take *your* perspective and ask them what color or colors you see. (Color figure online)

from about 10 cm and ask a friend to look at it simultaneously from about five meters (i.e., peering over your shoulder). Importantly, ask them to take *your* perspective, not their own, then ask them, “What color or colors do I see”? We suspect they will say “yellow”.² They will have great difficulty in genuinely being able to take your perspective. At a distance of greater than approximately four or five meters, the small squares cannot be resolved. Instead, a single uniform yellow square is usually perceived. *To take another person’s perspective has to mean computation of how this kind of stimulus will change as a person’s position in space changes, otherwise it is not a perspective.*

It is tempting to say that this experiment/demonstration could not possibly generate a perspective-taking effect. One might contend that, of course, the reason the square is perceived as red and green at a short viewing distance is because of a basic sensory mechanism; a function of a low-level process that is not available at the far position. One might additionally think that although perspective-taking is not possible in this specific situation, it *is* possible for most situations. Indeed, we as humans often consider another person’s (visual) “perspective.” However, the point here is that it is not possible to take another person’s perspective for *any* situation. The problem faced in the colored square example is the same for all perspective-taking situations. One’s own visual perspective is *always* as a result of a sensory process. In fact, some of these sensory processes are due to properties of the retina. For example, the number of rods at the fovea, which processes information where an observer is fixating, is relatively low compared with the number that occur at around 15–22 degrees eccentricity (Curcio, Sloan, Kalina, & Hendrickson, 1990; Osterberg, 1935). This has the odd consequence, well known to star gazers, that resolution of an object under very low luminance conditions is actually better when an observer looks away from the stimulus by around 15–22 degrees. This kind of mechanism is not, of course, available to an observer who is attempting to take the perspective of another individual, even if the observer knew this fact about the retina. It would require a homunculus to represent what the agent can see.

This homunculus cannot, however, help with another major problem for the perspective-taking notion. One of the central points to come out of all the work on the change-blindness phenomenon is that the very detailed visual representation we experience is something of an illusion (e.g., O’Regan, Rensink, & Clark, 1999; Simons & Levin, 1997). We appear to have a complete “picture” of the world and use our

² The perception of a uniform yellow square can depend on the quality of the color print, or the luminance of the device. For instance, the lower the luminance of the squares the less likely that yellow will be perceived; dark yellow is effectively brown. If yellow is not perceived, or “green” is stated, try asking your friend whether they can see any red. Red is very unlikely to be seen irrespective of luminance.

attentional spotlight to view different parts of the scene. However, despite this subjective impression, our visual representation is known to be extremely impoverished. This is why large changes to parts of a display often go unnoticed unless attention is located at that position/object. The critical point here is that *our own* perspective, our own visual experience, is wholly determined by our attention.³ There is no picture that passively sits there whilst our attentional spotlight traverses it. The picture or percept is generated by the spotlight; it is a *consequence* of attention, and no other observer can know where our attention is. Thus, we can never know what another person is seeing or experiencing. Even if we are observing a person who appears to be engrossed in a book, perhaps making notes in the margin—they may actually be attempting to listen to another person's conversation and using the book as a ruse. In the dot-perspective task or the gaze-cueing procedure, the issue of what the agent is attending to is relatively unambiguous because there is not much that the agent can be looking at. However, as soon as more than one object is within an agent's line-of-sight, it will be unclear what she is wishing to see, if anything. This, of course, is magnified when the agent is located within a real-world scene with multiple objects. At best, computation of another's viewpoint only illuminates those objects that are currently located on the side of a scene that a person is currently faced towards. One has to reiterate that the perspective-taking notion is about ToM—the mental state of others. A system that cannot represent what another person is attending to cannot help with ToM computations. One might even say that two homunculi are required, one to represent what is being seen (as above), the other to represent what is being attended to.

Not a question of semantics

One might contend that the issue only concerns what is exactly meant by the phrases and vocabulary employed within the field. One might further argue that the present authors have constructed a straw man; perspective-taking researchers don't *actually mean* that humans represent another person's perspective in a "literal" "optical sense," despite writing this. However, it is a moot point as to what authors do mean or infer when they state that a "visual experience" is computed. One can only, after all, evaluate what is said and written, not what a person may privately think. Despite this, it does seem that perspective-taking researchers are indeed referring to and mean processes beyond nonvisual representation of others viewpoint. As we stated in our first section, an effect based on the computation of "perspective" had already been extensively reported by a number of authors with respect to the gaze-following paradigm (e.g., Nuku & Bekkering, 2008; Teufel et al., 2010).

³ For a concise background on the relationship between attention and awareness, see Lamme (2003).

Importantly, the basic gaze-cueing phenomenon had been explained with mechanisms associated with ToM (e.g., Hietanen & Leppanen, 2003; Mathews, Fox, Yiend, & Calder, 2003; Holmes, Richards, & Green, 2006). For example, it was an explicit ToM component that motivated research assessing whether the expression of a gazing agent's face can increase the gaze-cueing effect. If its expression is fearful, the argument went, then one might expect a face to shift attention more rapidly to "motivationally significant objects" (Mathews et al., 2003)—that is, "I need to know what that person is scared of." The important point is that nonvisual-experience ToM processes were already assumed to be occurring when an agent gazes towards an object or location. That is, these authors (e.g., Nuku & Bekkering, 2008; Teufel et al., 2010) were explicit in stating that observers make *attributions*, they *infer* (spontaneous or not), what an agent can see. However, the very fact that effects of what an agent sees were reported subsequent to this, but are now referred to as "perspective-taking," suggests that researchers now felt that something additional was happening above and beyond attributions and inferences—something that had not been considered before. That something has to be, as argued, an agent's actual perspective—their visual experience.

