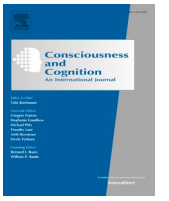
Consciousness and Cognition 102 (2022) 103352

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog



A return of mental imagery: The pictorial theory of visual perspective-taking

Geoff G. Cole [a,*], Steven Samuel [b], Madeline J. Eacott [a]

[a] Centre for Brain Science, University of Essex, UK
[b] Department of Psychology, University of Plymouth, UK

ARTICLE INFO

Keywords:
Mental imagery
Perspective-taking
Perceptual simulation
Pylyshyn
Kosslyn

ABSTRACT

The pictorial theory of mental imagery was a central concern of cognitive science during the latter years of the last century. Proponents of the theory argued that images are reinterpreted by the same processes that act upon perceptual inputs. This idea has recently re-emerged within the context of visual perspective-taking. The *perceptual simulation* theory argues that an observer not only generates an image of what another individual sees but the image is used by the perceptual system in a bottom-up manner. Based on the assumption of Kosslyn and colleagues, we argue that a minimum requirement of a pictorial theory of visual perspective-taking is that observers must faithfully represent relative distance between different points of a scene as would be viewed from an alternative position. The available evidence does not however support this. We conclude that the latest attempt to give mental imagery causal status in a cognitive process is unwarranted.

1. The mental imagery debate

In his essay *Optics,* Descartes (1637/2003) wrote that what is in our head when we perceive an object "bears some resemblance to the objects from which it proceeds". He warned however that "we must not think that it is by means of this resemblance that the picture causes our sensory perception of these objects – as if there were yet other eyes within our brain with which we could perceive it". For some authors at least (e.g., Pylyshyn, 2002), the eyes-in-the-head metaphor re-emerged with the mental imagery debate, a central concern of cognitive science during the 1970 s, 80 s, and 90 s. The debate concerned the fundamental question of how information is represented in the brain/mind.

Kosslyn and colleagues (e.g., Kossyln, Pinker, Smith, & Shwartz, 1979) argued that not only does a mental image look like a picture, i.e., it seems to have a spatial structure, its representational format is also spatial. Kosslyn et al. did however suggest that this medium is functional rather than literal. Evidence for the so-called "pictorial theory" of mental imagery was most often taken to be results from a large series of behavioural experiments showing that the time it takes an observer to move attention around a mental image is related to distance (that was seemingly) traversed. For example, reaction times are longer to move the "mind's eye" from the top of an imagined map to the bottom, as opposed to the middle (e.g., Kosslyn, Ball, Reiser, 1978). This, Kosslyn et al. argued, is because the bottom is further away on the image. In other words, there was deemed to be a one-to-one mapping between what is perceived and the representational medium. Furthermore, rather than being a by-product, mental images were said to play a *causal* role in problem solving and task performance (e.g., Kosslyn et al., 1999). Although it is clearly the case that forming images assists in a variety of tasks

* Corresponding author at: Centre for Brain Science, University of Essex, Wivenhoe Park, CO4 3SQ, UK.
E-mail address: ggcole@essex.ac.uk (G.G. Cole).

https://doi.org/10.1016/j.concog.2022.103352
Received 15 October 2021; Received in revised form 29 April 2022; Accepted 3 May 2022
Available online 19 May 2022

(see Antonietti, 1991), such as 'how many windows are in your house?', the critical issue concerns the processes that are occurring in the brain during imagery.

Amongst the many thousands of words written on the "great debate", the crux of the matter can be found in Pylyshyn (1973). Whilst discussing the picture metaphor, Pylyshyn (p9) stated, "however metaphorically one interprets the notion of picturing a recalled scene in one's mind, the implication is always that whatever is retrieved must be perceptually interpreted (or reperceived) before it becomes meaningful. In other words, the appearance of a memory image precedes its interpretation by the usual perceptual processes…..but what can serve as the input to such a perceptual process? Whatever it is it must be very much like the pattern of sensory activity which takes place at various levels of the nervous system when some sensory event token occurs". It is for these reasons that Kosslyn et al. (1979), when advocating for the pictorial theory, stated that an image can be "looked at" and "inspected". Although Kosslyn and Pomerantz (1977, p54) argued that "images, unlike pictures, are not in need of much fundamental perceptual processing", some such processing is said to occur. Furthermore, the partly unanalysed information that is subject to this kind of processing was said to be stored in long-term memory.

