What Piaget didn't know: The paradox of Visual Perspective-taking.

Geoff G. Cole[1] and Steven Samuel[2]

1. Geoff G. Cole, Senior Lecturer, Centre for Brain Science, University of Essex, ggcole@essex.ac.uk. https://www.essex.ac.uk/people/COLEG19909/Geoff-Cole
2. Steven Samuel, Senior Lecturer, Department of Psychology, City St. George's University, London, steven.samuel@city.ac.uk. https://www.citystgeorges.ac.uk/about/people/academics/steven-samuel

Abstract:

Research on visual perspective-taking, the ability to know what another person can see, effectively began with the work of Piaget in the 1950s. Although many authors have examined the various aspects of the phenomenon, it has never been made clear what it exactly means to take another's viewpoint and what processes occur when a person attempts to do so. Here, we suggest that the current conceptions of perspective-taking do not reflect the empirical data. These data not only show that humans are poor at knowing what others can see, we even have great difficulty reflecting upon our *own* perspective. We conclude that a theory of visual perspective-taking is needed.

Keywords: Perspective-taking; Piaget; Theory of Mind; Mental imagery

Seven years ago the present authors carried out what might be called a *Google count experiment*. We were interested in how many hits the following search terms would generate: "theory of attention", "theory of memory", "theory of emotion", "theory of motivation", and "theory of intelligence". These produced 16.5, 13.5, 7.5, 7.3, and 7.4 million hits respectively. These huge numbers are not surprising. If there is anything out there *Google* will find it. We then made the searches more niche by typing in: "theory of motor control", "theory of visual working memory", "theory of episodic memory", "theory of perception and action", and "theory of social perspective-taking". These returned 372, 78.4, 151, 375, and 22.2 thousand respectively. Although far less than the millions generated by the previous searches, tens and hundreds of thousands of hits were still found.

These searches were all intended as a comparison for our critical 'trial' in which we typed in "theory of visual perspective-taking". This returned one single hit, a reference to a peer-reviewed article by Qureshi, et al. (2010). Being relatively new to the field of perspective-taking, and the broader issue of *Theory of Mind*, we were extremely surprised, especially when one considers that the closely related "theory of social perspective-taking" returned over 22 thousand hits. We began to realise that no one, including the Qureshi et al. authors, seem to have tackled the question of what it means to take another person's perspective. What exactly is happening when we say to ourselves "I wonder what they are looking at"?

What does visual perspective-taking actually mean?

As many of us know, research on visual perspective-taking began with the seminal work of Piaget and Inhelder in the 1950s (Piaget & Inhelder, 1956). They had a table-top model constructed showing three mountains adjacent to each other. On the top of one was a house, on another a cross, and a third had no object. Piaget and Inhelder would then place an agent (i.e. toy doll or bear) at

different locations around the model and a child was asked to decide how the scene looks from those alternative positions. This allowed the authors to examine, for example, how perspective-taking ability develops with age. As Piaget and Inhelder did, the many thousands of later perspective-taking researchers tended to examine this ability in terms of *function*. What has not been addressed is the question of *representation*.

When we take the perspective of another agent, what processes are occurring? Ward, et al. (2019) summarised thinking at the time; "A recent proposal is that VPT [visual perspective-taking] takes a (quasi-)perceptual form, "painting" a mental image of the content of another person's viewpoint onto one's perceptual system that can stand in for one's own perception" (p.874). Ward et al. further added that perspective-taking "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" (p.874). Others (e.g. Moll & Kadipasaoglu, 2013) refer to the representation as being like a "snapshot" that occurs "in a literal, i.e. optical sense of the term" (p.1). The notion of representing others' phenomenal *experience* is also commonly invoked. Thus, Capozzi, et al. (2014) stated that, "in simple perspective-taking tasks, one's own and others' visual experience influence each other" (p.1). This is all unequivocal. When we take another's visual perspective, the argument is that we represent what they can *see* in a "literal", "optical sense". That is why it's called *perspective-taking*. We represent the *percept* of others.