In summary, we suggest that the perspective-taking field has not been forthcoming in describing what it actually means to take another person's perspective. Partly because of this, many authors have inadvertently suggested that humans can experience another's visual perception. We argue that it is not possible to take the perspective of another person. This would require computation and representation of both a person's sensory system and attentional system; two homunculi would be needed.

Mental imagery

In the following section, we argue that there are close parallels between the perspective-taking notion and the 'great debate' (Block, 1981) that occurred within cognitive psychology from the early 1970s. We will suggest that the problems that afflicted one side of the debate also afflict the notion that we can take the perspective of others. We also argue that the other side of the debate provides a good explanation of what is happening in so-called perspective-taking.

The mental imagery debate meant different things to different people. For some authors it meant whether imagery uses the same brain regions that are used when viewing a real object/scene; for other authors it meant whether images are based on a symbolic representation or a depictive representation; and for others it meant whether images have functional utility or are a nonuseful by-product of reasoning. The crux of the matter, however, was the issue of what can be concluded from a variety of experiments in which participants were required to make judgements and decisions based on what they

“saw” in their mental image. In perhaps the central experiment, participants were required to move their mental spotlight across an image from one position to another. Results typically showed that longer distances take longer to traverse. For instance, when imagining a map of the U.S., it takes the “mind’s eye” longer to move from Boston to Los Angeles than from Boston to New York. The reason such RT differences occur when looking at a real (and linear) map is that the map, of course, has a spatial structure. Kosslyn and colleagues (e.g., Kosslyn, 1981) argued that this is also the case with mental images. That is, images were said to be based on a spatial representation; they had distance; they were “depictive.” Thus, the processes involved when “looking at them” were considered to be analogous to perception.

Consider the following demonstration. Generate a mental image of the letter “C” and rotate it 90 degrees anticlockwise. What letter do you now “see”? You are likely to provide the correct answer (“U”) relatively easily. The critical question of the mental imagery debate was how does an observer arrive at the answer to this and related demonstrations? Kosslyn and colleagues (e.g., Kosslyn, 1981) argued that observers effectively “read off” the resultant image in order to arrive at the correct response. As Kosslyn (1981) stated, an image “once formed, can be operated upon in various ways” (p. 46). In other words, it is the very nature of the images that generated responses. Images were considered as having intrinsic optical and geometric properties that are functional in the reasoning process. Pylyshyn (2002), for example, in contrast argued that something very different is happening. In his view, observers simply *know* what shape appears when a “C” is rotated anticlockwise. They then use this knowledge to *simulate* what would happen if they actually saw the letter rotate. Observers are not of course aware of this process, but they are essentially making the image do what they want based on what they know. In this view, mental images are much like having a thought; we can generate limitless images some of which need not bear much resemblance to actual percepts. The phenomenal experience is real enough, and observers can certainly “see” an image and “rotate it,” but in Pylyshyn’s view the stimulus is giving them nothing that they can use to provide the correct response. It is in this sense that mental images were said to be “epiphenomenal” rather than functional. This is much like individuals who state that they are not able to generate mental images but can still correctly state what color their car is; they just know the answer. Thus, for Pylyshyn, it is not the intrinsic nature of mental images that generates the paradigm results, but what observers know about how the world works. Essentially, Kosslyn argued that an image is an intermediary that enables correct answers to occur, whereas for Pylyshyn it is the correct answers that enables the images to occur. Hence, when there is no correct answer, there is no (correct) image.

The fundamental problem with all “depictive” explanations of mental imagery is the same one faced by the suggestion that

an agent’s visual experience is computed; an internal “eye” is required to view the scene. As Kosslyn and colleagues sometimes acknowledges (e.g., Kosslyn, Pinker, Smith, & Shwartz, 1979), suggesting that a mental image is a real image only pushes the problem back a step. Thus, the homunculus required in Kosslyn’s view of mental imagery is the same homunculus required for perspective-taking, one that can compute all types of sensory processes including color constancy, contrasts at different spatial frequencies, and so on.

Knowledge not vision in perspective-taking

In the standard account of the dot-perspective paradigm, the agent’s visual experience is considered as the central component that generates the consistency effect; it is its perception that interferes with the participant’s own representation of the scene. It is much like any interference paradigm in which a task-irrelevant stimulus slows down target processing. In the case of the dot perspective paradigm, the agent’s visual experience is said to be the irrelevant stimulus. Our preferred explanation however posits that Pylyshyn’s account of the data observed in mental imagery experiments also explains the data observed in perspective-taking experiments. That is, participants make an assumption of what the agent *knows*; it is knowledge that is the important component that drives the effect rather than vision.

This knowledge explanation of the perspective-taking results is perhaps best seen in the central experiment of Zhao et al. (2015), described more fully in our first section. Recall that participants are presented with a photograph of an agent sitting behind a table on which an ambiguous number is located (see Fig. 2). In Zhao et al.’s variation of this ambiguous number paradigm, participants are asked one question: “What number is on the table?” Aside from the presumed perspective-taking effect when the agent is looking at the number, the most interesting condition for the present argument is the effect that occurs when the agent looks away from the number. In this condition, he is appearing to gaze over the shoulder of the participant/photographer at an object in the distance. That is, he is clearly not looking at the number. Despite this, 13% of participants still reported the number from the “perspective” of the agent. As with the barrier manipulation described previously, this cannot be due to the agent’s visual perspective being computed, because he is not viewing the number. However, it is clear that the agent did at some point, probably a couple of seconds previously, see the number. Indeed, it’s highly unlikely that he never saw the number and doesn’t know its identity. The agent obviously just knows what the number is, and an observer knows this.