The notion that perceptual processes can act upon mental imagery has re-emerged in the past decade, appearing in some high profile journals (e.g., *Current Biology*). The idea has been applied to the phenomenon of visual perspective-taking.

2. Visual perspective-taking and mental imagery

As described in a review by Flavell (1992), empirical work on visual perspective-taking began with the three-mountains task of Piaget and Inhelder (1956). Children aged 4–11 years were shown a physical model of three mountains and asked to indicate (via the choosing of a photograph) the viewpoint of a doll placed at various locations around the model. Amongst a number of findings, Piaget and Inhelder described how younger children would often choose the photograph that represented their own viewpoint, rather than that of the doll's. One of the many issues that was initiated by this work is the question of how the perspective of another agent influences (adult) responses when the participant is not explicitly asked to consider the alternative viewpoint. This has often been examined with, for instance, the *gaze cuing* paradigm in which reaction time is facilitated when targets that a participant is asked to detect or discriminate are being looked at by an agent (i.e., a "gaze cue") present in a display (Friesen & Kingstone 1998). Importantly, this effect was not simply assumed to occur because the agent shifts an observer's attention to the target location. What the agent actually *perceives* was also shown to be important. For example, Teufel, Alexis, Clayton, and Davis (2010), Nuku and Bekkering, (2008), and Morgan, Freeth, and Smith (2018) all showed that the gaze cueing effect can be reduced or even abolished if the agent faced towards the target but could not see it, if for instance they had their eyes closed. Importantly, these authors argued that observers *infer* or *attribute* seeing to an agent and that this "seeing" is a crucial element of the facilitation effect. Although the whole process concerned the vision of another person, there was no suggestion that an observer represents the sensory aspects or experience of that person.

The conception of visual perspective-taking as an attribution or an inference of anothers' vision has however undergone something of a revision in the past decade. This began with the influential work of Samson, Apperly, Braithwaite, Andrews, and Bodley Scott (2010) who argued that humans spontaneously represents what another person can see. The authors showed that when judging the number of discs located on the walls of a virtual room, a participant's response is slower if an agent, located in the centre, sees a different number of discs. Similarly, Ward, Ganis, and Bach (2019) found that reaction time to discriminate a target letter was not only dependent on how the participant saw the target (in terms of rotation from its canonical view; Shepard & Metzler, 1971) but also on how another agent saw it. The central process involved in these kind of effects has been described as *perceptual simulation* (Ward et al. 2019), and mental imagery is said to be instrumental in the process of assuming an alternative viewpoint. As Bach (2021) put it, "One assumption that is often implicit in how people think about perspective taking is that it involves generating a more-or-less realistic mental image of what we would see if we were in the place of the other person and looked through their eyes". Ward et al. further describe perceptual simulation as the "painting" of a "mental image of the content of another person's viewpoint onto one's own perceptual system". The authors add that in this view of perspective-taking, the process "not only remaps the other's spatial reference frame to one's own (e.g., that one's own left is another's right), but also derives the other's view on an object in the same way that one would perceive it oneself. "Seeing" the content of another's perspective in this manner could then—in a bottom-up fashion—drive all processes that operate on perceptual input". The authors also add that "another's perspective can "stand in" for one's own sensory input". This is then said to enable an observer to "recognize items that would be more difficult to recognize from their own perspective" (Ward et al. 2019). Moll and Kadipassaoglu (2013) describe the supposed image-like representation in these terms; visual perspective-taking generates a "snapshot" of a viewpoint "in a literal, i.e., optical sense of the term". This is all said to occur because visual perspective-taking is "quasi-perceptual" (Ward et al. 2019).

This conception of visual perspective-taking is therefore centrally concerned with the simulation of perceptual experience and the involvement of mental imagery. For this reason, the perceptual simulation account can be described as the *pictorial theory of perspective-taking*. Although the Samson et al. (2010) article motivated much of the recent work, Amorim (2003) was the first author to formally describe how perspective-taking involves mental imagery. This, he argued, included operations that directly correspond to those presented in Kosslyn's theory of mental imagery (e.g., Kosslyn, 1987).