We are poor at taking other perspectives.

The problem with these conceptualisations however is that humans are poor at knowing what another person can see. Look at the adjacent figure and take the visual perspective of the man looking into the mirror. Put yourself in his shoes and decide whether he/you can see his/your outstretched hand in the reflection. What about his/your waist? It's a trick question that we have enjoyed asking at many a conference presentation in the past five or six years. The man cannot see himself at all. This little-known phenomenon, sometimes referred to as the *Venus effect*, is commonly used in the film and TV industry. When we see a character looking at themselves in a mirror the actor can't actually see themselves. If they could we the viewer would see the camera. (This is true unless they are standing only a couple of inches from the mirror). In experimental psychology it is usually taken as an example of how poor humans are at reasoning with mirrors (i.e. angles of reflection; Bertamini & Soranzo, 2018). We however take it as a good example of how poor humans are at taking another's visual perspective.

In our view, the only experiments that reveal visual perspective-taking processes are those in which participants are *explicitly asked*, or effectively asked, to make a decision on what a display would look like from a different position. One of the oddest aspects of the visual perspective-taking field is that there is very little adult empirical work in which this occurs. Indeed, one has to look at other fields to find experiments that directly index what can be seen from an alternative location. One such example is Rock, et al. (1989). Participants sat at a round table which had two additional (and empty) chairs placed 90° to the left and 90° to the right of the participant's position. In the middle of the table was a bent wire frame and the task was to decide what the frame would look like if the participant was seated in one of the other two chairs. Ability to take the alternative viewpoint was measured in various ways across five experiments. For example, in one variant, participants

were asked to decide which of four alternative images of the frame was the correct viewpoint. Rock et al. found that performance was little better than chance. The difficulty of taking an alternative viewpoint is also supported by results from mental rotation experiments. Here, the tasks tend to feel effortful and reaction times are typically in the seconds (e.g. Bethell-Fox & Shepard, 1988).

The *Venus effect*, Rock et al. (1989) findings, and mental rotation experiments, do not however concur with an abundance of other more recent work arguing that not only can we effectively represent others' vision but this occurs *spontaneously*. This view comes from a number of experiments, beginning with Samson, et al. (2010), in which observers are required to detect and/or discriminate a target presented in a visual display. The important manipulation is that the display also includes an image of a human agent who, on some trials, happens to be looking at the target the participant has to detect/discriminate. On other trials, this agent is *not* looking at the target or, in some variants of this paradigm (Cole, et al. 2016), they are not able to see the target because their view is obstructed with, for instance, a physical barrier. Results consistently show that reaction time to process the target is facilitated when the agent is facing towards it. The common assumption, that these kind of results reveal spontaneous perspective-taking, is because these effects occur when participants are not explicitly attempting to take the agent's perspective. They are instead simply trying to discriminate the target/s as soon as they can.

How can we explain the difference between the findings from these 'spontaneous perspective-taking' experiments and the Rock et al. results, as well as the *Venus effect*? The reason is that the former are not measuring perspective-taking. Indeed, it is difficult to know what exactly is being shown in those experiments and a number of authors, including the present two, have attempted to answer this (e.g. Cole et al. 2016; Conway, et al., 2017; Samuel et al., 2021b). One idea is that the agent acts as an attention shifting cue that moves attention towards the target/s. In this scenario, reaction time will necessarily be facilitated when the agent is looking at the same object (i.e. target) as the participant. We should add that the attentional shift notion was mentioned by Samson et al. (2010) in the first account of spontaneous perspective-taking. They presented the idea as a *mechanism* that could assist the process. However, one person's explanation is another person's confound. If the so-called spontaneous perspective-taking effect is indeed due to the agent shifting attention towards the target this would not constitute a perspective-taking phenomenon, it would be an attentional cueing effect.