This knowledge-driven account concurs with results from experiments in which a participant’s belief about what an agent can see is manipulated (e.g., Furlanetto et al., 2016; Morgan et al., 2018). Indeed, our account must predict these

results precisely because knowledge of what the agent can see, and thus what she knows, has been manipulated in those experiments. This account was effectively put forward by Nuku and Bekkering (2008) and Teufel et al. (2010) when they referred to participants “inferring” what the agent knows. This explanation may also account for findings from experiments in which a physical barrier, located between the agent and target, does modulate the consistency effect (Baker et al., 2016). The knowledge account predicts a consistency effect because the participants know that the agent cannot see beyond the barrier. Again, participants may simply have knowledge as to what the agent knows, rather than what he can see. Furthermore, this alternative account may still concur with data from other occluding barrier experiments that show the occluder has no effect on responses (e.g., Cole et al., 2015); observers in all such experiments don’t know what the agent has seen and thus knows.

In a following section, we propose a number of future experiments. It is worth noting here, however, that when there is no ambiguity as to what an agent knows, then a particular pattern of data should be expected. For example, as suggested above, no matter where the model in Fig. 2 is looking, even when looking behind her back, we can be almost certain that she has at some point seen the number on the table; it could not easily be placed there without her knowledge. Thus, when the participant can be confident that the agent has viewed a stimulus, we can firmly predict which response the participant will give. Indeed, one could imagine an experimental paradigm in which observer knowledge of what the agent knows could be manipulated. This could include the standard dot-perspective experiment. In the basic paradigm it is not always clear what the agent has seen and thus knows about the dots in the room. That is, whilst an observer knows what the agent knows in the consistent condition (i.e., “she knows there are two dots in the room”), this observer cannot be sure in the inconsistent condition (i.e., “When did the dots appear?” “Has she looked behind her?”). Indeed, it is possible that knowledge uncertainly contributes to the basic dot-perspective findings.

The knowledge account can also explain the rubbernecking effect (Hietanen, 2002), in which gaze cues shift attention more rapidly when the gazing agent’s head is oriented 90 degrees to its body orientation, as opposed to when its body and head are both oriented the same way. In the former case, the agent appears to be straining somewhat to view a stimulus. As mentioned earlier, this has been explained with reference to a ToM process, specifically intention (Moors et al., 2015). Presumably, the attribution is that the agent really wants to see the stimulus being viewed. This is a clear case of knowledge (of an agent’s motivation) influencing the degree to which an agent shifts attention. However, it is worth noting that Gardner, Bileviciute et al. (2018) found that this avatar-stance manipulation does not modulate perspective-taking in

the dot-perspective task. This concurs with the perspective-taking notion because the agent sees the same number of dots in both the rubbernecking and nonrubbernecking conditions.

One also has to remember that, at the risk of emphasizing the obvious, when we consider the visual perspective of another individual, or the viewpoint from any object, we are not located at that position; we are displaced from it at our own location. In order to perform anything akin to taking a perspective, we have to “get there.” This is a basic truism. We cannot have a literal visual experience of being there because we are physically not. The nearest alternative to the experience of being there is to have a mental image. Indeed, to have anything like the visual experience from a position that we are not currently in has to mean that some form of imagery is required, otherwise it is not a perspective. However, it is interesting to note, but never mentioned in the perspective-taking literature, that when a person attempts to take an alternative perspective, there is no phenomenal experience of a different viewpoint being generated; a new “picture” does not occur. This can be contrasted with our ability (although not in everyone) to generate a mental image. At least with imagery, there is a real sense of a visual perception, of “seeing” something that is not present. In contrast, when we perspective-take our visual experience does not change, and we predominantly saccade towards the objects that an agent is looking towards (Kuhn et al., 2018).

The agent as a directional cue

Not only does the agent (e.g., in the ambiguous number paradigm; Zhao et al., 2015) have knowledge about the number and we as the observer know what he knows, he provides us with an important cue. Although he is not looking at the number, results are clearly being driven by his presence; no perspective-taking-like data occur when he is not there (in the control condition). We suggest an account involving two stages that help to generate the effect. In the first, the seeing agent acts as an anchor or reference point that makes its position in space particularly salient (see Epley, Keysar, Van Boven, & Gilovich, 2004). Attentional workers would say that the agent attracts attention. In the second stage, mental rotation-type processes occur in which the observer orients to the reference point and represents the direction in which the agent faces. Note that the rotation process acts upon a symbolic representation of the scene, rather than pictorial. The task stimuli are then represented from this position/side. Subsequently, this representation is facilitated relatively to other representations leading to, for instance, shorter RTs. We state that the observers represent the “direction” from the resultant position because one has to account for the finding that the identity of an ambiguous number is likely to be given relative to this direction, even when a participant is not currently viewing the number—for example, if the agent is

gazing over the shoulder of the participant/photographer (Zhao et al., 2015) or if the agent was looking towards the number, but could not see it because of an occluding object—for instance, if she was holding a newspaper. Thus, this account is subtly but importantly different to the standard perspective-taking notion because it is not based on a perspective. One of the central assumptions of our alternative account is that any stimulus that acts as a directional reference point can induce perspective-taking-like effects as when an agent does. For example, Millett, D'Souza, and Cole (2019) reported an experiment in which the agent in the ambiguous number paradigm was omitted (in one condition), but her chair remained. Other reference objects were included such as a monitor, keyboard, mouse, book, and phone placed on the table. These faced the chair, as with any office desk scene, thus providing the participant with a directional cue suggestive of which way the number should be read. Result showed that the number was reported from the direction/viewpoint of the chair to the same degree as when an agent sat in the chair. As noted previously, however, this is not to say that there is nothing special about having an agent in a display when observers are asked to make judgements that rely on the coding of spatial relationships (e.g., Becchio, Del Giudice, Dal Monte, Latini-Corazzini, & Pia, 2013; Bertamini & Soranzo, 2018; Tversky & Hard, 2009). Thus, a nonhuman reference point may not induce effects to the same degree as a human agent.