What then must a pictorial theory of visual perspective-taking predict? Although not all picture theorists would agree with this defining criterion, Kosslyn et al. (1979) argued that imagery is 'pictorial' or 'depictive' in the sense that "parts of the surface image correspond to parts of the represented thing, and the interpoint spatial relations among the thing's parts are preserved in the image". In other words, an image should be "capable of preserving relative metric distances between portions of objects" (Kosslyn, et al., 1979). For visual perspective-taking to be similarly pictorial (at least in the Kosslyn sense), behavioural responses (i.e., reaction time) also need to be related to distance 'perceived' by an agent when a participant is attempting to take that agent's viewpoint. For instance,

when an observer looks at a real scene from position X (i.e., not perspective-taking), reaction time may be say 20% slower to move attention from location A to location C, as opposed to moving it from locations A to B (see Eriksen & Murphy, 1987) This effect should also occur when the observer stands at location Y and performs the same task but asked to do so from the perspective of an agent standing at position X. Furthermore, the effect should occur when the observer never stood at location X. In other words, the process should be concerned with perspective-taking as opposed to memory (although memory will of course be sometimes involved). The proposed effect should occur for exactly the same reason as to why mental imagery is (supposedly) pictorial; the representational format codes for real-world distance.

Notice that this threshold for perspective-taking being quasi-perceptual is very low. It concerns the representation of space; the distance between two points. It does not include responses that correspond to, for instance, how an object's colour may change as an agent's viewing position changes. Because a pictorial representation does not necessarily code for depth, it similarly does not also have to include depth information. However, the coding of flat-plane one-to-one spatial information is a necessary condition of mental imagery being involved in visual perspective-taking, because all pictures represent spatial relations. If there is no spatial distance effect then the representational medium is not pictorial, in other words, no perceptual simulation.

3. Perspective-taking research

Despite the vast amount of work in the past decade, we posit that the field of "perspective-taking" is not actually concerned with the perspective of other agents. Because perspective-taking is by definition concerned with an agent's *perspective* (i.e., what they perceive), we argue that the only experiments relevant to perspective-taking are those that attempt to directly measure what would be *perceived* at an alternative viewpoint. Many current paradigms however use *attention* paradigms (rather than perception) with their emphasis on reaction time. In essence, reaction time is used as a proxy to assess what another agent sees, as in the Samson et al. (2010) and Ward et al. (2019) procedures. Standard visual attention work (e.g., attentional cueing; Posner, 1980) is not of course concerned with the *perception* of an observer; it is not about what another person can *see*. Indeed, the vast majority of attention work assesses phenomena that observers are not conscious of themselves (e.g., negative priming; Tipper, 1985). This is because of how attention is often defined, i.e., differential responses to stimuli that cannot be explained by sensory processes. It follows therefore that the predominantly attentional paradigms currently used to examine alternative 'perspectives' are similarly not concerned with the perception of an agent. This assumption is supported by a number of experiments showing that 'perspective-taking' effects still occur when an agent cannot see the critical stimuli. For instance, Cole, Atkinson, Le, and Smith (2016) replicated the Samson et al. (2010) paradigm and data pattern both when the agent could see the discs and when their view was obstructed by a physical barrier. Wilson, Soranzo, and Bertamini (2017) also reported a similar effect when the agent was blindfolded (see also Kuhn, Vacaityte, D'Souza, Millett, & Cole 2018). Whilst other studies *have* reported an effect of obscuring the critical stimuli (Baker, Levin, & Saylor, 2016; Furlanetto, Becchio, Samson, & Apperly, 2016), there are enough *visibility manipulation* studies to suggest that perspective-taking paradigms are not concerned with perspectives.