We are poor at taking our <u>own</u> perspective.

There is an episode of *Father Ted* in which *Dougal* is having great difficulty understanding why the cows he can see in a field are tiny. *Father Ted* tries to explain that they are not tiny, just far away. *Dougal* is managing to achieve what most of us have great difficulty with. He can reflect upon his own perspective and override size constancy mechanisms such that a scene to him is viewed in 'proximal', 2-D, x, y co-ordinates.

Asking whether a person can reflect upon, i.e. take, their own perspective is an odd thing because we rarely have to consider our perception. The world is there in front of us. However, when we are required to do so, it is difficult for many people. This is what we have to do when we draw a scene we are looking at, a task which many find difficult. One has to reduce the scene to retinal x, y co-ordinates. This is why the drawing aid known as the *grid technique* exists. It allows the difficulty of representing depth to be more easily achieved. For example, two objects which may be far from each other in depth may be immediately adjacent to each other in 2-D retinal co-ordinates. The very fact that many of us find drawing difficult and the grid technique exists shows how difficult it is for us to reflect upon and take our own perspective. This is supported by the empirical literature.

Perdreau and Cavanagh (2011), for example, presented participants with two objects, one of which was located within a linear perspective display making it look smaller. The task was to adjust the size of the object until it was the same proximal/retinal size as the other. The experimenters emphasised the need to consider the items as pictorial 2-D-like. This was achieved by telling the participants to imagine they were using their fingers to measure the objects. Results showed that the size of the linear perspective object was consistently overestimated. Moreover, Samuel and colleagues (2021a) showed that adults are resistant to agreeing that their visual experience is like a photograph they themselves had just taken. The difficulty of taking our own perspective is of course going to be magnified when we attempt to consider the perspective of another agent.

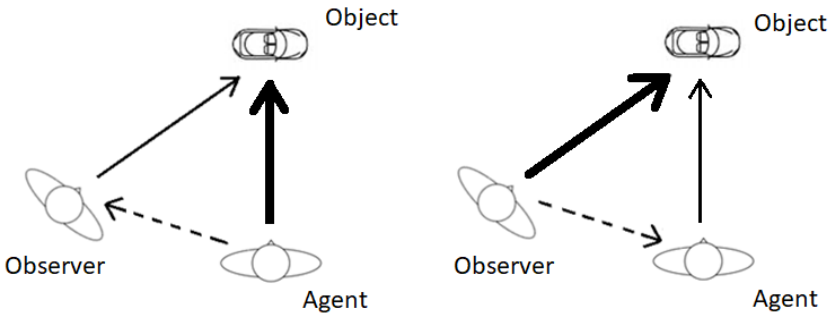
Not only does this challenge the notion that we can effectively take the visual perspective of others, it also fails to support the recent idea that mental imagery assists the process. Ward et al. (2019) have proposed what they refer to as *perceptual simulation* to account for the mechanism enabling visual perspective-taking. The authors stated that visual perspective-taking, "takes the form of a perceptual simulation that can drive subsequent processing like actual input, allowing it to be integrated into recent perceptual accounts of imagery". One of the central findings from mental imagery research is that reaction time to traverse an image with our "minds eye" is linear with the "distance moved". As many older readers will remember, the interpretation of this well-established observation was the very basis of the so-called *great debate* (or the *Kosslyn-Pylyshyn debate*) that raged within cognitive science during the 1970s, 80s, and 90s. If mental images are indeed involved in visual perspective-taking, then findings from visual perspective-taking experiments should reflect those from mental imagery work. That is, the perspective-taking representation should also be in the form of a pictorial 2-D format. However, as described, humans are poor at using x, y co-ordinates when attempting to reflect upon their own perspective. Furthermore, Rock et al. (1989) addressed the possibility of visualisation being used in their bent wire frame experiment described above. They concluded that, "Subjects are unable to perform this task unless they make use of strategies that circumvent the process of visualization" (p.185).