Future experiments

In the following section, we outline two broad strands of research. In the first, experiments are proposed that are directly concerned with the question of whether perspective-taking occurs when an observer is not explicitly attempting to consider an alternative viewpoint (i.e., spontaneously). We argue that the proposed paradigms are necessary if we want to know whether the perspective-taking notion is correct or not. The experiments are agnostic as to the processes involved; the question is simply whether spontaneous perspective-taking occurs. In the second, we briefly set out a model of what it actually means and what may be happening when an observer consciously attempts to take the perspective of another person (i.e., nonspontaneously). This therefore is concerned with *representation* of an alternative viewpoint.

Assessment of spontaneous perspective-taking

One of the fundamental problems with the various perspective-taking paradigms employed so far is that they are rarely designed to index *vision*. Perspective-taking clearly concerns perception—what another person can see. The dot-perspective paradigm, however, is what would normally be

referred to as an attentional paradigm. An observer is given a task (i.e., a dot number judgement), an irrelevant stimulus is presented (i.e., the agent), and relative RT is recorded. It is an attentional paradigm that is intended to index vision. Of course, relative RT can be used to measure effects that occur at various levels in the system, including phenomena that are due to perception, attention, decisions, memory, and motor processes. Sometimes, perhaps for practical rather than theory-generating reasons, it is good enough to know that a difference occurs irrespective of where. However, if we want to claim that a phenomenon is due to a particular mechanism, then an appropriate paradigm is clearly required. Since students of the perspective-taking notion are concerned with what others can *see*, that is, their perspective, it makes sense to employ paradigms that can only be due to an assumed perception. The problem with RT measurement is that an observed effect could be due to mechanisms and processes that occur at any of the stages mentioned above. For this reason, manual RT measures are often avoided in work on perception. Indeed, rather than passively measure a particular effect, some authors argue that manual RT responses can influence or even generate the phenomenon being examined. Milliken and Tipper (1998) made the point that it is like posing the question of whether lights inside fridges are constantly illuminated. The empirical data certainly support the theory that they are (lights are always on when the door is opened). Of course, it is the act of looking, of measuring, that generates the effect being examined. Few non-RT experiments exist in the perspective-taking literature. Examples include a variant of the ambiguous number paradigm (Zhao et al., 2015), change-detection experiments (Cole et al., 2015; Morgan et al., 2018), the assessment of what can be seen in a mirror (Bertamini & Soranzo, 2018), and a variant of the “sandbox” assessment of false belief (Marshall, Gollwitzer, & Santos, 2018).

A further problem with almost all the experiments that have been used to examine spontaneous perspective-taking is that the participant and the agent can see the same thing (in the consistent conditions). Paradoxically, this confounds any effect of what the agent can see with what the participant can see. Recall that the whole issue is about whether we take the perspective of the agent, not our own, so both seeing the same stimulus is experimentally problematic. What is required is a paradigm in which the agent can see something that the participant cannot. We then examine whether that stimulus influences participant responses—a stimulus the participant cannot see. In principle, we need something equivalent to the scenario in which a physical barrier is placed between the participant and the critical stimulus, but the view between agent and the stimulus is unobstructed. We say this is required in principle because this type of set up could not of course possibly generate a perspective-taking effect—such an effect would really be a case of Feeling the Future. Thus, the participant and agent simply have to see the same stimulus. However, as we

describe below, it is possible to present the same stimulus to both participant and agent, but they perceive that same stimulus very differently.

In order to test the pictorial theory of mental imagery, Slezak (1995) designed and used a number of animal figures. For instance, look at the duckling shown on the monitor in Fig. 4 for three or four seconds and form a mental image of the animal before you read on. Without looking back at the figure, rotate the duckling through 90 degrees in an anticlockwise direction. What animal do you now see? It is unlikely that you are able to correctly state a rabbit (if you did not initially spot it when you first looked at the figure). This suggests that observers cannot rotate an image and then “read off” the image that results, thus supporting Pylyshyn’s knowledge account of mental imagery; when you don’t a priori know the answer, you can’t form the image.

Unlike the well-known ambiguous figures (e.g., young/old lady), this kind of Slezak animal is not at all apparent unless rotated in a flat-plane orientation; it is somewhat hidden or invisible to the participant. In order to test the spontaneous perspective-taking claim, one could present participants with displays in which an agent views such a figure lying down (i.e., agent is positioned horizontally; see Fig. 4) and the experimenter measures the degree to which, if at all, participants responses are influenced by the figure seen by the agent (i.e., at the different rotation). This could be indexed in a variety of ways. For instance, after viewing Fig. 4, a stem completion task could be presented in which participants are asked to complete “_ a _ b _.” If the agent’s viewpoint is represented, participants should be more likely to respond “rabbit” compared with when the agent is not present. A control condition would, on some trials, present a rabbit in the correct upright orientation, as seen by the participant and in isolation (i.e., no agent). This would assess/ensure that such a stimulus can indeed prime observers to indicate “rabbit.”