When one considers the importance of assessing what an agent can *see*, as opposed to how the observer's attention is manipulated, the paradigms and fields most relevant to perspective-taking are those in which an experimenter effectively asks a participant to indicate what they can see from the alternative position. Particularly relevant are studies on object recognition (e.g., Hayward & Tarr, 1997; Farah, Rochlin, & Klein, 1994), mental rotation (e.g., Hochberg & Gellman, 1977), and reference frames (e.g., Tarr & Pinker, 1990). These are all concerned with alternative viewpoints, what is sometimes referred to as 'updating'. Yet the discipline of 'perspective-taking' does not typically consider any of this work. This may be because the field is more associated with a *developmental* tradition rather than *vision*.

Amongst the many interpretations of updating research, the one overriding finding is that observers have some difficulty in being able to consider a different perspective to the one being seen. Reaction times are typically in the seconds (e.g., Bethell-Fox & Shepard, 1988) and the tasks tend to feel effortful. This does not concur with the conception of perspective-taking being "spontaneous" as Samson et al. (2010) suggested. Furthermore, as slow as updating is, there is one aspect of the paradigms used that make them easier to perform than everyday perspective-taking. Participants are often shown a sample and asked to match it to a number of alternative viewpoints (e.g., Shepard & Metzler, 1971). However, when observers consider other viewpoints in everyday situations, it is not a multiple choice task; there are of course no alternative viewpoints to choose from. Instead, the viewer simply attempts to take the alternative perspective. The present authors do however acknowledge that humans are of course far more practiced at attempting to take the perspective of other agents compared to, for instance, rotating Shepard and Metzler –type figures. Indeed, as showed by Bethell-Fox and Shepard (1988), response times in these paradigms do show large practice effects.

Very few *alternative viewpoint* experiments exist that do not use reaction time as a proxy for perspective and do not also use some form of recognition task. One such study is Rock, Wheeler, and Tudor (1989). In order to test the 'object-centred' theory of object recognition, Rock, et al. asked participants to draw a three-dimensional wire frame, located on a table in front on them, as it would look if they were positioned 90 degrees to the side. Results showed that independent judges could not match the drawings to images of the correct viewpoint compared with control drawings. A second experiment ruled out the possibility that participants were unable to faithfully reproduce the stimulus (i.e., poor drawing ability). Samuel, Hagspiel, Eacott, and Cole (2021) also examined the ability of observers to take a perspective without forced-choice recognition. The authors presented participants with images containing an agent looking at two lines pinned to a wall. The participants were asked to judge how long the lines appeared to be from the perspective of the agent. Using adjustable sliders, participants consistently failed to judge that a line closer to another agent would appear larger than a line further away. These findings do not concur with the notion that visual perspective-taking involves the "painting" of a "mental image of the content of another person's viewpoint onto one's own perceptual system".

Another empirical issue directly relevant to the pictorial theory of visual perspective-taking is the question of whether humans can make effective judgments about *their own* pictorial representations (Perdreau & Cavanagh, 2011). As described by Rock's theory of Constructive Perception (Rock, 1983), proximal representations are the 2-D, or 'flat plane', retinal projections that occur before size constancy mechanisms 'correct' the image. Many vision scientists have made the point that visual artists are particularly adept at rendering objects within scenes in accordance with their proximal, as oppose to real, size. For example, one proximal representation principle is that distant objects need to be drawn relatively small. Perdreau and Cavanagh (2011) presented artists and non-artists with two objects, one of which was placed within a linear perspective scene so that it looked smaller. Participants were required to adjust the size of the object until it was the same proximal size as the other. To emphasise the fact that the task was concerned with pictorial-like 2-D size, they were told to imagine they were using their fingers to measure the objects. Results showed that both artists and non-artists consistently overestimated the size of the object presented in linear perspective. The same inability to judge our own proximal perspective was also found by Samuel, Hagspiel, Cole, and Eacott (2021). The paradigm was essentially the same as the two-lines-on-the-wall procedure described above (Samuel et al., 2021) with the exception that observers were not required to take an agent's perspective; they stood in the location themselves.