Although the present authors have been critical of the mental imagery idea it does at least represent an attempt to describe what exactly is happening in visual perspective-taking.

A theory of visual perspective-taking is now needed.

There is however a paradox. As difficult as visual perspective-taking is (as described), humans *do* have a good impression of what someone else can see. We are not totally blind to the viewpoint of others. Furthermore, judgements on alternative perspectives are better when we are asked to take an *agent's* perspective, as opposed to a non-agent (e.g. imagine sitting in a different chair). However, whatever is happening, the representation is not visual. As we have pointed out previously (Cole, et al. 2022), the suggestion that mental imagery is involved generates the same problems that afflicted one side of the mental imagery debate. In line with Pylyshyn's view in that debate (Pylyshyn, 2002), we suggest that when humans take another agent's perspective, we are representing their *knowledge* of a scene rather than their *vision*. This knowledge tells us that, for instance, an object which may be on our right is on the left of an agent. We have therefore rejected the Ward et al. (2019) notion that visual perspective-taking is more than simply remapping other's spatial reference frame to one's own (e.g. that one's own left is another's right).

What is now needed is a theory of visual perspective-taking. A formal account of what might be happening. As we explain in Cole, et al. (2020), there are a number of issues that any such theory must consider. It must, for example, tackle the issue of attention. What a person perceives, i.e. there *own* perception, is very much based on what they are attending to (see Lamme, 2003) and no one else can know what their internal spotlight is focused on. At the most fundamental level, any theory must also specify the relationship between the perspective taker, the agent, and the object/s in view. The left panel of the adjacent figure is based on the current conception, although, as noted, it is not entirely clear what that conception is. Here, primacy is given to the agent's view with the observer's (i.e. the perspective-taker) view being downgraded. Something is then taken from the agent, as shown with the dotted line. The right panel is an alternative. Here, the image reminds us that primacy must be given to the observer and we attribute a perspective to the agent. Any model should also stipulate what attributes of a visual scene can and cannot be reliably resolved from the alternative view.



Conclusions

In the past decade the two present authors have published dozens of experiments across a number of peer-reviewed papers assessing visual perspective-taking. In these articles we have raised many concerns with the interpretations of empirical work and in the ways the phenomenon has been conceived. For example, far from being 'spontaneous', tasks in which an alternative viewpoint needs to be computed, such as mental rotation, are effortful and difficult. Furthermore, although the attempt to take another's' perspective seems like a straight-forward notion, it is not entirely clear what processes are involved. In contrast to what has been argued by many other authors, we suggest that humans cannot *simulate* another agent's perception. We do not represent a "snapshot" of an alternative view "in a literal, i.e. optical sense of the term". Instead, we have suggested that observers generate knowledge of what another agent is likely to be seeing.

Given that visual perspective-taking effectively began with empirical research by one of the most eminent cognitive psychologists to have ever lived, it is surprising that Piaget did not elaborate on what is happening when a person attempts to take an alternative viewpoint. His work on perspective-taking, like the vast majority of that which followed him, was concerned with function. Consequently, no theory or description of representation has ever been put forward. In August 2025, the authors repeated the *Google count experiment* described at the beginning of this article. Whilst the comparison terms largely generated the same number of hits, "theory of visual perspective-taking" had jumped from the one single hit to 22. Whilst falling a little short of the 16.5 million for "theory of attention", it's a start.

References

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

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.

Capozzi, F., Cavallo, A., Furlanetto, T., & Becchio, C. (2014). Altercentric intrusions from multiple perspectives: Beyond dyads. *PLoS One*, *9,* e114210.

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., Samuel, S., & Eacott, M. (2022). A return of mental imagery: The pictorial theory of visual perspective-taking. *Consciousness & Cognition, 102*, 103352.

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*, 454-465.

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7,* 12–18.

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

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.

Pylyshyn Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25*, 157–238.

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

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

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

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