Relatedly, any paradigm could be used where the agent and participant view a scene in a different orientation, together with a paradigm in which responses are known to be different as a function of stimulus orientation. Thus, one could also

Fig. 4 The duckling presented by Slezak (1995)

imagine a classic mental rotation experiment (Shepard & Metzler, 1971) in which an agent is viewing one of the sample objects. Crucially, its perspective happens to be the correct view of the target object. One should expect responses to be facilitated under this condition relative to when the agent cannot see the correct orientation. Evidence based on mental rotation was recently reported by Ward, Ganis, and Bach (2019). Participants were required to indicate whether a letter “R” was presented in its correct canonical orientation or as its mirror image. The letter would also be rotated in a flat-plane orientation by various degrees. Importantly, an agent could also be present in the display looking at the item. As expected, the authors found that RT increased as the letter’s orientation increased away from its canonical position. This, however, was not only the case for letters relative to the participant, but also relative to the agent; an effect clearly suggestive of perspective-taking. To reiterate the central point we want to make here, in order to know whether an observer is representing what another person can see, it is more expedient to employ a stimulus/target that looks different to the agent and observer. Notice that the critical stimulus (i.e., target) in the Ward et al. paradigm, although not hidden or invisible, is still somewhat different for the participant and agent. Indeed, one can argue that the change in the perception of the letter from the two positions is more different than the change in the perception of the red discs in the dot-perspective paradigm. To put another way, the phenomenal experience of a red disc does not change much when viewed from the position of the agent or observer. For instance, it stays relatively round. When a stimulus looks relatively similar one cannot be sure whether a response is based on what the agent can see or what the observer can see. This was illustrated in Fig. 3; an observer may *think* they are taking another person’s perspective, but it is very much their own. We will also suggest that future work should always consider including a control condition in which a physical barrier is placed between the agent and target (i.e., Cole et al., 2015). If any effect is truly due to the computation of another’s perspective, any “perspective-taking” effect under examination should be abolished when the agent cannot see the target. Such a control was not, however, included in the Ward et al. work.

The principle of a participant not knowing something that is seen by another agent was also employed in another classic mental imagery experiment (see Pylyshyn, 2002). In this case, the other agent was the participant’s own “mind’s eye.” A variant of this experiment could be employed to assess whether an agent’s perspective is taken. In the basic (mental imagery) experiment, participants were asked to imagine holding a red colored filter right in front of their eyes. They were then asked what colored tint the visual scene has. The answer was, of course, invariably red. To recall, the critical question for the imagery debate was, Why is it red? Is it red because of some intrinsic property of the image that results from the filter, or is it because

participants *know* what happens to a scene when we look through a red filter? Participants were then asked to imagine they were looking through *two* filters, the red filter plus a yellow filter, and then asked, “What colored tint do you now see?” Typically, participants do not know what the answer is or should be. This demonstration is usually taken as evidence to support the Pylyshyn position; when participants “saw” the red tint, whilst holding the red filter only, this was because they knew how the scene should look. They were effectively forcing the scene to look that way based on their knowledge, not because the filter, and resultant scene, helped to generate the correct answer. The standard dot-perspective agent could be presented such that it is viewing stimuli through either one or two differently colored filters. Participants could then undertake, for instance, a Stroop task in which ink color of the word is different for the participant and agent. If the agent’s actual perspective is computed, then RTs should be slower when the participant sees a neutral colored word and the agent sees an incongruent colored word versus the scenario in which both see a neutral colored word. Importantly, the participant sees the same stimulus in both conditions; it is only the agent who sees different stimuli. Variants of this colored filters method could be used with any experimental paradigm that is based on surface color perception. For instance, the color congruency flanker paradigm in which the distracting effect created by letters placed either side of a target letter (Eriksen & Eriksen, 1974) is greater if the letters are the same color as the target letter (Baylis & Driver, 1992). The proposed filters experiments are not much different from the classic mental imagery studies they are based on. This is intentional. As Pylyshyn did so with the mind’s eye, we predict that an agent looking through filters will not modulate responses. This is because, as we have stated, it is not possible to represent the visual experience of another individual.

Readers may have complained (in the [Mental Imagery](#) section) that relying on the mental imagery debate to assist with the question of perspective-taking cannot get us any further because that debate has not itself been solved. Indeed, many disagree with Pylyshyn and instead support the contentions of Kosslyn in his view that mental images are pictorial. Moreover, mental images do appear to help in the reasoning process. If one is asked which is larger, a mouse or an elephant, imagery is not needed. The answer is represented in a semantic form. However, if we are asked which is higher off the ground, the tip of a horse’s tail or its rear knees (Kosslyn et al., 1979), or how many windows are in a room familiar to us, we do seem to use mental images to assist. We do not seem to possess the answer to this in any symbolic form, which then leads to an image. Moreover, despite the homunculus requirement, Finke and Kosslyn (1980) did report that acuity of the “mind’s eye” decreases towards the periphery, as does acuity of our actual vision. However, even if people can use mental images, rather than knowledge, to provide a correct response, and even if Kosslyn turns out to be correct, our job as perspective-taking

researchers is much easier than that faced by Pylyshyn and Kosslyn; we only have to show that experiments such as the filters procedure produce the predicted effect in a perspective-taking paradigm (in people who have no knowledge of how colored filters work) to know whether we can indeed take another’s perspective. It’s an empirical question.

A related notion, and one that joins the mental imagery debate and perspective-taking idea, concerns the important issue of cognitive penetrability. If a process or mechanism of cognition cannot be altered by any information, expectation, or knowledge that a person has, then that process is said to be cognitively impenetrable (see Firestone & Scholl, 2016; Pylyshyn, 1999). Many authors argue that this is the case for low-level vision. Thus, one’s pure perception of a stimulus does not change irrespective of what we believe or are told; the percept remains the same. Recall that Nuku and Bekkering (2008) reported an absence of gaze following when a gazing agent could not see anything. Although these authors did not invoke any notion of visual *experience*, the theory of spontaneous perspective-taking necessarily assumes that the agent’s perspective in these paradigms is taken. The logical conclusion of this is that since the participant is looking at a person who cannot see, their own vision is somehow compromised, and in a way that could be measured with a psychophysical experiment. For instance, when viewing a person wearing a blindfold, thresholds for detecting a low contrast Gabor patch, presented, for instance, in the periphery, should be raised. Note that the theory that vision is noncognitively penetrable predicts no successful manipulation of an observer’s vision. However, one does not have to take a position on vision and cognitive penetrability to apply this test. Again, it’s an empirical question. We will add that many readers may find the suggestion that a participant’s vision could possibly be compromised when viewing a blindfolded individual somewhat ridiculous. We agree. If this is the case, then the reader must necessarily be agreeing with our position; it is not possible to assume the perception of another person.