Experiments that have examined human ability to correctly judge the size of an image projected onto a mirror also reveal how poor observers are at estimating flat-plane pictorial-like representations. For example, Lawson, Bertamini, and Liu (2007) asked observers to estimate the length of sticks when seen in a mirror or through a window. The authors found that size judgements were relatively good (for both window and mirror conditions) when participants were asked to judge the real physical size of the sticks. This was the case irrespective of viewing distance. This effect however contrasted performance when participants were asked to consider the *projected* size of the same objects, as seen on a mirror or a window. Overall, object size was overestimated by approximately 60%. The error occurred despite the authors placing great emphasis on what a judgment of projected size actually means. As Lawson et al. noted, these data concur with the common misconception (see Bertamini & Parks, 2005) that an image of one's face or body in a mirror becomes smaller as the distance between ourselves and the mirror increases. (This can be illustrated with the typical modern phone. When turned off and used as a [black] mirror it tends to show one's head fitting neatly into the rectangular display area. This projection does not change size irrespective of how far away the phone is placed from the observer).

These experiments suggest that not only do observers have difficulty taking another's perspective (i.e., 'updating' research) they also have difficulty considering their own proximal viewpoint. Essentially, these latter tasks index an observer's ability to successfully 'inspect' a pictorial representation of a scene; a representation that faithfully depicts relative size. Results show that observers are not adept at doing this. If observers have difficulty being able to efficiently interrogate their own 2-D image, they are unlikely to be able to interrogate such an image based on another agent's perspective.

4. Propositions rather than pictures

There is of course a paradox. Whilst the above results show how difficult it can be to adopt a different viewpoint, our own experience tells us that we can in many situations successfully "take the perspective of others". Indeed, Tversky and Hard (2009) empirically showed that observers can effortlessly represent where two objects are in relation to each other from the perspective of an agent seen sitting opposite facing them. Similarly, most observers will know that the number "9" will be seen as "6" from an alternative position (but see, Samuel, Eacott, and Cole, 2022, for such performance). Although the present article is primarily a critique of the *perceptual simulation* account and does not propose an alternative theory of perspective-taking (beyond the general principle of *constructing* rather than *looking* at any image that might be generated; see the next section), we suggest that knowledge represented in a propositional form is at least a more parsimonious alternative. This explanation, outlined below, takes the critical element of what some authors (e.g., Hochberg & Gellman, 1977) argued to be the central component of successful mental rotation performance. We also suggest that the difference between the Samuel et al. (two lines) paradigms and that of Tversky and Hard may also reveal the critical factor that enables an observer to take an alternative perspective, but also illustrates why the pictorial theory of visual perspective-taking is untenable.

As employed in cognitive science, propositions are structures that represent exact relations between the entities. For example, in the case of vision, a chair might be coded as being "to the left" of a desk or a car might be represented (amongst other representations) according to its colour, e.g., "the car is red". Although these relations can be described using natural language, propositional structures are abstract. For instance, "the car is red" can be represented in non-language terms such as with morse code or semaphore. Rather than relying on an image to generate a response, participants could use a network of propositions to compute where objects and their parts are in relation to each other from the viewpoint of an agent. Indeed, mental rotation researchers would often refer to "landmarks" that allow objects to be rotated (Hochberg & Gellman, 1977). Although the classic Shepard and Metzler stimuli were relatively abstract, they still possessed distinct features at certain points. One part of the object may for example include a 90 degree corner, whilst another part would include an end point where a protruding part stops. The relationship between these two parts, and others within the object or scene, can then be compared to determine how they will look when rotated. Whilst this can occur when a scene has distinct landmarks (i.e., virtually every scene in the world), including the Tversky and Hard paradigm, it cannot occur when there are no such landmarks. This was the situation in the Samuel et al. procedure because the stimuli landmarks were identical; the only difference was how long the lines were. Importantly, a pictorial representation would have enabled successful perspective-taking in that experiment because a pictorial representation codes for distance, i.e., size (Kosslyn et al., 1979). Thus, the pictorial theory of visual perspective-taking predicts that the distance between the two end points of each line (i.e., their length) presented in the Samuel et al. experiment would have been successfully represented, leading to correct responses. This was not however the case. As with classic mental rotation experiments, participants did not seem to be interrogating an image. Because the lines all possessed the same features (i.e. two ends)

they could not be distinguished so easily. One does have to acknowledge however that different participants often state they use different strategies on the same task, including mental rotation and visual perspective-taking.