A brief model of perspective-taking

Despite all the problems we have outlined in the present article, it is certainly true that humans can in some broad sense represent/consider/assume/take a “different perspective” to that of our own. We simply know that the number located on the table in Fig. 2 is seen as 6 to the agent. This is clearly a case of us representing what another person knows—an effect one might call “perspective-taking”.

The present authors contend that when humans “take another person’s perspective, or assume the perspective from any position except their own, the observer, of course, sees the objects and their features as they are themselves currently seeing them. This, as we have stated, has to be the case, and to further reiterate, we reject any notion of imagining the

alternative scene, with its reference to picture generation. We posit that during the perspective-taking process, observers employ basic corrections to account for the different position. Take the red/green squares example described previously and shown in Fig. 3. When you ask your friend to look at the stimuli from a distance, but from your perspective they perceive the features of the stimulus as they currently see them, their percept of the world does not change (i.e., they see a yellow square). However, adjustments that take into account the alternative position can be made for certain object properties. These adjustments are likely to be based on their knowledge and experience of how objects change according to viewing position. The most obvious correction is that objects in a scene appear progressively larger the closer we are to them, and objects that were located peripherally are no longer visible. Thus, the yellow square is represented (symbolically)—that is, inferred, as being larger. Furthermore, *form*, which includes size, can be recomputed; form may thus be said to be part of the correction. It is what might be called a *correction token*.⁴ Thus, an observer will know that a 9 will appear as 6 from an opposite position. A further correction token would be relative position. Prior to the ambiguous number work of Zhao et al. (2015) and Surtees et al. (2016), Tversky and Hard (2009) presented images of an agent sitting at a table, facing the participant, gazing at two objects. Data showed that observers can effortlessly represent where objects are in relation to each other with respect to the agent's position. One could also add *depth* as another correction token, since relative position of stimuli will often (although not always) involve the computation of how close objects are to an agent. The more subtle corrections required to generate the “correct” visual representation are however unlikely to be made. Thus, colors do not become more desaturated, and apparent contrasts are not reduced as distance is increased; color is not part of the correction. It is not a correction token.

If observers do make such corrections, one could empirically examine what adjustments can be made. The basic paradigm would be one in which a participant is required to perform a perceptual judgement (not RT) and told to do this from the perspective of an agent. The experimenter then analyzes what the participant can and cannot correctly reconstruct. As stated, an initial prediction is that form, size, and depth can be reconstructed, but color cannot. Luminance can also be reasonably added to the latter. This whole question is about attribution of vision. What can an observer know about what another person can see? Starting with the seminal work of Piaget and Inhelder (1956), the large literature on the so-called perspective change problem already provides data on one particular correction token. Specifically, relative position of different objects (e.g.,

Huttenlocher & Presson, 1973; Tversky & Hard, 2009). The important point is the emphasis on the notion that perspective-taking is concerned with what other people can see—what we can know that others know about the world from their position. We acknowledge that the proposed work posits a strict criteria (a “high bar”) as to when one should conclude that a perspective has been represented. This, however, is a direct consequence of what it has to mean to represent another's perspective. Such a computation has to mean representing these subtle properties; otherwise, as stated, it would not be a perspective.

Conclusions

In this article we have raised a number of issues concerned with current work on spontaneous perspective-taking. First, we suggest that it has not yet been established whether humans do, in fact, spontaneously represent the viewpoint of others. Although there is certainly evidence suggestive of this, there are too many reports of the perspective-taking-like data when agents cannot see the critical stimuli. At best, the phenomenon seems to occur under limited experimental circumstances. Second, the field has not been clear in describing what exactly it means to assume another's perspective. One consequence is that it is difficult to test the theory because no theory has been forthcoming. A further consequence is that the vocabulary used suggests, indeed states, that humans can *experience* what another person sees. We argue that authors should consider placing greater emphasis on what agent's know, and not on their “visual experience.” Relatedly, researchers need to outline how their conception of perspective-taking differs from previous work suggesting that humans spontaneously *infer* what others see. We have also suggested that the perspective-taking notion raises the same problems that made Kosslyn's position on the mental imagery debate untenable. In other words, some mechanisms or “eye” is required at the alternative position. Finally, we have outlined a number of experiments that will test whether humans can spontaneously represent what another person sees. We have also posited a theory of what actually occurs when a person attempts to take another person's perspective.

References

- Amorim, M. A. (2003). “What is my avatar seeing?”: The coordination of “out-of-body” and “embodied” perspectives for scene recognition across views. *Visual Cognition*, 10(2), 157–199. <https://doi.org/10.1080/713756678>
- Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review* 116, 953–970. <https://doi.org/10.1037/a0016923>
- Baker, L. J., Levin, D. T., & Saylor, M. M. (2016). The extent of default visual perspective taking in complex layouts. *Journal of*

⁴ We borrow this term from motion perception researchers who have often examined so-called correspondence tokens (e.g., Green, 1986). These are those properties of a stimulus that enable the motion system to calculate that an object presented at time T1 is the same object that occurs at a different location at time T2.

- Experimental Psychology: Human Perception and Performance*, 42(4), 508. <https://doi.org/10.1037/xhp0000164>
- Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics*, 51(2), 145–162. <https://doi.org/10.3758/BF03212239>
- Becchio, C., Del Giudice, M., Dal Monte, O., Latini-Corazzini, L., & Pia, L. (2013). In your place: Neuropsychological evidence for altercentric remapping in embodied perspective taking. *Social Cognitive and Affective Neuroscience*, 8(2), 165–170. <https://doi.org/10.1093/scan/nsr083>
- Bertamini, M., & Soranzo, A. (2018). Reasoning about visibility in mirrors: A comparison between a human observer and a camera. *Perception*, 47, 821–832. <https://doi.org/10.1177/0301006618781088>
- Block, N. J. (1981). *Imagery*. Cambridge, MA: MIT Press.
- Bukowski, H., Hietanen, J. K., & Samson, D. (2015). From gaze cueing to perspective taking: Revisiting the claim that we automatically compute where or what other people are looking at. *Visual Cognition*, 23, 1020–1042. <https://doi.org/10.1080/13506285.2015.1132804>
- Capozzi, F., Cavallo, A., Furlanetto, T., & Becchio, C. (2014). Altercentric intrusions from multiple perspectives: Beyond dyads. *PLOS ONE*, 9(12), e114210. <https://doi.org/10.1371/journal.pone.0114210>
- 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. <https://doi.org/10.1016/j.actpsy.2016.01.007>
- Cole, G. G., Atkinson, M. A., D'Souza, A. D., & Smith, D. T. (2017). Spontaneous perspective taking in humans? *Vision*, 1(2), 17. <https://doi.org/10.3390/vision1020017>
- Cole, G. G., Smith, D. T., & Atkinson, M. A. (2015). Mental state attribution and the gaze cueing effect. *Attention, Perception, & Psychophysics*, 77(4), 1105–1115. <https://doi.org/10.3758/s13414-014-0780-6>
- Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, 43(3), 454. <https://doi.org/10.1037/xhp0000319>
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, 292(4), 497–523. <https://doi.org/10.1002/cne.902920402>
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87(3), 327. <https://doi.org/10.1037/0022-3514.87.3.327>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, 16(1), 143–149. <https://doi.org/10.3758/BF03203267>
- Erle, T. M., & Topolinski, S. (2017). The grounded nature of psychological perspective-taking. *Journal of Personality and Social Psychology*, 112(5), 683. <https://doi.org/10.1037/pspa0000081>
- Finke, R. A., & Kosslyn, S. M. (1980). Mental imagery acuity in the peripheral visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 126. <https://doi.org/10.1037/0096-1523.6.1.126>
- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, 39. <https://doi.org/10.1017/S0140525X15002800>
- 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. <https://doi.org/10.1037/0012-1649.17.1.99>
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 1030. <https://doi.org/10.1037/0096-1523.18.4.1015>
- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5(3), 490–495. <https://doi.org/10.3758/BF03208827>
- 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. <https://doi.org/10.1037/xhp0000138>
- Gardner, M., Bileviciute, A., & Edmonds, C. (2018). Implicit mentalising during Level-1 visual perspective-taking indicated by dissociation with attention orienting. *Vision*, 2(1), 3. <https://doi.org/10.3390/vision2010003>
- Gardner, M. R., Hull, Z., Taylor, D., & Edmonds, C. J. (2018). ‘Spontaneous’ visual perspective-taking mediated by attention orienting that is voluntary and not reflexive. *Quarterly Journal of Experimental Psychology*, 71(4), 1020–1029. <https://doi.org/10.1080/17470218.2017.1307868>
- Green, M. (1986). What determines correspondence strength in apparent motion? *Vision Research*, 26(4), 599–607. [https://doi.org/10.1016/0042-6989\(86\)90008-8](https://doi.org/10.1016/0042-6989(86)90008-8)
- Hietanen, J. K. (2002). Social attention orienting integrates visual information from head and body orientation. *Psychological Research*, 66(3), 174–179. <https://doi.org/10.1007/s00426-002-0091-8>
- Hietanen, J. K., & Leppanen, J. M. (2003). Does facial expression affect attention orienting by gaze direction cues? *Journal of Experimental Psychology-Human Perception and Performance*, 29(6), 1228–1243. <https://doi.org/10.1037/0096-1523.29.6.1228>
- Holmes, A., Richards, A., & Green, S. (2006). Anxiety and sensitivity to eye gaze in emotional faces. *Brain and Cognition*, 60(3), 282–294. <https://doi.org/10.1016/j.bandc.2005.05.002>
- Huttenlocher, J., & Presson, C. C. (1973). Mental rotation and the perspective problem. *Cognitive Psychology*, 4(2), 277–299. [https://doi.org/10.1016/0010-0285\(73\)90015-7](https://doi.org/10.1016/0010-0285(73)90015-7)
- Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology*, 1, 213. <https://doi.org/10.3389/fpsyg.2010.00213>
- Kosslyn, S. M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review*, 88(1), 46. <https://doi.org/10.1037/0033-295X.88.1.46>
- Kosslyn, S. M., Pinker, S., Smith, G. E., & Shwartz, S. P. (1979). The how, what, and why of mental imagery. *Behavioral and Brain Sciences*, 2(4), 570–581. <https://doi.org/10.1017/S0140525X00064530>
- Kuhn, G., Vacaiyte, I., D'Souza, A. D., Millett, A. C., & Cole, G. G. (2018). Mental states modulate gaze following, but not automatically. *Cognition*, 180, 1–9. <https://doi.org/10.1016/j.cognition.2018.05.020>
- Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7(1), 12–18. [https://doi.org/10.1016/S1364-6613\(02\)00013-X](https://doi.org/10.1016/S1364-6613(02)00013-X)
- Marshall, J., Gollwitzer, A., & Santos, L. R. (2018). Does altercentric interference rely on mentalizing?: Results from two level-1 perspective-taking tasks. *PLOS ONE*, 13(3), e0194101. <https://doi.org/10.1371/journal.pone.0194101>
- Mathews, A., Fox, E., Yiend, J., & Calder, A. (2003). The face of fear: Effects of eye gaze and emotion on visual attention. *Visual Cognition*, 10(7), 823–835. <https://doi.org/10.1080/13506280344000095>
- Maylor, E. A. (1985). Facilitatory and inhibitory components of orienting in visual space. In M. I. Posner & O. S. M. Morin (Eds.), *Attention and performance XI* (pp. 189–204). Hillsdale, NJ: Erlbaum.