Another possibility is that participants applied naïve and erroneous theories of how vision works to the task. For instance, participants may have believed that vision works to 'correct' for the appearance of size differences between identical objects, effectively 'enlarging' the further object relative to the closer one. This could explain why about half of participants in the Samuel et al. paradigm not only failed to judge that the closer line would appear longer to the agent but actually judged that the *further* line as appearing longer. Even if this were contained within a mental image it would certainly not be faithful to the agent's vision, but a product of sheer imagination. In any case it is just as likely that one could arrive at this error without imagery, erroneous or otherwise, but through inaccurate geometric calculation and the application of (false) logic.

There is one further problem with the argument that mental imagery assists visual perspective-taking. This concerns phenomenology. The present discussion has so-far assumed that an image *is* generated when a person attempts to take an alternative viewpoint. It has not however been established whether this is the case. Unlike mental imagery, in which an observer does at least have the strong experience of seeing something that is not literally there, during perspective-taking no alternative experience seems to be generated. Although this claim is based on the authors' own experiences of perspective-taking, and a handful of anecdotal reports from colleagues, the scene remains the same. Of course, *attention* to the relevant objects and parts of a scene changes, as shown with numerous 'perspective-taking' studies (e.g., Samson et al. 2010), but experience of a scene does not seem to change. Indeed, many scenarios that require an observer to consider an alternative viewpoint can use non-pictorial knowledge. For example, when viewing a theatre seating plan such knowledge tells us that back row locations will not be ideal and front row seats may induce some neck straining. Neither of these will require the generation of an image. Although the generation of mental imagery is more associated with the rotation of single objects (during Shepard and Metzler-type paradigms), as opposed to viewing a scene from an alternative location, whether mental imagery is indeed employed during visual perspective-taking is of course an empirical question.

5. A different kind of perceptual simulation

Despite the fact that mental images are conceived of as passive representations that can be "looked at", Pylyshyn often reminded us (e.g., Pylyshyn, 2002) that it is the *observer* who actively generates images. Pylyshyn maintained that when a person, such as a participant in a mental imagery experiment, is asked to generate an image, they *simulate what it would be like* to actually see the object/scene. In this view, the contents of an image are solely based on what an observer believes or knows. Part of the difficulty in appreciating this is that it *feels like* we "look over" images to glean information, and as noted earlier, images do help in the reasoning process. The constructive, as opposed to interpretive, explanation is perhaps more apparent when we imagine a non-static situation such as running through a busy shopping arcade. The image experienced is clearly under top-down control. We are constructing it, not looking at it. If imagery *is* involved in visual perspective-taking, we argue that the process suggested by Pylyshyn can account for its generation. An observer *simulates* what the scene would look like if they were standing at the alternative position. Readers will notice that this explanation seems no different to the pictorial theory of visual perspective-taking. Indeed, it is referred to as the *perceptual simulation* account by proponents of the account. The critical difference is in the way that the representation of the alternative scene is exploited. The *perceptual simulation* account argues that the image is used "in a bottom-up fashion" and that it drives "all processes that operate on perceptual input" (Ward et al. 2019). In other words, during perspective-taking an observer forms a mental image of what the agent is seeing which is then processed by the visual system, thus enabling successful computation of the alternative view.

The problem here is that the notorious homunculus is required; Descartes' metaphorical eyes in the head that "see" the images. This was the problem that plagued the pictorial theory of mental imagery. Kosslyn et al. (1979) argued however that the need for an inner perceptual system was no more the case for mental imagery than it is for standard visual perception. Indeed, the present authors do have some sympathy with this particular Kosslyn et al. position. Irrespective of how an image is projected onto the primary visual cortex, whether through actual perception, migraine aura, or mental imagery, the observer has an experience of seeing something. No homunculus is required for this; activity in the primary visual cortex generates a perceptual experience. However, a homunculus *is* required in the pictorial theory of perspective-taking. The homunculus needs to have access to the agent's perceptual machinery. This is a prerequisite for any perspective-taking system that uses perceptual mechanisms to process information pertaining to an alternative viewpoint "in a bottom-up fashion". One has to remember that the 'picture' we see in front of us when we look out into the world (i.e., vision) is of course based on sensory processing. This includes a whole array of processes from edge detection to colour and luminance constancy (not to mention all the other constancies). One also has to note that the *perceptual simulation* account argues that images do not just happen to co-occur when an alternative viewpoint is attempted, they are said to be instrumental in the process; instrumental in a bottom-up manner. Indeed, if they were not thought of in this way, imagery would not of course be mentioned; images would be no more important than any other phenomenon that just happens to co-occur when a person is asked to take an alternative perspective.

6. Is the "picture theorist" a straw man?

We have effectively argued that there is real danger of researchers re-running the mental imagery debate of the 1970 s, 80 s, 90 s, and early 2000 s with talk of mental imagery assisting visual perspective-taking. In his extensive review of the debate, Pylyshyn (2002) included a section tackling the criticism that he had erected a straw man; that no one actually believes in the pictorial theory of mental imagery. Similarly, the straw man argument has on occasion been put to the present authors. Fellow students of visual perspective-taking will often state that when other authors write that, for instance, observers can represent an agent's vision as being photographic-like ("snapshot") in a "literal" "optical" sense, they don't *literally* mean "literal". However, one can only evaluate what is

said and written rather than what researchers may perhaps have meant and if there is a difference, we believe that it is important to clarify. It is for this reason that we have recently argued that a formal theory of visual perspective-taking is now required (Cole, Millett, Samuel, & Eacott, 2020; see also Cole & Millett, 2019). What cannot be up for dispute however is the notion that mental imagery is involved.

Perhaps the issue has been one of language and terms. It is not always easy to describe the phenomenon in which observers represent what another person can see; their visual experience. Metaphors are always useful. However, we must not fall into the trap of mixing up metaphors with real mechanisms. As Pylyshyn (2002, p180) noted, "it is always the picture *metaphor* that people retreat to in the face of the implausibility of the literal version of the picture theory". Fortunately, there is enough non-metaphorical description of the perceptual simulation notion to allow predictions to be tested. For instance, Ward et al. (2020) wrote that the process allows observers to "recognize items that would be more difficult to recognize from their own perspective". This is an empirical question.

Supporters of the pictorial theory also have to explain why the change in the way that visual perspective-taking is now described has occurred. After all, gaze cueing research, well-known to perspective-taking researchers (who cross over into developmental work), had already been arguing that the gaze following effect was not just an attention shift effect but due to observers *attributing* seeing to an agent (Nuku & Bekkering, 2008; Teufel, et al., 2010). What has occurred in the past decade that warrants a reconception of visual perspective-taking? It can only be the belief that the more recent experiments actually show something more than mere attribution of seeing.