- Michael, J., & D'Ausilio, A. (2015). Domain-specific and domain-general processes in social cognition: A complementary approach. *Consciousness and Cognition*, 18, 434–437. <https://doi.org/10.1016/j.concog.2014.12.009>
- Michael, J., Wolf, T., Letesson, C., Butterfill, S., Skewes, J., & Hohwy, J. (2018). Seeing it both ways: Using a double-cueing task to investigate the role of spatial cueing in Level-1 visual perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance*, 44, 693–702. <https://doi.org/10.1037/xhp0000486>
- Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics*, 68(2), 327–337. <https://doi.org/10.3758/BF03193680>
- Millett A. C., D'Souza, A. D. C., & Cole, G.G. (2019). Attribution of vision and knowledge in 'spontaneous perspective taking'. *Psychological Research*, in press. <https://doi.org/10.1007/s00426-019-01179-1>
- Milliken, B., & Tipper, S. P. (1998). Attention and inhibition. *Attention*, 191–221.
- Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in Human Neuroscience*, 7, 558. <https://doi.org/10.3389/fnhum.2013.00558>
- Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, 132, 297–326. <https://doi.org/10.1037/0033-2909.132.2.297>
- Moors, P., Germeyns, F., Pomianowska, I., & Verfaillie, K. (2015). Perceiving where another person is looking: The integration of head and body information in estimating another person's gaze. *Frontiers in Psychology*, 6, 909. <https://doi.org/10.3389/fpsyg.2015.00909>
- Morgan, E. J., Freeth, M., & Smith, D. T. (2018). Mental state attributions mediate the gaze cueing effect. *Vision*, 2(1), 11. <https://doi.org/10.3390/vision2010011>
- Nielsen, M. K., Slade, L., Levy, J. P., & Holmes, A. (2015). Inclined to see it your way: Do allocentric intrusion effects in visual perspective taking reflect an intrinsically social process? *The Quarterly Journal of Experimental Psychology*, 68(10), 1931–1951. <https://doi.org/10.1080/17470218.2015.1023206>
- Nuku, P., & Bekkering, H. (2008). Joint attention: Inferring what others perceive (and don't perceive). *Consciousness and Cognition*, 17(1), 339–349. <https://doi.org/10.1016/j.concog.2007.06.014>
- O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of 'mudsplashes'. *Nature*, 398(6722), 34. <https://doi.org/10.1038/17953>
- Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica* (Supp. 6), 11–97.
- Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London, UK: Routledge & Kegan Paul.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 531–556). Hillsdale, NJ: Erlbaum.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain sciences*, 1(4), 515–526. <https://doi.org/10.1017/S0140525X00076512>
- Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(3), 341–365.
- Pylyshyn Z. W. (2002) Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25, 157–238. <https://doi.org/10.1080/02643290442000130>
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, 18(3), 849–860.
- 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. <https://doi.org/10.1037/a0018729>
- Santesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Agents and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 929. <https://doi.org/10.1037/a0035175>
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703. <https://doi.org/10.1126/science.171.3972.701>
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1(7), 261–267. [https://doi.org/10.1016/S1364-6613\(97\)01080-2](https://doi.org/10.1016/S1364-6613(97)01080-2)
- Slezak, P. (1995). The “philosophical” case against visual images. In P. Slezak, T. Caelli, & R. Clark (Eds.), *Perspectives on cognitive science, Volume 1: Theories, experiments, and foundations*. New York, NY: Ablex.
- Stephenson, B., & Wicklund, R. A. (1983). Self-directed attention and taking other's perspective. *Journal of Experimental Social Psychology*, 19, 58–77. [https://doi.org/10.1016/0022-1031\(83\)90005-7](https://doi.org/10.1016/0022-1031(83)90005-7)
- 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.
- Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception, & Psychophysics*, 72(3), 695–705. <https://doi.org/10.3758/APP.72.3.695>
- Torralva, T., Kipps, C. M., Hodges, J. R., Clark, L., Bekinschtein, T., Roca, M., . . . Manes, F. (2007). The relationship between affective decision-making and theory of mind in the frontal variant of fronto-temporal dementia. *Neuropsychologia*, 45(2), 342–349. <https://doi.org/10.1016/j.neuropsychologia.2006.05.031>
- Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, 110(1), 124–129. <https://doi.org/10.1016/j.cognition.2008.10.008>
- Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*, 29, 874–880.
- Wilson, C. J., Soranzo, A., & Bertamini, M. (2017). Attentional interference is modulated by salience not sentence. *Acta Psychologica*, 178, 56–65. <https://doi.org/10.1016/j.actpsy.2017.05.010>
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128. [https://doi.org/10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5)
- Zhao, X., Cusimano, C. J., & Malle, B. F. (2015). In search of triggering conditions for spontaneous visual perspective taking.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.