7. Conclusions

The notion that mental images can be 'reperceived' or 'reinterpreted' has a long and difficult history. The latest incarnation of the idea suggests that images are exploited by the perceptual system to aid visual perspective-taking. This does not however concur with the available evidence. Paradigms that index what another agent can see, as opposed to where the observer's attention is directed, reveal that the representation of an alternative viewpoint is not based on pictorial information. It is also unclear how visual perspective-taking mechanisms can make use of any image that the observer forms, if observers do indeed generate mental images. How could the content of another's perspective initiate "processes that operate on perceptual input" in a "bottom up fashion"? What can serve as the input to the perceptual process that are said to occur? To paraphrase Pylyshyn (1973; see the present Section 1), whatever it is it has to be like the pattern of sensory information that the visual system operates upon during visual perception. Indeed, the *perceptual simulation* account of visual perspective-taking does explicitly state that all the processes that operate on perceptual input also operate on the content of another's vision. One positive aspect of the *perceptual simulation* account is that it has encouraged researchers to consider the question of what is actually occurring when a person attempts to adopt the perspective of another agent. Incorporating a causal role for mental imagery into any explanation is however unwarranted.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Geoff G. Cole: Conceptualization, Writing - original draft. **Steven Samuel:** Writing - review & editing. **Madeline J. Eacott:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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*, 157–199.
Antonietti, A. (1991). Why does mental visualization facilitate problem-solving? *Advances in Psychology, 80*, 211–227.
Bach, P. (2021). Laboratory website. https://www.actionprediction.org/perspective-taking Accessed June 2021.
Baker, L. J., Levin, D. T., & Saylor, M. M. (2016). The extent of default visual perspective taking in complex layouts. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 508–516.
Bertamini, M., & Parks, T. E. (2005). On what people know about images on mirrors. *Cognition, 98*, 85–104.
Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 12–23.
Cole, G. G., Atkinson, M., Le, A., & Smith, D. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica, 164*, 165–168.
Cole, G. G., Millett, A., Samuel, S., & Eacott, M. (2020). Perspective taking: In search of a theory. *Vision, 4*, 30.
Cole, G. G., & Millett, A. (2019). The closing of the Theory of Mind: A critique of spontaneous perspective-taking. *Psychonomic Bulletin & Review, 26*, 1787–1802.
Descartes, R. (1637/2003). La dioptrique, in: Descartes, R., Discours de la méthode, Leyden: Ian Maire.
Eriksen, C. W., & Murohy, T. D. (1987). Movement of attentional focus across the visual field: A critical look at the evidence. *Perception & Psychophysics, 42*, 299–305.
Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*, 490–495.
Farah, M. J., Rochlin, R., & Klein, K. L. (1994). Orientation invariance and geometric primitives in shape recognition. *Cognitive Science, 18*, 325–344.
Flavell, J. H. (1992). Perspectives on perspective taking. In H. Beilin, & P. B. Pufall (Eds.), *Piaget's theory: Prospects and Possibilities*. New Jersey: Lawrence Erlbaum.

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*, 158–163.
Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1511–1521.
Hochberg, J., & Gellman, L. (1977). The effect of landmark features on mental rotation times. *Memory & Cognition, 5*, 23–26.
Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review, 94*, 148–175.
Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 47–60.
Kosslyn, S. M., & Pomerantz, J. R. (1977). Imagery, propositions and the form of internal representations. *Cognitive Psychology, 9*, 52–76.
Kossyln, S. M., Pinker, S., Smith, G. E., & Shwartz, S. P. (1979). The how, what, and why of mental imagery. *Behavioral and Brain Sciences, 2*, 570–581.
Kuhn, G., Vacaityte, I., D'Souza, A., Millett, A., & Cole, G. G. (2018). Gaze following and mental state attribution. *Cognition, 180*, 1–9.
Lawson, R., Bertamini, M., & Liu, D. (2007). Overestimation of the projected size of objects on the surface of mirrors and windows. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1027–1044.
Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in Human Neuroscience, 7*, 558.
Morgan, E. J., Freeth, M., & Smith, D. T. (2018). Mental state attributions mediate the gaze cueing effect. *Vision, 2*, 11.
Nuku, P., & Bekkering, H. (2008). Joint attention: Inferring what others perceive (and don't perceive). *Consciousness and Cognition, 17*, 339–349.
Perdreau, F., & Cavanagh, P. (2011). Do artists see their retinas? *Frontiers in Human Neuroscience, 5*, 171.
Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London, UK: Routledge & Kegan Paul.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.
Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25*, 157–238.
Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
Rock, I., Wheeler, D., & Tudor, L. (1989). Can we imagine how objects look from other viewpoints? *Cognitive Psychology, 21*, 185–210.
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*, 1255–1266.
Samuel, S., Hagspiel, H., Eacott, M., & Cole, G. G. (2021). Visual perspective-taking and image-like representations: We don't see it. *Cognition, 210*, Article 104607.
Samuel, S., Hagspiel, H., Cole, G. G., & Eacott, M. (2021). 'Seeing' proximal representations: Testing attitudes to the relationship between vision and images. *PLoS ONE, 16*, Article e0256658.
Samuel, S., Eacott, M. & Cole, G. G. (2022). Visual perspective taking without visual perspective taking, *Journal of Experimental Psychology, Learning, Memory, and Cognition,* (forthcoming).
Tarr, M., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science, 1*, 253–256.
Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental state attribution drives rapid, reflexive gaze following. *Attention, Perception, & Psychophysics, 72*, 695–705.
Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 37*, 571–590.
Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition, 110*, 124–129.
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 sentience. *Acta Psychologica, 178*, 56–65.